



NATIONAL RESEARCH UNIVERSITY
HIGHER SCHOOL OF ECONOMICS

Ilya Prakhov, Denis Sergienko

MATCHING BETWEEN STUDENTS AND UNIVERSITIES: WHAT ARE THE SOURCES OF INEQUALITIES OF ACCESS TO HIGHER EDUCATION?

BASIC RESEARCH PROGRAM

WORKING PAPERS

SERIES: EDUCATION
WP BRP 45/EDU/2017

SERIES: EDUCATION

Ilya Prakhov¹, Denis Sergienko²

MATCHING BETWEEN STUDENTS AND UNIVERSITIES: WHAT ARE THE SOURCES OF INEQUALITIES OF ACCESS TO HIGHER EDUCATION?³

It is assumed that a perfect balance between student academic achievement and university quality is beneficial both for students and higher education institutions (HEIs). Matching theory predicts the existence of perfect matching between the two groups in the absence of transaction costs associated with university enrollment. However, in this study we show cases of mismatch situations in Russia under the Unified State Exam (USE) – the standardized student admission mechanism. This research studies the reasons for this phenomenon for minimal transaction costs and the emergence of unequal access to HEIs. Based on data on Moscow high school graduates who entered university, the determinants of the mismatch between the quality of universities and applicant abilities are assessed. It is shown that although in most cases favorable matching results are established, the individual student achievement results themselves are subject to the influence of school and family characteristics. Thus, inequality of access can be formed at stages preceding HEI enrollment.

JEL Classification: I21, I24, I28

Keywords: matching, mismatch, admission, accessibility of higher education, the Unified State Exam

¹ Ph.D., Research Fellow at Center for Institutional Studies, National Research University Higher School of Economics, Moscow, Russia. ipra@inbox.ru

² Research Assistant at Center for Institutional Studies, National Research University Higher School of Economics, Moscow, Russia. denis-sergienko@ro.ru

³ The article was prepared within the framework of the Basic Research Program at the National Research University Higher School of Economics (HSE) and supported within the framework of a subsidy by the Russian Academic Excellence Project '5-100'.

Introduction

During the past decade, a number of papers on the accessibility of higher education under the recently introduced Unified State Exam (USE) have appeared in Russia. Researchers examined various aspects that simplify or restrict access to higher education. In general, the research on this topic is devoted to the study of the effectiveness of specific educational strategies within the framework of the transition from high school to university. For example, it was shown that students from the most affluent families can benefit from more effective investment in pre-entry coaching and, therefore, they can have greater opportunities for university choice [Uvarov, Yastrebov, 2014; Prakhov, Yudkevich, 2017]. Despite the high prevalence of attending extra classes in general [Burdyak, 2015], applicants from disadvantaged households find themselves in a less favorable position, because they have less resources to pay for pre-entry courses or classes with tutors and they often choose inefficient ways of additional training. Another barrier to higher education is educational mobility. Although the introduction of USE increased the level of educational mobility of university applicants [Slonimczyk et al, 2017], in some cases family income and characteristics of regional socio-economic and educational development can have a strong influence on the decision to move and accordingly, on the accessibility of higher education in general [Prakhov, Bocharova, 2016; Anikina et al., 2014; Pitukhin, Semenov, 2011]. Finally, the characteristics of the family and regional factors are important predictors of university choice in the context of its selectivity (a choice of elite vs non-elite HEIs) [Prakhov, 2016a]. Recent studies emphasize the role of family factors (mainly income and parental education) in choosing a university and the accessibility of higher education in general. In other words, not only the individual results of USE determine educational outcomes, but also the type of institution and the admission strategy.

Nevertheless, USE is the main mechanism for admission to universities.⁴ In an ideal situation the applicant's academic achievement, expressed in her individual USE scores, will correspond to the level of the university where she will enroll. This situation is considered the most desirable, because the future student enters a university which fully corresponds to her abilities. In addition, when the levels of student and university quality match (regardless of the level of student achievement), she has the greatest chance of graduation [Light, Strayer, 2000]. However, other situations are possible when such an *ideal (perfect) matching* is absent. First, the applicant's ability might be lower than the level of university selectivity, i.e. less

⁴ The proportion of students enrolled through the alternative pathways is relatively small: in our sample around 87% of students were admitted solely on the basis of their USE results.

capable applicants may enter universities with higher requirements for students (*overmatching*). In this case, it can be assumed that the student being enrolled occupies the place of a more capable student, which reduces the efficiency of resource allocation in higher education. Such a situation is, however, favorable for the applicant, since studying in a more selective institution (and accordingly, being surrounded by more capable students) will allow her to invest in human capital and get a higher return on higher education.

A less favorable situation is the reverse, because it may indicate the existence of inequalities in access to higher education and unfair distribution of entrants among HEIs. This is when a more capable entrant chooses a less selective institution (*undermatching*) and receives an education of lower quality than that she could claim on the basis of her individual abilities. Therefore, we think it important to understand why such mismatches between the applicant's abilities and the university quality can occur, and what the reasons for the emergence of undermatching for USE and standardized admission procedure are

This paper studies the factors determining the probability of such undermatching. For the Russian higher education system the study of matching mechanisms as the final result of selection on the basis of USE scores is especially relevant since the institution of admission to higher education underwent significant changes in 2009. The earlier system of separate double high school and university examinations (when high school graduates were first required to sit the final school exams, and then, during the admission campaign they had to pass other, university-specific exams) was replaced by USE on different subjects which is passed at high schools and recognized by all Russian HEIs. Thus, in the context of a significant institutional transformation, it is important to know what potential barriers (other than those described above) can limit access to higher education with a standardized admission system.

Thanks to the unification of the evaluation criteria and the assessment scale, the introduction of USE opened up, among other things, ample opportunities from a research perspective. With the help of USE, it is now possible to assess the contribution of various factors on the effectiveness of high school graduates and to compare the quality of teaching among schools, and the quality of admission to various HEIs. In particular, USE makes it possible to evaluate how effectively university selection works, which factors determine the 'failures' in the functioning of this mechanism, and what barriers to higher education exist. These questions are addressed in this paper using a methodology of matching mechanisms and an educational production function.

This study analyzes the mechanism of matching between the quality of HEIs and the abilities of students admitted to these institutions under USE. The paper is organized as

follows. Section 1 deals with theoretical and empirical studies devoted to matching quality. We also present research on factors affecting student performance which can act as possible barriers to higher education. Section 2 describes the empirical model. Section 3 presents the results of regression analysis, where we evaluate the factors that affect undermatching when enrolling in HEIs. It is shown that although in most cases favorable matching situations are established between individual USE results and the quality of the university, the USE scores themselves are subject to the influence of school and family characteristics. Section 4 concludes.

1. Matching and the accessibility of higher education

Matching models are used in different fields, such as the analysis of the labor market in medicine [Roth & Peranson, 1990; Roth, 1984], sport [Frechette et. al., 2007], HR [Roth, 1991; Haruvy et. al., 2006], as well as in studies of university admission. The research methodology of the current study is based on two methodological approaches: a two-sided matching model between applicants and universities, and the educational production function. In this section we start with the theoretical propositions of matching and describe the definition and the conditions for a stable match. Then we apply the matching approach to university admission. After we describe possible non-equilibrium mismatch situations and their determinants based on the results of empirical studies. Next, we provide the links on how matching is concerned with the accessibility of higher education and introduce the educational production function and its determinants.

The model of matching was first proposed by Gale and Shapley [1962]. They formulated a set of conditions for matching, according to which a match between the university and the entrant is a situation where: (1) each applicant is either admitted to one of the universities or remains unadmitted; (2) each university either admits a certain number of students, or does not accept anyone; (3) the applicant submits the application and enrolls in the university, and the university accepts her; (4) the number of students admitted to universities is lower or equal to the quota of available places in the university.

Under the standardized admission test (in our case – USE) an applicant sends her applications to her desired universities and waits for the admission results. The universities rank the applications on the basis of USE scores and make an offer to the eligible applicants (their number should not exceed the quota). The applicant either accepts one of the offers, if there are more than one, or remains unadmitted. The resulting distribution between students and universities is called matching.

According to Gale and Shapley [1962], the assignment of applicants, or matching, can be either stable, or unstable: “A stable assignment is called optimal if every applicant is at least as well off under it as under any other stable assignment” [ibid, p. 10]. For USE we call the optimal assignment a perfect match when applicant ability corresponds to university quality. However, unstable situations or matching imperfections may occur and this problem deserves special attention. There are papers on the quality of matching and the phenomenon of mismatches between student ability and the quality of the university. Such a mismatch is often regarded as a sign of irrational admission to higher education and as a signal of inefficiency and injustice in the higher education system [Bowen and Bok, 1998; Cooper and Liu, 2016]. Moreover, it is empirically proven that a match is one of the determinants of college graduation rates: students whose skill levels match the university quality have a much better chance of successfully graduating than those undermatched [Light and Strayer, 2000]. Another important note to the matching to the universities of a better quality is a correlation between college quality, completion rates and future earnings [Dillon and Smith, 2016].

Since matching imperfections are regarded as adverse admission outcomes, we need to find out what the reasons for this imperfect situation are. In other words, why bright high school graduates choose HEIs of a lower quality than those that could accept them, and on the contrary, how some students with a relatively low level of ability manage to get into top universities. Most studies examine two alternative reasons for mismatches, often called ‘money’ or ‘grit’ [Cooper and Liu, 2016]. The first reason is concerned with borrowing constraints: income plays an important role in college choice [Griffith and Rothstein 2009; Smith, Pender, and Howell 2013; Lincove and Cortes, 2016]. For example, Hoxby and Avery [2013] found that students with a high level of educational achievement from low-income families are more likely to submit applications and consequently enroll to HEIs offering a relatively lower level of education than they could claim. The authors identified several possible reasons for this type of mismatch between able students from low-income families and universities. The main factors for such a mismatch are either a lack of awareness about their admission opportunities or the presence of possible cultural, social or family problems, which drive these students away from universities of a quality corresponding to their abilities.

The second factor which determines mismatches during the admission campaign is grit which is influenced by non-financial parental influence and/or tastes for education [Cooper and Liu, 2016]. For example, the role of information asymmetry in studies of matching in university admission is cited in a number of works [Hoxby and Turner, 2013; Dillon and Smith, 2017; Lincove and Cortes, 2016]. Dillon and Smith conclude that more informed students are much less likely to enter colleges whose educational quality is lower than

students' own abilities, and on the contrary, these students are more likely to enter colleges of a higher level, unlike their less informed colleagues. Hence, one of the main predictors of mismatch are family characteristics related to financial wellbeing, e.g., family income, and cultural and social capital, and information awareness, which can be dependent on other family factors. In our study in addition to the family characteristics we examine the relationship between individual achievement, school characteristics and the probability of a mismatch.

We argue that individual achievement is closely related to the balance between perfect matching and undermatching. Student educational progress itself is not a random variable, but depends on a number of indicators. Since students are admitted to HEIs on the basis of their USE results, factors influencing educational achievement may also have an impact on access to higher education. In other words, in our framework a mismatch is not the only mechanism of unequal access to higher education: inequalities may be set up before admission and influence the accessibility of higher education through the exam scores. Consequently, we need to estimate what characteristics influence student achievement, because high achievers have a better chance of admission than students with lower scores.

To do this, we used the educational production function [Hanushek, 1968]. Hanushek presented a conceptual model of the process of educational production which allows for the establishment of a statistical relationship between the inputs to education and the measure of its output – individual abilities. Usually student abilities are expressed in the individual results of student achievement, which can be reflected, for example, in the scores on a standardized test. Certainly, this proxy can be a noisy indicator of individual abilities, but under certain assumptions it reflects the resources invested in student's human capital. In general, the educational production function has the following form [Hanushek, 1968, p. 14]:

$$A_{it} = F(B_i^{(t)}, S_i^{(t)}, E_i), \text{ where}$$

A_{it} is the vector of educational achievement of a student i at time t ,

$B_i^{(t)}$ is the vector showing the characteristics of student i 's background at time t (including the socio-economic status of student i , her family and classmates),

$S_i^{(t)}$ is the vector of school factors affecting the student i at time t ,

E_i is the vector of innate abilities of the student i .

For almost half a century, the educational production function has undergone many changes, but on the whole, it can be reduced to the following general form [Harris, 2010, p. 403]:

$$A_{it} = f(S_{it}, S_{it-1}, \dots, F_{it}, F_{it-1}, \dots, I_i, \varepsilon_{it}), \text{ where}$$

A_{it} is the vector of educational achievement of a student i at time t ,

S_{it}, S_{it-1}, \dots are vectors of school factors affecting the student i at time t and other previous time periods,

F_{it}, F_{it-1}, \dots are vectors of family and other non-school factors that affect the student i at time t and other previous time periods,

I_i is a fixed student contribution,

ε_{it} is an error.

Factors included in the educational production function are usually divided into two main groups: school and non-school. These factors may give some advantage or limit access to higher education when the applicants are ranked on the basis of their educational achievement. Among the non-school factors are the influence of family income [Hanushek et al., 2014], the number of books at home [Woessmann, 2003], parental education [Woessmann, 2003; Davis-Kean, 2005], pre-entry coaching [Prakhov, 2016b, 2017]. Among the school factors are the so-called teacher factors. Many studies show that student performance is related to the cognitive abilities of teachers [Eide et al., 2004; Hanushek, Rivkin, 2006; Hanushek et al., 2014; Rockoff et al., 2011] but teacher education, gender, and experience have an ambiguous impact on student educational achievement [Hanushek, Rivkin, 2006]. In addition to teacher factors, time spent in the classroom [Jez, Wassmer, 2015], and class size [Cho et al., 2012; Achilles, Finn, 1999; Rivkin et al., 2005] also matter.

2. Data and methodology of study

This paper is based on data obtained from a longitudinal study of students in Moscow schools, started in 2012, when the participants were in the 9th grade and were choosing their further trajectory of learning and their life in general⁵. The data used in this study currently has three waves. During the first wave, more than 5,000 schoolchildren were interviewed. The second wave took place two years after the start of the project, when one part of the respondents was in the final year in high schools and these students faced with a choice of university and further specialization, while another part of survey participants either was receiving a secondary special education, or terminated their studies at the 9th grade of the school (this group was not concerned with university choice and was excluded from the future

⁵ This is a part of a number of panel surveys united in the project ‘Trajectories in Education and Careers’ run by the Center for Cultural Sociology and Anthropology of Education, National Research University Higher School of Economics. See <https://trec.hse.ru/en/> for more details.

analysis). The third wave was carried out a year later, i.e. when former high school graduates had been either admitted to university and became 1st year students, or failed to do so.

We use data on Moscow high school graduates only, i.e. students who study in the region with the largest and most developed regional higher education market in the country. Most of these students were admitted to a university located in the same region (Moscow), where they had studied at school. According to the focus of this study only on those who studied in Moscow, we manage to avoid a number of transaction costs associated with admission to HEIs, e.g. costs of moving, living in another city. In other words, we exclude the regional variation in socio-economic development, the differences in the development of local higher education markets and the corresponding transaction costs which can have a serious impact on the choice of university. All students from the sample have equal opportunities for applying for university (document submission). Consequently, we formally assume equal access to higher education based on the results of USE (i.e., the income factor is eliminated, because students do not have to change the region for the purpose of studying or they do not have to spend additional money for living separately from their parents). The study of the educational choice of Moscow students allows us to study the barriers to higher education in conditions of low transaction costs concerned with university choice and further study.

Data is from those students who graduated from high schools and were admitted to the first year of the university on a state-subsidized place, i.e. we look only at those students who study for free and who competed for this opportunity with their USE scores. Studying on a tuition-free basis is important, because the level of university is based on the average USE score of students enrolled on state-subsidized places, characterized by the competition for them. The sample consists of 826 observations. We added information on the level of the university selectivity (an indicator of university quality); this level is expressed as the average USE score among students who were admitted to a state-subsidized place. This information was obtained from the Monitoring of admission quality – 2015⁶. In addition, we added information on subjects which was taken into account during admission for every student⁷ in order to calculate their average individual USE scores. This information was taken from the directory, which includes information about universities in Moscow and the Moscow region [Kuznetsova, 2016]. Two variables reflecting university quality (the average USE score among admitted students) and individual student abilities (the individual USE average for the

⁶ This is an annual survey of the Russian universities. See <https://ege.hse.ru/> for more information.

⁷ For example, scores in Russian, Mathematics and Physics only were taken into account for some specializations in Physics or Computer science, while for specialization in Humanities one could submit her scores in Russian, History and Foreign language.

individual subjects) were calculated. As a result of the adjustments, the total sample size was 718 observations.

The graphic matching model is shown in Fig. 1. The individual students' scores (x) are on the horizontal axis (the average score for subjects that were taken into account when enrolling in HEI). On the vertical axis is the rating of the university quality. The ideal matching takes place when the individual USE score coincides with the rating of the university. The exact coincidence of the points is on the line of ideal matching, i.e. when $y = x$. In this paper, for an ideal matching certain intervals are constructed (see below), which are also considered to be an ideal fit. The ideal matching area in Fig. 1 is limited by the boundaries of ideal matching ($y = x \pm \alpha\sigma_x$, where σ_x is a standard deviation of the individual USE score, and α is a matching coefficient: for the line of perfect matching $\alpha = 0$; the larger the value of α , the wider the boundaries of perfect matching). Other areas on the graph reflect a mismatch between individual results and the HEI quality; overmatching, when the university quality is significantly higher than the individual USE points, and undermatching, when applicants with sufficiently high USE scores are in HEIs with a low average USE score among admitted students. In the empirical models we assess the factors that lead to the emergence of the latter situation.

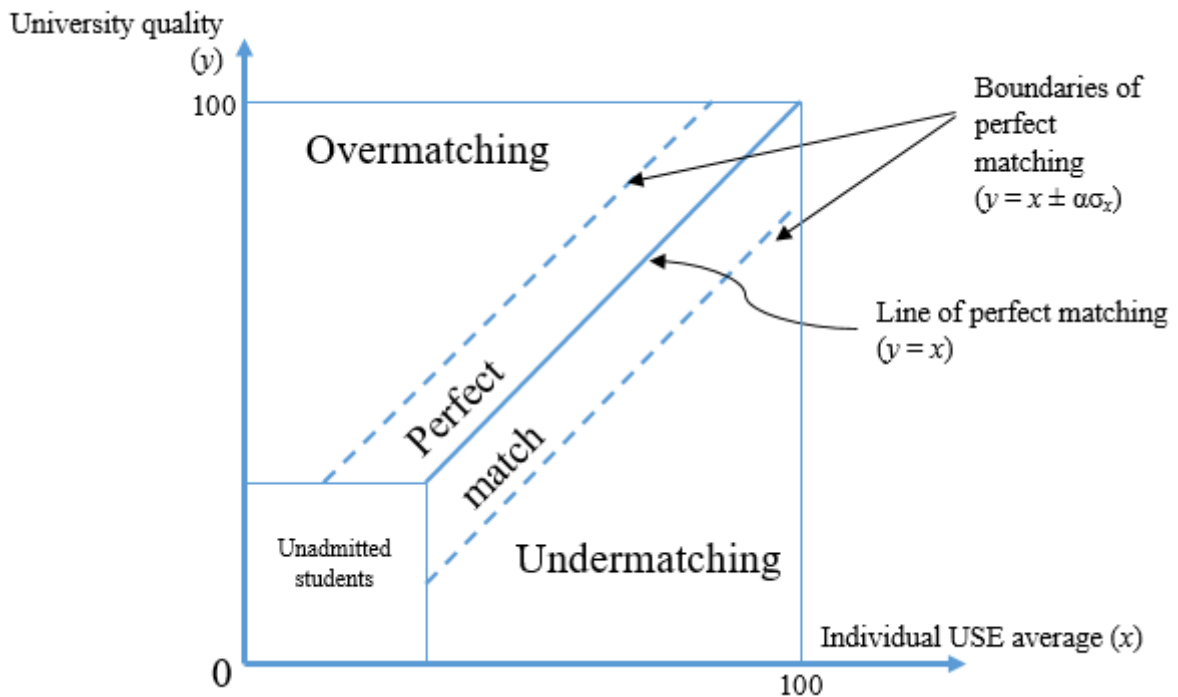


Fig. 1. The graphic model of matching

We assume the neutral role of the university in the process of matching, we examine the matching process from the students' side as students are the main decision-makers in this

process. This corresponds to the nature of USE, in most cases students are admitted only on the basis of their USE scores. In other words, universities cannot introduce their own requirements and manipulate admission (matching) process.

There are a few alternative models of matching. For example, the incline of the boundaries of perfect matching may not be constant (in our case: 45°) and may depend on the score. They can be narrowed with an increase in the individual USE average score. Another variant of matching is student orientation towards last year's admission scores and the adjustment of their university choice (having their USE results) to the past levels of admission. The third explanation of matching is a broader one, if the brightest students belong to top universities and their less able counterparts are enrolled in bottom universities, in general, this can be regarded as a perfect match. In this paper we study the case of simultaneous matching with same-year student scores and university quality with a continuity of scores (x and y).

In order to find out whether the student's choice is a match, in the first stage the difference between the student's individual USE average and the level of university quality was calculated: $\Delta = x - y$. Using Δ , we can formally determine mismatching. Thus if $\Delta \ll 0$, then the personal average score of the USE applicant is much lower than the quality of admission to the university, which corresponds to overmatch. Conversely, when $\Delta \gg 0$, the applicant's individual USE score exceeds the average score for the university where she was enrolled, which is undermatching. Descriptive statistics on the variable Δ are presented in Table 1.

Table 1. The difference between the individual average USE score of the student and the average USE score among admitted students. Descriptive statistics

Variable	Min.	Max.	Mean	Std.Dev.	Obs.
Δ	-27.8	36.4	-0.63	8.47	718

Table 1 shows the mean value of the variable Δ which is close to 0 and corresponds to a perfect match. For a number of entrants there are very significant discrepancies between individual USE results and the quality of the university they entered.

At the next stage, binary variables which estimate the probability of undermatching were created ($\Pr(\Delta > \alpha\sigma_x)$). In the regression probit models, they are considered as dependent variables. Variables take the value of 1 if $\Delta > \alpha\sigma_x$, and 0 otherwise. The case where $\Pr(\Delta > \alpha\sigma_x) = 1$ corresponds to undermatching. In the regression models we use the coefficients of matching (α) which equal 0.5, 0.75 and 1 in different specifications. Descriptive statistics for dependent variables are presented in Table 2.

Table 2. Dependent variables. Descriptive statistics

Variable	Value	Obs.	Frequency
$\Pr(\Delta > 0.5\sigma_x)$	1	200	27.9%
	0	518	72.1%
$\Pr(\Delta > 0.75\sigma_x)$	1	134	18.7%
	0	584	81.3%
$\Pr(\Delta > \sigma_x)$	1	98	13.6%
	0	620	86.4%

Since the standard deviation of x is approximately 8.47, the boundaries of perfect matching are given as follows: for $\Delta > 0.5\sigma_x$: [-4.235; 4.235]; for $\Delta > 0.75\sigma_x$: [-6.353; 6.353]; for $\Delta > \sigma_x$: [-8.47; 8.47]. When establishing the narrowest border (for $\alpha = 0.5$), 200 entrants were undermatched, which is 27.9% of the sample. In the ‘wider’ borders, at $\alpha = 0.75$, were 134 respondents (about 18.7%) and for the ‘widest’ borders of matching ($\alpha = 1$), there were 98 entrants (13.6% of the sample). Thus, on the whole, favorable relations were established between individual USE results and the quality of admission to HEIs, as the majority of students are perfectly match or overmatched.

Note that individual USE scores represent a function and are associated with the student’s innate abilities and are influenced by a set of external characteristics (family and school factors). That is why, in the second part of the empirical analysis, an assessment of the educational production function is conducted. Individual average USE scores are used as the dependent variable. Descriptive statistics are presented in Table 3.

Table 3. The individual USE average score for the subjects taken into account in admission. Descriptive statistics

Variable	Number of obs. (Frequencies)			Min.	Max.	Mean	Std.Dev.
	41-60	61-80	81-100				
x	94 (13.1%)	370 (51.5%)	254 (35.4%)	42.33	97.33	74.79	11.70

The regression models can be presented in the following form:

$$\Pr(\Delta > \alpha\sigma_x) = f(F'\beta) \quad (1)$$

$$x = F'\gamma \quad (2), \text{ where}$$

F is a vector of independent variables,

$f(\cdot)$ is a probit distribution function,

β, γ are vectors of regression coefficients.

Independent variables (F) are gender, high school status, high school ranking, a student's change of school, class specialization, the birthplace of the student, the level of the mother's education, family status (single-parent family), the number of books at home. In addition, in the probit models (1), the individual USE average score is used as an independent variable. Descriptive statistics of independent variables are presented in Table 4.

Table 4. Independent variables. Descriptive statistics

Variable	Min.	Max.	Mean	Std.Dev.
Gender (=1, if male)	0	1	0.46	0.50
Individual USE average (out of 100)	42.33	97.33	74.79	11.70
School status (=1, if a special status)	0	1	0.54	0.50
School ranking (position)	1	350	208.5	125.45
School change (=1, if yes)	0	1	0.1	0.30
Class specialization (=1, if any)	0	1	0.8	0.40
Place of birth (=1, if Moscow)	0	1	0.7	0.46
Mother's education (=1, if higher education)	0	1	0.77	0.42
Single-parent family (=1, if yes)	0	1	0.91	0.28
Number of books at home	5	650	291.7	220.25

The explanatory variables include student individual, school and family characteristics. Individual factors are gender, individual USE result and a place of birth. School features are high school status, its ranking, a student's change of school, and class specialization. Family factors include mother's education, single-parent family, and the number of books at home.

Next we briefly describe the main characteristics of the independent variables. The share of female students among respondents who are in the final sample of 718 observations is 54%. Among the study sample, about 70% of students were born in Moscow. Individual USE scores range from 42 to 97 with a mean value of 75.

The status of the high school is a binary variable that takes the value 1 if the student has not studied in an ordinary school and 0 otherwise. Around 46% of the respondents were enrolled in high schools without any status. The variable 'school rating' reflects the school position in the Ranking of Moscow schools in the 2014/2015 academic year⁸. In this ranking, the top 300 schools in Moscow are ranked according to the results of their educational activity. The higher the position, the 'better' the school. For all schools which are not included in this ranking, a position of 350 is assigned.

⁸ This ranking is official and formed by the Moscow Department of Education. See: <https://www.mos.ru/dogm/function/ratings-vklada-school/rating-2014-2015/>.

The regressions also reflect whether the student changed school after the 9th grade. About 10% of students changed their secondary schools. The class specialization is given by a dummy variable which is 1 if there is a certain specialization (for example, with extra classes in mathematics or languages). 80% studied in classes with a specialization.

As one of the family characteristics, the level of mother's (stepmother's) education is considered. In 77% of families, the mother (stepmother) has a degree. Another family characteristic is the family status. In the sample, 9% students have a single-parent family with no one substituting one of the biological parents (a stepmother or a stepfather). The number of books at home can influence the educational achievement and is considered as an independent variable. This variable takes values from 5 to 650. On average, a respondent has 292 books at home.

Correlation coefficients were calculated for all independent variables. The results are presented in the correlation matrix (see Appendix 1). Only weak (up to 0.3) correlations are observed, there is no strong correlation between any factors.

3. The results of regression analysis

The results of the regression analysis of the probit models (1) are presented in Table 5. For the convenience of interpreting the results of the probit regression, average marginal effects are reported⁹.

Table 5. The results of probit-regressions assessing the likelihood of an undermatch situation

Independent variables	Dependent variables					
	Pr($\Delta > 0.5\sigma_x$)		Pr($\Delta > 0.75\sigma_x$)		Pr($\Delta > \sigma_x$)	
	1	2	3	4	5	6
Gender	0.017 (0.041)	—	0.044 (0.034)	0.062** (0.028)	0.048** (0.028)	0.048** (0.023)
Individual USE average	0.018*** (0.002)	0.017*** (0.002)	0.013*** (0.002)	0.013*** (0.001)	0.010*** (0.001)	0.010*** (0.001)
School status	-0.007 (0.042)	—	0.005 (0.034)	—	0.012 (0.028)	—
School ranking	0.001*** (0.000)	0.000*** (0.000)	0.000** (0.000)	0.000*** (0.000)	0.000*** (0.000)	0.000** (0.000)

⁹ In addition to the correlation matrix presented in Appendix 1, the variance inflation factor (VIF) was calculated to exclude the possible multicollinearity of independent variables. For all explanatory factors in the regressions presented, the VIF is approximately equal to one. Thus, this proves the absence of a linear interdependence between the variables.

School change	0.055 (0.075)	—	0.052 (0.064)	—	0.074 (0.059)	—
Class specialization	0.012 (0.049)	—	-0.006 (0.041)	—	-0.011 (0.035)	—
Place of birth	0.040 (0.041)	—	0.047 (0.032)	—	0.028 (0.026)	—
Mother's education	0.077* (0.045)	0.073* (0.040)	0.017 (0.039)	—	0.022 (0.031)	—
Single-parent family	0.046 (0.068)	—	0.001 (0.059)	—	0.046 (0.036)	—
Number of books at home	0.000 (0.000)	—	0.000 (0.000)	—	0.000 (0.000)	—
Observations	518	718	518	718	518	718
<i>McFadden R</i> ²	0.400	0.244	0.379	0.141	0.379	0.145
Standard errors in parentheses. Levels of significance: *** – 1%-level, ** – 5%-level, * – 10%-level.						

The first important finding is that the individual USE mean score is statistically significant in determining the probability of a mismatch in all model specifications, regardless of the width of the boundaries of perfect matching. Higher USE results increase the probability of being in a weaker institution. One of the possible interpretations of this is the assumption that the higher the results of the USE, the more opportunities the applicant has to choose. Accordingly, this choice can be more conscious, i.e. an entrant with high USE scores interested in a certain specialization may not find a university of the appropriate (corresponding) quality. On the other hand, with higher USE scores, the value of the variable Δ can also increase, which will lead to this entrant being undermatched.

There is a statistically significant relationship between the school ranking and the probability of undermatching in all model specifications. The lower the ranking of the school, the higher the probability of undermatching. Since the school ranking can be considered as a proxy for the quality of high school education, it can be concluded that the higher the quality of high secondary education, the lower the probability of undermatching.

Third, some individual and family characteristics matter. Table 5 shows that for the widest limits of the perfect matching (columns 4-6), male students are more likely to be undermatched than female students. Mother's education is significant under narrow matching (columns 1-2). With wider borders of perfect matching, this variable becomes insignificant.

The remaining variables are insignificant in all the model specifications regardless of the matching coefficient¹⁰.

As the school ranking has little marginal effect on the probability of being undermatched, and as a significant share of students graduated from high schools that were not included in the Top-300 rating (264 out of 718, or 37.8%), it was suggested that the variance between schools included in the rating is low. However, the effect of being in a non-ranked high school might be larger. To test this hypothesis, additional alternative regression models were estimated: instead of high school ranking, a dummy variable reflecting graduation from the high school from Top-300 was added. This new variable equals 1 if the school is ranked among the best, and 0 otherwise. The results of the regression analysis (average marginal effects) for the new variable are presented in Table 6. More detailed results of the regression analysis can be found in Appendix 2.

Table 6. A summary of additional probit regressions

Independent variable	Dependent variables		
	$\Pr(\Delta > 0.5\sigma_x)$	$\Pr(\Delta > 0.75\sigma_x)$	$\Pr(\Delta > \sigma_x)$
High school from Top-300	-0.104** (0.045)	-0.064* (0.037)	-0.078** (0.017)
Observations	518	518	518
<i>McFadden R</i> ²	0.395	0.374	0.377
Standard errors in parentheses. Levels of significance: *** – 1%-level, ** – 5%-level, * – 10%-level.			

Graduation from one of the best schools is sufficient and significantly influences the probability of choosing a HEI which is better suited to the abilities of the applicant. Schooling of this kind reduces the probability of undermatching by 6-10%.

USE generally copes with its task of providing a fair mechanism for admission to higher education, since even when establishing narrow matching borders, a relatively small proportion of students enter weaker institutions from the selectivity point of view. However, it is worth noting the influence of a high school quality on the matching outcomes. In the context of the low transaction costs associated with admission, with the homogeneity of regional socio-economic development and having all students from the same regional education market, the variation in the quality of schooling can also influence undermatching.

¹⁰ Besides, the additional models with the inclusion of the family income status were tested. As expected, income variables turned out to be insignificant, since this study examines students from the same region, and the variation in the family income in the sample is quite low. That is why the financial situation of the family finally was excluded from all regression models.

As noted, in all specifications of model (1) the individual USE average score is significant. Since USE results are not random, but are a function of school, family and other characteristics, it is also necessary to identify which factors are the determinants of student performance. Such an analysis will reveal how the USE result is influenced by factors not directly related to student abilities. It will be possible to draw conclusions about the limitations of the chances of obtaining high USE scores and the barriers of access to higher education, even in conditions of low transaction costs. In addition, when assessing the production function in education, for the first time in Russia a ‘strategic’ factor is used as the dependent variable. In other words, this variable reflects the results of only those subjects that are required for enrollment, i.e. subjects in which the student is most interested in the results. The results of regression analysis presented in Table 7 most accurately reflect the contribution of various resources to student performance¹¹.

Table 7. The determinants of student academic performance of schoolchildren.
Dependent variable: the individual USE average score (x)

Independent variables	1	2
Constant	72.857*** (2.557)	68.590*** (2.304)
Gender	-6.352*** (0.960)	-6.354*** (0.964)
School status	5.571*** (0.988)	5.903*** (0.978)
School ranking	-0.011*** (0.004)	—
High school from Top-300	—	2.005* (1.029)
School change	0.349 (1.721)	0.571 (1.729)
Class specialization	2.189* (1.180)	2.245* (1.184)
Place of birth	1.053 (1.034)	1.102 (1.038)

¹¹ In the previous Russian studies on the student performance, the educational production function was assessed by a regression of USE points earned in the Russian language, Mathematics (compulsory subjects), and the average score for all subjects taken on a number of characteristics. At the same time, a number of subjects that the student sit could not be taken into account by higher education institutions when enrolling: for example, Mathematics, which is a mandatory USE subject for being graduated from the high school, is usually not required when enrolling in a Humanities specialization. Thus, such estimates of the educational production function in education could be biased, since they did not reflect the strategic nature of the dependent variable used.

Mother's education	3.012** (1.173)	3.102*** (1.182)
Single-parent family	-1.990 (1.813)	-1.607 (1.813)
Number of books at home	0.004* (0.002)	0.004* (0.002)
Observations	518	518
R ²	0.196	0.190
Standard errors in parentheses. Levels of significance: *** – 1%-level, ** – 5%-level, * – 10%-level.		

Individual factors (gender), school characteristics and family factors determine educational outcomes measured by the student's USE score. Female students have USE scores 6.4 points higher than male students.

The magnitude of the previous factor is comparable to the coefficient for the type of school. Those applicants who studied in special schools (such as in schools with in-depth study of certain subjects) receive an average of 6 points more per subject compared to their peers from ordinary secondary schools. Among the school factors, in addition to the status of the school, the class specialization is statistically significant, which brings students slightly more than 2 USE points per subject. Students who graduated from high schools from the Top-300 gain 2 USE points. The exact rank of the school in the ranking also affects USE results. The difference in USE results for the student from the 1st school in the ranking and the 300th school can exceed 3 points.

Among family factors, those students whose mother has higher education, pass USE exam with 3 more points. The number of books at home also influences the USE results. Similar conclusions were obtained in the works of Woessmann [2003] and Hanushek et al. [2014]. School and non-school factors have approximately equal contribution to the final USE results. Such factors such as place of birth, school change and family structure (single-parent family) do not have a significant impact on the final USE results.

Thus, despite the relatively small proportion of applicants who are undermatched, the USE results which are the basis of matching, are influenced by a number of family and school characteristics. In other words, in conditions of low transaction costs, USE as an institution copes with the problem of the accessibility of higher education in terms of ensuring matching between the quality of the applicant and the university. On the other hand, the situation of unequal access to higher education may arise much earlier than the enrollment stage due to the presence of statistically significant between the effectiveness of the applicant and the characteristics of the family and school, which raises questions about the accessibility of higher education even within the same region.

4. Conclusion

This paper examined the mechanism of matching the quality of the applicant and the university using USE. On the basis of this mechanism, the possibility for unequal access to higher education was explored. The results show that during the admission process, favorable conditions for ideal matching or overmatching, when the level of HEIs far exceeds the personal educational achievements of students, are observed. Thus, depending on the coefficient of matching, in the least favorable situation (undermatching), when applicants find themselves in universities with a lower level than the personal abilities of the student, is relevant for 14 to 28% of entrants.

The probability of undermatching is affected by individual USE scores and the ranking of the high school. Nevertheless, it was empirically shown that USE results themselves are influenced by a number of school characteristics (school type, class specialization, school quality) and family characteristics (mother's education, number of books at home). These results indicate the presence of factors not directly related to student's innate abilities, which are important determinants of academic performance. From the point of view of ensuring equal access to higher education, it can be concluded that, although most entrants find themselves in HEIs corresponding to their abilities, the inequality of access to higher education is laid at earlier stages – in the family and in the school.

Note that these results were obtained under assumption of low transaction costs associated with admission to university, since the empirical part of the study was based on data on students who graduated from Moscow high schools and entered universities in the Moscow region. Thus, a number of barriers related to the family income status, regional economic development and the limited choice of the university in the regional higher education markets were initially leveled. Since even in conditions close to ideal, the inequality of access to higher education arises at pre-entry stages, the inclusion of students from different regions (and when the corresponding regional variation is included) can lead to even greater effects with an increase of the share of mismatched students. Therefore, conducting an analysis of matching on the Russian higher education market in general would be useful.

The results make it possible to identify problems associated with the implementation of the existing mechanism for access to higher education, allow for the development of policy recommendations (including managerial decisions in education) and for making decisions by individual households, and to identify areas for further possible research.

First, in all models, there was a significant effect of the quality of high school education both on USE results, and on the probability of undermatching. Thus, it is worth paying

attention to the implementation of policies to improve the quality of school education and which are oriented towards a possible smoothing of inequalities between graduates from schools with special status and those from ordinary schools. We do not mean ‘averaging’ the quality of teaching in Russian schools, we are talking about the need to ‘pull up’ backward educational institutions to the middle level. For example, it is possible to conduct additional work to improve the skills of school teachers.

Secondly, despite the fact that we did not reveal a significant contribution of family factors to the probability of undermatching (with the exception of mother’s education in a number of models), family factors are significant in determining USE results. Therefore, it is worth paying attention to family investments in human capital. It has been shown that in the presence of mother’s higher education and with a significant number of books at home, students show higher results in terms of USE. Thus, the level of the cultural and social capital of the family is related to the academic performance. Therefore, programs should be developed to support families with insufficient social and cultural capital to increase their awareness. It is also possible to organize support in the form of additional school activities with children from such families.

Among other things, the influence of the gender on the results of USE and the likelihood of undermatching was revealed. This conclusion should be taken cautiously, as according to some studies this difference between male and female students is primarily attributed to the difference in their preferences and tastes [Zafar, 2013]. The determinants of such gender differences would also be an interesting topic for possible future research.

References

1. Anikina E.A., Lazarchuk E.V., & Chechina V.I. (2014). Accessibility of higher education as a socio-economic category (in Russian). *Fundamental research*. 12-2, 355-358.
2. Bowen W., & Bok D. (1998). *The shape of the river: Long-term consequences of considering race in college and university admissions*. Princeton: Princeton University Press.
3. Burdyak A. Y. (2015). Additional school subjects’ lessons: motivation and popularity (in Russian). *The Monitoring of Public Opinion: Economic and Social Changes Journal*. 2(125). 96-112.
4. Cho, H., Glewwe, P., & Whitler, M. (2012). Do reductions in class size raise students’ test scores? Evidence from population variation in Minnesota's elementary schools. *Economics of Education Review*, 31(3), 77-95.
5. Cooper, R., & Liu, H. (2016). *Money or Grit? Determinants of MisMatch by Race and Gender* (No. w22734). National Bureau of Economic Research.

6. Davis-Kean, P. E. (2005). The influence of parent education and family income on child achievement: the indirect role of parental expectations and the home environment. *Journal of family psychology, 19*(2), 294.
7. Dillon, E. W., & Smith, J. A. (2017). Determinants of the match between student ability and college quality. *Journal of Labor Economics, 35*(1), 45-66.
8. Eide, E., Goldhaber, D., & Brewer, D. (2004). The teacher labour market and teacher quality. *Oxford Review of Economic Policy, 20*(2), 230-244.
9. Finn, J. D., & Achilles, C. M. (1999). Tennessee's class size study: Findings, implications, misconceptions. *Educational evaluation and policy analysis, 21*(2), 97-109.
10. Fréchette, G. R., Roth, A. E., & Ünver, M. U. (2007). Unraveling yields inefficient matchings: evidence from post-season college football bowls. *The RAND Journal of Economics, 38*(4), 967-982.
11. Gale, D., & Shapley, L. S. (1962). College admissions and the stability of marriage. *The American Mathematical Monthly, 69*(1), 9-15.
12. Griffith, A. L., & Rothstein, D. S. (2009). Can't get there from here: The decision to apply to a selective college. *Economics of Education Review, 28*(5), 620-628.
13. Hanushek, E. A. (1968). *The education of negroes and whites* (Doctoral dissertation, Massachusetts Institute of Technology).
14. Hanushek, E. A., Piopiunik, M., & Wiederhold, S. (2014). *The value of smarter teachers: International evidence on teacher cognitive skills and student performance* (No. w20727). National Bureau of Economic Research.
15. Hanushek, E. A., & Rivkin, S. G. (2006). Teacher quality. *Handbook of the Economics of Education, 2*, 1051-1078.
16. Harris, D. N. (2010). Education production functions: Concepts. In B. McGaw, P. L. Peterson, & E. Baker (Eds.), *International encyclopedia of education* (pp. 402–406). Amsterdam: Elsevier.
17. Haruvy, E., Roth, A. E., & Ünver, M. U. (2006). The dynamics of law clerk matching: An experimental and computational investigation of proposals for reform of the market. *Journal of Economic dynamics and control, 30*(3), 457-486.
18. Hoxby, C., & Turner, S. (2013). Expanding college opportunities for high-achieving, low income students. *Stanford Institute for Economic Policy Research Discussion Paper, (12-014)*.
19. Hoxby, C., & Avery, C. (2013). The missing 'one-offs': The hidden supply of high-achieving, low-income students. *Brookings papers on economic activity, 2013*(1), 1-65.

20. Jez, S. J., & Wassmer, R. W. (2015). The impact of learning time on academic achievement. *Education and Urban Society*, 47(3), 284-306.
21. Kuznetsova I., & Shilova O. (2016). Moscow and Moscow District Universities 2016-2017 (in Russian). Moscow: Eksmo.
22. Light, A., & Strayer, W. (2000). Determinants of college completion: School quality or student ability?. *Journal of Human Resources*, 299-332.
23. Lincove, J. A., & Cortes, K. E. (2016). *Match or Mismatch? Automatic Admissions and College Preferences of Low-and High-Income Students* (No. w22559). National Bureau of Economic Research.
24. Pitoukhin E. A., & Semenov A. A. (2011). Analysis of inter-regional mobility of school-leavers entering to the universities (in Russian). *University Management: Practice and Analysis*. 3. 82-89.
25. Prakhov, I. (2016). The Barriers of Access to Selective Universities in Russia. *Higher Education Quarterly*, 70(2), 170-199.
26. Prakhov, I. (2016). The Unified State Examination and the Determinants of Academic Achievement: Does Investment in Pre-entry Coaching Matter?. *Urban Education*, 51(5), 556-583.
27. Prakhov, I., & Bocharova, M. (2016). *Socio-Economic Predictors of Student Mobility* (No. WP BRP 34/EDU/2016). National Research University Higher School of Economics.
28. Prakhov I. A., & Yudkevich M. M. (2017). University admission in Russia: Do the wealthier benefit from standardized exams?. *International Journal of Educational Development*.
29. Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (2005). Teachers, schools, and academic achievement. *Econometrica*, 73(2), 417-458.
30. Rockoff, J. E., Jacob, B. A., Kane, T. J., & Staiger, D. O. (2011). Can you recognize an effective teacher when you recruit one?. *Education*, 6(1), 43-74.
31. Roth, A. E. (1984). The evolution of the labor market for medical interns and residents: a case study in game theory. *Journal of political Economy*, 92(6), 991-1016.
32. Roth, A. E., & Peranson, E. (1999). The redesign of the matching market for American physicians: Some engineering aspects of economic design. *The American Economic Review*, 89, 748-780.
33. Roth, A. E. (1991). A natural experiment in the organization of entry-level labor markets: regional markets for new physicians and surgeons in the United Kingdom. *The American economic review*, 415-440.
34. Slonimczyk, F., Francesconi, M., & Yurko, A. V. (2017). Moving On Up for High School Graduates in Russia: The Consequences of the Unified State Exam Reform.

35. Smith, J., Pender, M., & Howell, J. (2013). The full extent of student-college academic undermatch. *Economics of Education Review*, 32, 247-261.
36. Uvarov A. G., & Yastrebova G. A. (2014). Schools and socioeconomic standing of Russian families as competing factors in promoting social inequality in Russia (in Russian). *Universe of Russia*. 2. 103-129.
37. Wößmann, L. (2003). Schooling resources, educational institutions and student performance: the international evidence. *Oxford bulletin of economics and statistics*, 65(2), 117-170.
38. Zafar, B. (2013). College major choice and the gender gap. *Journal of Human Resources*, 48(3), 545-595.

Appendices

Appendix 1. Correlation matrix

Independent variables		1	2	3	4	5	6	7	8	9	10
Gender	1	1.000									
Individual USE average	2	-0.267	1.000								
School status	3	0.016	0.279	1.000							
School ranking	4	0.019	-0.201	-0.260	1.000						
School change	5	-0.005	0.032	0.067	-0.031	1.000					
Class specialization	6	0.055	0.091	0.080	-0.043	0.031	1.000				
Place of birth	7	-0.038	0.074	0.050	-0.038	-0.073	0.067	1.000			
Mother's education	8	0.052	0.140	0.076	-0.166	0.029	0.058	-0.029	1.000		
Single-parent family	9	0.023	-0.007	0.060	-0.117	-0.020	0.018	0.000	0.126	1.000	
Number of books at home	10	-0.118	0.139	0.055	0.002	-0.003	-0.017	0.077	0.160	0.009	1.000

Appendix 2. The results of additional probit regressions

Independent variables	Dependent variables		
	$\Pr(\Delta > 0.5\sigma_x)$	$\Pr(\Delta > 0.75\sigma_x)$	$\Pr(\Delta > \sigma_x)$
Constant	-1.579*** (0.168)	-1.176*** (0.135)	-1.003*** (0.116)
Пол	0.015 (0.041)	0.044 (0.034)	0.047* (0.028)
Individual USE average	0.018*** (0.002)	0.012*** (0.002)	0.010*** (0.001)
School status	-0.021 (0.042)	-0.005 (0.034)	0.007 (0.028)
High school from Top-300	-0.104** (0.045)	-0.064* (0.037)	-0.078** (0.032)
School change	0.042 (0.074)	0.043 (0.063)	0.063 (0.057)
Class specialization	0.009 (0.050)	-0.009 (0.041)	-0.014 (0.035)
Place of birth	0.040 (0.041)	0.047 (0.033)	0.030 (0.026)
Mother's education	0.076* (0.045)	0.017 (0.039)	0.025 (0.030)
Single-parent family	0.024 (0.071)	-0.017 (0.063)	0.037 (0.040)
Number of books at home	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
Observations	518	518	518
<i>McFadden R</i> ²	0.196	0.190	0.196
Standard errors in parentheses. Levels of significance: *** – 1%-level, ** – 5%-level, * – 10%-level.			

Ilya Prakhov

Ph.D., Research Fellow at Center for Institutional Studies, National Research University
Higher School of Economics, Moscow, Russia.

Email: ipra@inbox.ru

Denis Sergienko

Research Assistant at Center for Institutional Studies, National Research University Higher
School of Economics, Moscow, Russia.

Email: denis-sergienko@ro.ru

**Any opinions or claims contained in this Working Paper do not necessarily
reflect the views of HSE.**

© Prakhov, 2017

© Sergienko, 2017