NATIONAL RESEARCH UNIVERSITY
HIGHER SCHOOL OF ECONOMICS

Andrey Zakharov, Martin Carnoy, Prashant Loyalka

WHICH TEACHING PRACTICES IMPROVE STUDENT PERFORMANCE ON HIGH-STAKES EXAMS? EVIDENCE FROM RUSSIA

BASIC RESEARCH PROGRAM

WORKING PAPERS

SERIES: EDUCATION
WP BRP 13/EDU/2013

Andrey Zakharov[1], *Martin Carnoy*[2], *Prashant Loyalka*[3]

WHICH TEACHING PRACTICES IMPROVE STUDENT PERFORMANCE ON HIGH-STAKES EXAMS? EVIDENCE FROM RUSSIA[4]

This study examines the relationship between teaching practices aimed at raising student performance on a high stakes college entrance examination—the Russian Unified State Exam (USE) — and student performance on that test. The study uses data from a school/classroom survey of almost 3,000 students conducted in 2010 in three Russian regions. The analysis employs a student fixed effects method that estimates the impact of teaching practices used by students' mathematics and Russian language teachers on students' exam results. To test for possible heterogeneous effects of practices in different academic tracks, the study estimates the practices' effect on USE scores for students in advanced and basic level tracks. The study finds that the only strategy with positive effects on test outcomes is greater amounts of subject-specific homework geared to different types of test items, and that the most effective type of homework differs across tracks.

JEL classification: I21 (Analysis of Education).
Keywords: teaching practices, curriculum, student achievement, selection bias, student fixed effect, high-stakes examinations

[1] National Research University Higher School of Economics. International Laboratory for Educational Policy Research. Deputy Head. Email: ab.zakharov@gmail.com
[2] Stanford University. Vida Jacks Professor of Education. Email: carnoy@stanford.edu
[3] Stanford University. Freeman Spogli Institute for International Studies. Center Research Fellow. Email: loyalka@stanford.edu
[4] The authors wish to thank the National Research University Higher School of Economics and the U.S. Department of Education, Fund for the Improvement of Secondary Education, for supporting this research.

1. Introduction

Although a generation of research on "educational production" has greatly increased our knowledge of what works to improve student learning, there is little research on which teaching practices impact high school students who face high-stakes exit/entrance examinations. In most countries, including China, India, Russia, Germany, Brazil, and the United States, high school students must take entrance exams to qualify for college and particularly elite colleges (Carnoy et al., 2013). In other countries, such as France, Spain, and Italy, high school students are required to take high school exit exams to qualify for a degree.

Evidence suggests that, when preparing students for such high-stakes examinations, teachers use certain types of practices more than others to increase student achievement. Bishop (1996, 1997) showed that teaching practices in Canadian provinces with curriculum-based high school exit examinations were more likely to focus on more complex learning skills. Teachers also assign more homework related to the exam and give more practice exams compared to provinces without such examinations. Whether and which of these teaching practices in fact help students to improve their performance in a high-stakes environment, however, has not yet been rigorously tested in the empirical literature.

Given this lack of evidence, the goal of our study is to examine which teaching practices improve the performance of high school students on a high-stakes examination. We use a unique data set from a survey from three regions of Russia of almost 3,000 final-year (11^{th} grade) high school students who were preparing for the national college entrance exam (the Unified State Exam or USE) in 2010. We use the data and a cross-subject student fixed effects model (Clotfelter, Ladd, and Vigdor, 2010) to estimate the effect of three specific teaching practices on student examination outcomes: (a) the proportion of homework exercises targeting specific entrance exam items (hereafter known as "test-specific homework exercises"); (b) teachers' use of practice (or mock) tests; and (c) teachers' use of websites geared to help students prepare for the exam.

We find that of the three practices only "test-specific homework exercises" has a positive and significant effect on student performance. The effect is rather large—about 0.2 of a standard deviation (SD) in exam score. Further, we find that the effectiveness of test-specific homework exercises is greater for students in the advanced track when homework exercises are focused on more difficult test items. Similarly, the effectiveness of test-specific homework exercises is greater for students in the basic track when homework exercises are focused on easier test items. The results suggest that we can identify those teaching practices that improve high school student performance on high-stakes exams.

The structure of the rest of the paper is as follows: Section 2 provides background on research that is relevant to this study and on the Russian education system, in particular the USE examination. Section 3 describes the data. Section 4 discusses the estimation strategy. Section 5 presents the results. Section 6 discusses the results and concludes.

2. Background

Research on Teacher Impacts

Recent discussion on the effectiveness of school inputs in raising student outcomes focuses on teachers, showing that students with more effective teachers perform better on achievement tests (for example, Sanders and Rivers, 1996; Rockoff, 2004; Hanushek, Rivkin and Kain, 2005; Nye et al., 2004; Boyd et al., 2006). However, much of the emphasis in identifying effective teachers has been on teacher characteristics associated with higher student outcomes rather than on teaching practices. Such teacher characteristics are important for our study because they help identify the "quality" of teachers that should be controlled for in estimating the effect of teacher practice on student examination performance. For example, some studies suggest that greater teacher experience contributes significantly to student achievement (Ferguson and Ladd, 1996; Clotfelter et al., 2007; Rockoff, 2004; Hanushek, Rivkin and Kain, 2005). Other studies suggest that positive effects on student outcomes result from the quality of teachers' pre-service education (Clotfelter, et. al., 2007; Darling-Hammond, 2009; Goldhaber and Brewer 1997, 2000; Kukla-Acevedo, 2009; Monk, 1994) and teacher certification (Boyd et al, 2006; Clotfelter et al., 2007).

Alongside the emphasis on teacher characteristics, there has also been a history of trying to link teaching practices to student achievement (for a review of research in the United States, see Hill et al., 2005). More recently, researchers have used large-scale samples to measure the link between teaching practices and student test scores gains. In the United States, the effort has culminated in an extensive study of teacher effectiveness—a sample of 3,000 primary and middle school teachers in urban areas. In particular, the study collects information on teaching practices through videotaped observations and student survey responses. The aim of the study, however, was to create a composite indicator of an effective teacher rather than to estimate the impacts of specific teaching practices (Kane et al, 2013). In Botswana and South Africa, Carnoy et al. (2012) show that observational ratings of sixth-grade mathematics teaching quality have significant and large effects on mathematics achievement. Schwerdt and Wuppermann (2011) find that lecture-style teaching in U.S. schools significantly increased students' achievement. In contrast, von Klaveren's (2011) estimates using Dutch TIMSS data show that time spent lecturing in front of the class has no significant effect on student outcomes. Overall, while all of the above research on teacher practices informs our study, none of the research refers to high school students or to high-stakes examinations.

The Russian Context

Since 2009, Russia has required that all grade 11 (the final year of academic schooling) students take a national exit examination that also functions as a college entrance examination (the USE). In fact, most students who complete general (academic) high school sit for the USE (98 percent of students in 2009).[5] The scores on the USE (along with whether students won an award at a recognized academic competition and students' college choices) are the primary criteria used to match students into different colleges and majors. Because of the extremely high proportion of students taking the USE and the central role of the USE in college admissions, the exam is high-stakes.

Along with the USE being a high-stakes exam for students, it is high-stakes for teachers and principals. Teacher performance is assessed, in part, according to their students' USE scores. USE scores are also an important criterion that determines principal bonuses.[6] Furthermore, the reputation of schools is affected as they are ranked according to students' USE results.[7] Teachers and school principals therefore have strong incentives to use teaching practices that will maximize student performance on the USE.

The USE has two main features other than its high stakes that are important for our analysis of the impact of teaching practices on student performance. First, each student takes two mandatory subject-specific exams, Russian language and mathematics, as well as three subject-specific exams of his/her own choosing. As explained in the section on empirical strategy below, we use within-student variation across the mandatory subject-specific exams to help identify the causal impacts of teaching practices on student USE performance.

---

[5] 98.2 percent of students sat the Russian language USE and 98.1 percent sat the math USE. Estimated from http://www.ed.gov.ru/files/materials/11987/76rik.pdf

[6] See http://old.mon.gov.ru/files/materials/6772/model-nsot.pdf

[7] For example see http://ria.ru/sn_edu/20130423/930945392.html

Second, the subject-specific exams contain items of varying difficulty. The mathematics test includes two types of items: B and C. The B-type items are short-answer questions that require some basic analysis. The C-type items are also open-ended but of a higher level of complexity. They require students to give detailed answers and show their work.

The Russian language test also includes B and C-types of items. Both of them are of a high level of difficulty. The B-type items are short answer questions that evaluate students' linguistic competence. The C-type items engage the students in writing compositions and are supposed to reflect students' ability to communicate effectively.

3. Data

In May 2010, we conducted a survey of 2,927 students in 182 classes in 127 schools in three regions of Russia: Pskovskaya and Yaroslavskaya *oblasts* and Krasnoyarsky *krai.* The three regions were chosen because they represent a diversity of economic and educational contexts in a large and heterogeneous country. Krasnoyarsky *krai* in Siberia is Russia's second largest region (13 percent of the national territory) and is one of the richest in natural resources and industrial production. Yaroslavskaya *oblast* is located in the central part of the country, north of Moscow. Pskovskaya *oblast* is in the northwest of Russia and borders Estonia, Latvia, and Belarus. Both *oblasts* are small and, compared to Krasnoyarsky *krai,* are less developed economically.

Table 1. Descriptive Statistics for Russian Regions Participating in School Sample, 2010

Variable	**Krasnoyarsky krai**	**Pskovskaya oblast**	**Yaroslavskaya oblast**
Population (thousands)	2,829	671	1,271
Population with higher education (per 1000)	207	187	218
GDP per capita (thousand rubles)	371.0	124.7	183.6
Number of schools (fall 2009)	1124	269	470
Secondary school students, fall 2009 (thousands)	287	60	107
Number of students finished secondary and took USE in 2010	19,630	3,902	5,881
Russian language USE mean score (0-100)*	57.9	59.8	60.3
Mathematics USE mean score (0-100)*	40.7	43.4	44.7
Budget expenditures per student (rubles)	53,390	57,325	73,585

Source: Russian Regions School Sample, 2010.
* Including students who took USE after finishing vocational school or after finishing secondary schools in previous years. In Russia mean USE scores were 58.0 for Russian and 43.7 for math in 2010.

We chose the schools in each region using stratified random sampling. After obtaining a list of all the schools in each region, schools were sorted into strata by region, school type (regular, magnet etc.), settlement type (rural, urban, *oblast/ krai* center), by administrative district and high school size (the number of 11th grade students). In each stratum, schools were selected using simple random sampling. The sample included 14.5 percent of all schools in Pskovskaya *oblast,* 8.9 percent in Yaroslavskaya *oblast,* and 4.1 percent in Krasnoyarsky *krai*.

We surveyed three types of respondents: students, teachers, and school principals. The student survey questionnaire asked students about their individual and family background characteristics. We also obtained data from the school on students' 10th grade grades and later, on students' math and Russian-language USE scores (in summer 2010, after students took the USE). The teacher survey form asked math and Russian teachers about their background (including gender, birth year, education, teaching experience), qualifications, textbooks and teaching practices used. The principal survey included questions on the principal's characteristics, school characteristics (urban, rural, size), and the number of basic and advanced classes (in each subject). We also asked principals to provide information on each school's course curricula (in particular, hours taught in mathematics and Russian in each class in the 11th grade).

Outcomes

The USE scores from the two mandatory subject-specific exams, Russian language and mathematics, are the primary outcome variables in our subsequent analyses. The USE scores were originally reported on 100-point scales, but for ease of interpretation, we convert these 100-point scales into z-scores. Because the subject-specific exams are meant to test students' knowledge of the national (standardized) curriculum, the USE scores are considered valid measures of high school students' performance. External agencies create, proctor, and grade the exams, and the exams have been shown to be reliable.[8]

Teaching practices

Our treatment variables are teaching practices used by 11th grade teachers to help students prepare for the USE examination. Russian schools employ a variety of methods to maximize student performance on the USE. We examine the impact of the three main types of such practices: (a) the proportion of homework exercises targeting specific entrance exam items ("test-specific homework exercises"); (b) teachers' use of practice tests ("practice tests"); and (c) teachers' use of websites geared to help students prepare for the exam ("websites"). For "test-specific homework exercises", teachers were asked to report the percentage of homework exercises of different types (including exercises which were and were not connected to USE preparation) that were assigned to students

[8] Cheating sometimes takes place (Koshkin, 2011), however we found no evidence of this in 2010, the year of our data collection, or in the set of schools we surveyed.

during the past semester. We standardized the percentage share of each type of homework exercise and divided the standardized distribution into three categories: small (more than 0.5 SD below the mean), medium (from – 0.5 to 0.5 SD around the mean) and high (more than 0.5 SD above the mean). For "practice tests", teachers were asked about how frequently tests (in the USE framework) were given to students over the academic year. Since there has been a dramatic recent growth of Internet-resources devoted to the USE, the "websites" questions asked teachers whether they were familiar with the examples of USE items and information about the USE on the websites and (if yes) how frequently they used these websites in their teaching.

Teacher characteristics

Our analysis of teaching practices controls for teacher characteristics. For historical reasons, variation in teacher characteristics is low in Russia. Almost all teachers in schools are women (97% of Russian language teachers and 94% of mathematics teachers as of October, 2009—the proportion of each is about 2 percentage points higher in our sample).[9] Because teacher turnover has been low in the past decades, teacher age is highly correlated with teacher work experience (and therefore in our analyses, controlling for teacher age is similar to controlling for teacher work experience). About 96% of teachers have completed higher education and 93% have completed higher pedagogical education, so it is not meaningful to use teacher education as a proxy for teacher skills. Furthermore, essentially every 11th grade teacher has a degree with a specialization in the subject matter they teach.

We therefore control for only two types of teacher characteristics in our estimated model: teacher work experience (in years) and teacher qualification. In Russian schools, teachers can attain higher qualification categories ("second," "first," and "highest") by undergoing a certification process.[10] One feature of the certification process is its timing. The certification process usually takes place once during a 5-year period; further, a teacher with the first qualification category has to wait at least two years before he/she can apply for the highest category. Thus, teachers who have achieved the higher qualification categories usually have considerably more work experience. Another feature of the certification process is that the quality of each teacher's students' academic work is taken into account in evaluating the teacher for certification. In part because of this inclusion of student performance, some higher category teachers have less work experience than others.

Subject Coverage (Basic and Advanced Track)

[9] Estimated from http://www.ed.gov.ru/files/materials/12928/83rik.pdf
[10] A teacher's qualification (along with seniority) is associated with a higher level of salary.

An important factor that may be correlated with both teaching practices and student achievement is the track of students the teacher is teaching. For each subject (e.g. math or Russian), students can be in one of two tracks: basic or advanced. Whether a student is in the basic or the advanced track for a particular subject partially determines the number of hours per week spent in that subject. Track status and the related number of teaching hours spent on a given subject per week are therefore key control variables in our analysis. In the advanced math track, students take 6 to 8 hours of math a week while students in the basic track take 4 or 5 hours. In the advanced Russian language track, students take 3 hours a week. Students in the basic Russian language track take 1-2 hours a week.[11] We coded 4 hours of math and 1 hour of Russian as "basic subject study with the lowest exposure." Five hours of math and 2 hours of Russian were coded as "basic subject study with the highest exposure." We coded 6 or more hours of math and 3 or more hours of Russian as "advanced subject study."

Students and classroom characteristics

In our cross-subject student fixed effects analysis of the causal impacts of teaching practices on student achievement (see the Section 4 below for more details), we also control for certain (cross-subject) student characteristics. Most importantly, we control for students' grades in 10th grade subjects (algebra and Russian, specifically). We thus account for the possibility that students with higher grades in mathematics or Russian could be sorted into classes taught by math or Russian teachers who use more effective teaching practices. We also control for average peer grades in algebra and Russian—the average grades of each student's classmates in the 10th grade class. Classes with higher average grades probably influence teachers to step up the amount of subject-matter items they ask students to do and may influence the types of teaching practices they use. As we explain in detail below, our analytical strategy assumes that we do not need to control for student and family background variables that stay constant across mathematics and Russian subjects within the same student.

Descriptive statistics of the variables used in the analyses are shown in Table 2.

[11] To compare educational standards for basic and advanced Russian language and mathematics study in high school see http://www.ed.gov.ru/edusupp/metodobesp/component/9067/

Table 2. Descriptive Statistics for Russian Regions' Sample, 2010

	Total Sample (N=2927)				Mathematics advanced study (N=1101)				Both subjects basic study (N=1481)			
Variable	***Mathematics***		***Russian***		***Mathematics***		***Russian***		***Mathematics***		***Russian***	
	Mean	*SD*	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*
A. Non-Subject-Specific Characteristics												
USE test score (1-100)	45.5	14.6	61.2	11.0	51.7	14.6	64.3	10.0	42.0	13.4	59.2	11.2
Student's gender: male	0.41	0.49	---	---	0.44	0.50	---	---	0.40	0.49	---	---
0-100 books in home	0.54	0.50	---	---	0.47	0.50	---	---	0.58	0.49	---	---
Rural location	0.17	0.37	---	---	0.06	0.24	---	---	0.24	0.43	---	---
Town	0.43	0.50	---	---	0.48	0.50	---	---	0.40	0.49	---	---
City, regional center	0.40	0.49	---	---	0.46	0.50	---	---	0.35	0.48	---	---
School size	601	340	---	---	701	351	---	---	524	305	---	---
B. Subject-Specific Characteristic												
10th grade marks: "good" or "excellent"	0.54	0.50	0.60	0.49	0.62	0.48	0.69	0.46	0.49	0.50	0.55	0.50
Class average 10th grade marks	3.66	0.31	3.71	0.30	3.80	0.34	3.84	0.31	3.58	0.25	3.63	0.24
Possible to choose basic or advanced study in school	0.37	0.48	0.13	0.34	0.49	0.50	0.22	0.41	0.22	0.41	---	---
Basic subject study with the lowest exposure	0.22	0.41	0.42	0.49	---	---	0.44	0.50	0.34	0.47	0.50	0.50
Basic subject study with the highest exposure	0.41	0.49	0.38	0.49	---	---	0.35	0.48	0.66	0.47	0.50	0.50
Advanced subject study	0.38	0.48	0.20	0.40	---	---	0.21	0.40	---	---	---	---
Teacher experience: <=10 years	0.10	0.30	0.09	0.29	0.07	0.26	0.03	0.16	0.13	0.34	0.10	0.30
Teacher experience: 11-20 years	0.23	0.42	0.25	0.43	0.31	0.46	0.29	0.45	0.20	0.40	0.25	0.43
Teacher experience: 21-30 years	0.38	0.49	0.46	0.50	0.32	0.47	0.42	0.49	0.45	0.50	0.51	0.50
Teacher experience: >=31 years	0.28	0.45	0.20	0.40	0.30	0.46	0.26	0.44	0.22	0.42	0.14	0.34
Teacher has 2nd or no category	0.12	0.32	0.15	0.35	0.03	0.16	0.13	0.34	0.16	0.37	0.17	0.38
Teacher has the 1st category	0.46	0.50	0.37	0.48	0.39	0.49	0.34	0.47	0.56	0.50	0.38	0.49
Teacher has the highest category	0.42	0.49	0.48	0.50	0.59	0.49	0.52	0.50	0.28	0.45	0.44	0.50
Low amount of B-type exercises in homework	0.28	0.45	0.28	0.45	0.37	0.48	0.21	0.41	0.17	0.38	0.30	0.46
Medium amount of B-type exercises in homework	0.42	0.49	0.50	0.50	0.44	0.50	0.59	0.49	0.43	0.49	0.46	0.50
High amount of B-type exercises in homework	0.30	0.46	0.21	0.41	0.19	0.39	0.20	0.40	0.40	0.49	0.25	0.43
Low amount of C-type exercises in homework	0.37	0.48	0.36	0.48	0.17	0.38	0.33	0.47	0.54	0.50	0.37	0.48
Medium amount of C-type exercises in homework	0.26	0.44	0.38	0.48	0.23	0.42	0.41	0.49	0.21	0.41	0.36	0.48
High amount of C-type exercises in homework	0.37	0.48	0.26	0.44	0.60	0.49	0.27	0.44	0.25	0.43	0.28	0.45
Tests in USE frame 2-3 times a semester or less	0.21	0.41	0.17	0.37	0.21	0.41	0.18	0.38	0.23	0.42	0.14	0.35
Tests in USE frame once a month	0.26	0.44	0.34	0.47	0.19	0.39	0.28	0.45	0.27	0.44	0.42	0.49
Tests in USE frame 2-3 times a month	0.40	0.49	0.34	0.47	0.47	0.50	0.36	0.48	0.38	0.49	0.31	0.46
Tests in USE frame once a week or more	0.13	0.34	0.16	0.36	0.14	0.34	0.18	0.39	0.11	0.32	0.13	0.33
USE Web-sites used less than once a month	0.19	0.39	0.39	0.49	0.14	0.35	0.36	0.48	0.22	0.42	0.40	0.49
USE Web-sites used several times a month	0.34	0.48	0.37	0.48	0.39	0.49	0.43	0.50	0.33	0.47	0.35	0.48
USE Web-sites used several times a week	0.47	0.50	0.24	0.43	0.47	0.50	0.21	0.41	0.44	0.50	0.26	0.44

Source: Russian Regions Schools Sample, 2010.

4. Estimation Strategy

Estimating the impact of teaching practices on student performance can be complicated by omitted variables bias. Traditional (e.g. ordinary least squares or OLS) analyses of the relationship between school inputs and student outcomes often do not account for non-random assignment of students to schools. Many parents make an effort to choose their child's school (Lankford et al., 2002; Bonesrønning et al., 2005). Children from families with greater economic and cultural capital are likely to attend schools with better resources. In Russia, in particular, students' family characteristics are positively correlated with the quality of school resources (Kuzmina and Tyumeneva, 2011).

Moreover, students are generally not randomly assigned to teachers who have differing characteristics and teaching practices (Clotfelter et al, 2010). Selection into classes by student ability level is rather widespread in Russia. Our sample contains a number of schools with basic and advanced classes, particularly in mathematics. Higher social class students are more likely to be in advanced classes, and also in urban, regional capital city schools. According to our data, higher category teachers are more likely to teach higher social class students and students in the advanced mathematics classes (Table 3).

Table 3. Correlations of Measures of Student Family Background and Whether Student is in Advanced Mathematics or Russian with Teachers' Qualification

Variable	Math Teacher Lowest Qualification Category	Math Teacher Highest Qualification Category	Russian Teacher Lowest Qualification Category	Russian Teacher Highest Qualification Category
Total sample (N=2927)				
At least One Parent Higher Education	-0.119***	0.148***	-0.041**	0.142***
< 100 Books in the Home	0.079***	-0.021	0.038**	-0.086***
Student in Advanced Math	-0.219***	0.257***		
Student in Advanced Russian			-0.128***	-0.018
Advanced math study subsample (N=1101)				
At least One Parent Higher Education	-0.116***	0.149***	-0.039	0.111***
< 100 Books in the Home	0.066**	0.040	0.101***	-0.034
Advanced Russian study subsample (N=572)				
At least One Parent Higher Education	-0.121***	0.129***	-0.164***	0.176***
< 100 Books in the Home	-0.096**	0.061	-0.090**	-0.093**
Both subjects basic study subsample (N=1481)				
At least One Parent Higher Education	-0.089***	0.096***	-0.004	0.140***
< 100 Books in the Home	0.113***	-0.046*	0.013	-0.115***

Source: Russian Regions Schools Sample, 2010.
*** Statistically significant at the 1 percent significance level.
** Statistically significant at the 5 percent significance level.
*Statistically significant at the 10 percent significance level.

The non-random assignment of students makes it more difficult to estimate unbiased impacts of teaching practices on student achievement. To address issues of selection bias, researchers have used different types of student fixed effects models (Dee, 2007; Dee and Cohodes, 2008; Clotfelter et al, 2010; Schwerdt and Wuppermann, 2011; van Klaveren, 2011). Following Clotfelter et al (2010) we implement a cross-subject student fixed effect model that utilizes variation within the same student but across different subjects to identify the impact of different teachers with different characteristics/practices. The cross-subject student fixed effect model is derived from the traditional education production function:

$$y_{is} = T_{is}\alpha + x_{is}\beta + z_i\Upsilon + u_i + \varepsilon_{is},\ i = 1,\ \dots N,\ s = 1,\dots S \quad (1)$$

where N is the number of individuals;

S is the number of subjects (in our case S=2);

y_{is} is the USE score of student i in subject s;

T_{is} is a vector of treatment variables (teaching practices) that vary across students and subjects;

x_{is} is a vector of teacher, classroom and student characteristics that vary across students and subjects;

z_i is a vector of school, family, student, teacher and classroom characteristics that vary only across students but not across subjects;

u_i is a student-specific error term (that represents unobservable variation across students);

ε_{is} is an error term that varies across both students and subjects.

Traditional OLS approaches produce biased estimates α of the impact of the treatment T_{is} on the outcome y_{is} if the error term ($u_i + \varepsilon_{is}$) is correlated with both the treatment and the outcome.

The cross-subject student fixed effects model attempts to control for the problematic correlation between the error term that varies across students but not across subjects and the treatment and outcome variables. In particular, by subtracting from each variable in equation (1) within student cross-subjects average of that variable the model effectively eliminates $z_i\Upsilon$ and u_i (observable and unobservable factors that were constant across subjects but not across students):

$$y_{is} - \overline{y_i} = (T_{is} - \overline{T_i})\delta + (x_{is} - \overline{x_i})\phi + (\varepsilon_{is} - \overline{\varepsilon_i}), \quad (2)$$

where $\overline{y_i} = \frac{1}{S}\sum_{s=1}^{S} y_{is},\ \overline{x_i} = \frac{1}{S}\sum_{s=1}^{S} x_{is},\ \overline{T_i} = \frac{1}{S}\sum_{s=1}^{S} T_{is},\ \overline{\varepsilon_i} = \frac{1}{S}\sum_{s=1}^{S} \varepsilon_{is}$.

The above model (2) produces unbiased estimates of δ under a few assumptions. The first assumption is that coefficients for each variable are equal across the two subjects (Dee, 2005). This implies that the way in which the treatment (and other teacher characteristics) affects student achievement is the same across subjects. The second assumption is that the error term $(\varepsilon_{is} - \overline{\varepsilon_{\iota}})$ in equation (2) is uncorrelated with the regressors $(T_{is} - \overline{T_{\iota}})$. This means that unobserved student, classroom, or teacher characteristics that vary across subjects are not correlated with the teaching practice and student achievement (Schwerdt and Wuppermann, 2011).

The cross-subject student fixed effects model may not address biases stemming from the non-random assignment of teaching characteristics/practices to students with greater abilities in mathematics or Russian language. Controlling for cross-subject variation in student ability—as well as other cross-subject factors that may be correlated with the treatment and outcome variables—may address this threat to the internal validity of the causal estimates (Clotfelter et al, 2010). Indeed, in all of our analyses, we control for several important cross-subject factors such as student's grades, average peer grades, student's track status (basic subject study with the lowest exposure, basic subject study with the highest exposure, or advanced subject study), as well as a variety of teacher characteristics.[12]

5. Results

The results of our model show that some of the tools available to high school teachers do have a positive effect on student performance on high-stakes tests. According to our estimates in Table 4, a high amount of exposure to test-specific homework exercises positively impacts student performance on the USE. High amounts of B-types of test-specific homework exercises, for example, increase USE scores by about 0.10-14 SDs (Table 4, columns 2, 4, 6, 8) Similarly, high amounts of C-types of test-specific homework exercises increase USE scores by about 0.08-0.13 SDs (Table 4, columns 3, 4, 6, 8). Both sets of results are statistically significant at the 5 percent level.

Our results also show, however, that neither "practice tests" nor "websites" have a significant impact on high school student performance. According to Table 4 (columns 5, 6, 8), the impact of "practice tests" on USE scores is small and not significantly different from zero, even at the 10 percent level. The impact of "websites" on USE scores is also small and not significantly different from zero (Table 4, columns 7, 8). Taken together, the findings show that some teaching practices are distinctly more effective than others at improving high school student performance on high-stakes tests.

[12] We also addressed the clustered structure of the data. Students are grouped in classes and in schools. This potentially may produce biased estimates of standard errors. To eliminate this bias we use cluster correction of standard errors in all the models.

Table 4. Three Russian Regions: Cross-subject Student Fixed Effects Estimates of Teaching Strategies Directed at Improving Student USE Test Results, 2010.

Variables	1	2	3	4	5	6	7	8
Basic subject study with the lowest exposure (0=no, 1=yes)	0.08	0.08	0.07	0.07	0.07	0.06	0.07	0.06
	(0.07)	(0.07)	(0.07)	(0.06)	(0.07)	(0.07)	(0.07)	(0.07)
Advanced subject study (0=no, 1=yes)	0.22***	0.22***	0.21***	0.20***	0.22***	0.20***	0.23***	0.21***
	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.06)	(0.05)
Low amount of B-type exercises in homework (0=no, 1=yes)		0.00		-0.01		-0.01		-0.00
		(0.06)		(0.05)		(0.05)		(0.05)
High amount of B-type exercises in homework (0=no, 1=yes)		0.10*		0.13**		0.14**		0.13**
		(0.05)		(0.06)		(0.06)		(0.05)
Low amount of C-type exercises in homework (0=no, 1=yes)			-0.01	-0.02		-0.01		-0.01
			(0.04)	(0.04)		(0.04)		(0.04)
High amount of C-type exercises in homework (0=no, 1=yes)			0.08	0.12**		0.12**		0.13***
			(0.05)	(0.05)		(0.05)		(0.05)
Tests in USE frame 2-3 times a semester or less (0=no, 1=yes)					0.03	-0.02		-0.01
					(0.07)	(0.06)		(0.06)
Tests in USE frame 2-3 times a month (0=no, 1=yes)					-0.03	-0.05		-0.04
					(0.06)	(0.05)		(0.05)
Tests in USE frame once a week or more (0=no, 1=yes)					-0.03	-0.04		-0.01
					(0.07)	(0.07)		(0.06)
USE Web-sites used less than once a month (0=no, 1=yes)							0.07	0.07
							(0.06)	(0.05)
USE Web-sites used several times a week (0=no, 1=yes)							-0.06	-0.06
							(0.05)	(0.05)
Constant	0.28	0.15	0.21	0.01	0.24	0.01	0.40	0.13
	(0.55)	(0.54)	(0.55)	(0.55)	(0.56)	(0.56)	(0.51)	(0.52)
Tenth grade individual and class average grades included?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Teacher experience and category dummies included?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	5,850	5,850	5,850	5,850	5,850	5,850	5,850	5,850
Number of individual students	2,925	2,925	2,925	2,925	2,925	2,925	2,925	2,925
Adjusted R-squared	0.06	0.07	0.07	0.08	0.06	0.08	0.07	0.08

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Source: Russian Regions Schools Sample, 2010.

Among the coefficients in Table 4, we also find that the coefficient of students' track has a potentially large and significant impact on student performance. Specifically, students in the advanced track in either math or Russian language score much higher on the USE than students in the basic track. The large impacts further imply that teaching processes may in fact have heterogeneous effects across the two tracks. To test for this possibility, we estimated two additional cross-subject student fixed effects models: (a) one for students who were in the basic track in both subjects; and (b) one for students who were in the advanced mathematics track. We chose to look at the impacts of teaching practices on students in the advanced track mathematics because there was a much higher fraction of students in the advanced mathematics track compared to the advanced Russian in our sample (there were only seven advanced Russian classes).

For students in the basic track, we find that assigning high amounts of B-type and high amounts of C-type homework have a positive impact on students' USE performance. This positive impact is larger than in the model estimating the effects for the entire sample. High amounts of B-type homework, which is oriented to more basic test items, increases the performance of students in the basic track by 0.15-0.17 SDs (Table 5, columns 1, 3, 5, 7). Low and high amounts of C-type homework also increase students' USE performance (Table 5, columns 2, 3, 5, 7).

The positive effect of low amounts (as opposed to medium amounts—the left out category) of C-type homework on the USE scores can be explained by the observation that B-type and C-type homework appear to be "substitutes," in the sense that we find a high negative correlation (-0.48) between teachers in the basic track giving high amounts of B- and C-type homework. We also find a positive correlation (0.39) between teachers in the basic track giving high amounts of B-type homework and low amounts of C-type homework. The correlations suggest that teachers who give high amounts of C-type homework are less likely to give high amounts of B-type homework (and vice versa). Rather, teachers who give high amounts of B-type homework are likely to give low amounts of C-type homework (and vice versa). Thus one or the other type of homework in high amounts contributes to students' higher performance on the USE test.

To the contrary, neither a moderate or high amount of USE test practice nor USE website use has a positive impact on the USE performance of students in the basic track. When all the teacher practice variables are included in the estimate, website use even shows a negative and significant (at a 10 percent level) effect on test performance. This negative coefficient suggests that students assigned high amounts of website use do less well on the USE test. This is not an artifact of the correlation of teachers assigning high amounts of website use with either assigning high amounts of

B-type or C-type homework. Those correlations are very small. The result may mean that students assigned high amounts of website use are, in their time allocation at home, diverted from doing homework assignments into an activity that is not effective in improving their test scores. Thus, they score lower on the USE test than students who are assigned website use less frequently.

Table 5. Three Russian Regions: Cross-subject Student Fixed Effects Estimates of Teaching Strategies Directed at Improving Students' USE Test Results for Sample of Student Studying both Russian Language and Mathematics at the Basic Level, 2010.

Variables	1	2	3	4	5	6	7
Basic subject study with the lowest exposure (0=no, 1=yes)	0.05	0.08	0.06	0.06	0.07	0.07	0.09
	(0.08)	(0.09)	(0.08)	(0.09)	(0.08)	(0.08)	(0.08)
Low amount of B-type exercises in homework (0=no, 1=yes)	0.01		-0.01		-0.00		-0.02
	(0.06)		(0.05)		(0.05)		(0.05)
High amount of B-type exercises in homework (0=no, 1=yes)	0.15***		0.16***		0.17***		0.17***
	(0.05)		(0.06)		(0.06)		(0.05)
Low amount of C-type exercises in homework (0=no, 1=yes)		0.11**	0.09*		0.11**		0.12***
		(0.05)	(0.05)		(0.04)		(0.04)
High amount of C-type exercises in homework (0=no, 1=yes)		0.10	0.14**		0.16**		0.16**
		(0.07)	(0.06)		(0.07)		(0.06)
Tests in USE frame 2-3 times a semester or less (0=no, 1=yes)				-0.01	-0.09		-0.06
				(0.07)	(0.07)		(0.06)
Tests in USE frame 2-3 times a month (0=no, 1=yes)				-0.07	-0.12		-0.08
				(0.07)	(0.07)		(0.07)
Tests in USE frame once a week or more (0=no, 1=yes)				-0.05	-0.05		0.00
				(0.08)	(0.07)		(0.08)
USE Web-sites used less than once a month (0=no, 1=yes)						0.03	0.01
						(0.08)	(0.07)
USE Web-sites used several times a week (0=no, 1=yes)						-0.11	-0.13*
						(0.08)	(0.06)
Constant	-0.27	0.12	-0.25	-0.07	-0.25	0.05	-0.26
	(0.73)	(0.80)	(0.74)	(0.80)	(0.72)	(0.72)	(0.64)
Tenth grade individual and class average grades included?	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Teacher experience and teacher category dummies included?	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Dummy for schools that have only basic track and no advanced track classes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,200	2,200	2,200	2,200	2,200	2,200	2,200
Number of individual students	1,100	1,100	1,100	1,100	1,100	1,100	1,100
Adjusted R-squared	0.08	0.14	0.14	0.08	0.14	0.08	0.14

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
Source: Russian Regions Schools Sample, 2010.

Table 6. Three Russian Regions: Cross-subject Student Fixed Effects Estimates of Teaching Strategies Directed at Improving Students' USE Test Results for Sample of Student in the Advanced (Mathematics) Track, 2010.

Variables	1	2	3	4	5	6	7
Low amount of B-type exercises in homework (0=no, 1=yes)	-0.00		-0.05		0.00		-0.00
	(0.10)		(0.06)		(0.07)		(0.07)
High amount of B-type exercises in homework (0=no, 1=yes)	-0.01		0.07		0.10		0.08
	(0.10)		(0.08)		(0.08)		(0.08)
Low amount of C-type exercises in homework (0=no, 1=yes)		-0.28***	-0.28***		-0.30***		-0.31***
		(0.07)	(0.07)		(0.07)		(0.07)
High amount of C-type exercises in homework (0=no, 1=yes)		0.13**	0.15**		0.15**		0.15**
		(0.06)	(0.07)		(0.07)		(0.07)
Tests in USE frame 2-3 times a semester or less (0=no, 1=yes)				0.07	-0.02		-0.01
				(0.13)	(0.12)		(0.12)
Tests in USE frame 2-3 times a month (0=no, 1=yes)				0.07	0.09		0.11
				(0.11)	(0.10)		(0.10)
Tests in USE frame once a week or more (0=no, 1=yes)				-0.03	-0.02		0.02
				(0.11)	(0.10)		(0.11)
USE Web-sites used less than once a month (0=no, 1=yes)						0.03	0.05
						(0.10)	(0.09)
USE Web-sites used several times a week (0=no, 1=yes)						0.06	-0.01
						(0.08)	(0.06)
Constant	2.43***	1.83**	1.66**	2.22***	1.72**	2.49***	1.85**
	(0.83)	(0.79)	(0.76)	(0.80)	(0.81)	(0.87)	(0.87)
Tenth grade individual and class average grades included?	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Teacher experience and teacher category dummies included?	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,960	2,960	2,960	2,960	2,960	2,960	2,960
Number of individual students	1,480	1,480	1,480	1,480	1,480	1,480	1,480
Adjusted R-squared	0.08	0.08	0.09	0.07	0.09	0.08	0.10

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
Source: Russian Regions Schools Sample, 2010.

For the students in the advanced mathematics track, low amounts of C-type homework have a robust negative effect on students' USE performance and high amounts of C-type homework have a robust positive effect on students' USE performance (Table 6). More B-type homework does not contribute to higher USE performance. Thus, for students in the advanced track, doing (more complex) C-type homework has a positive influence on USE scores. The effects are quite large, particularly for medium amounts of C-type homework (the negative coefficient of low amounts of homework is the difference between low amounts of homework assigned and medium amounts assigned). The estimated effect of assigning medium amounts of C-type homework is about 0.30 SDs and of assigning high amounts of C-type homework (compared to assigning medium amounts) is about 0.15 SDs (Table 6, columns 2.3.5.7). The results also indicate that the exam performance of students in the advanced track is influenced by how they perform on the more difficult test items. Other teaching strategies (more USE tests or web-based USE practice) do not affect how these advanced mathematics students perform on the USE test.

In sum, the study's results for a sample of students from three Russian regions indicate that assigning higher amounts of homework targeting specific exam items on the high-stakes end-of-school/college entrance examination improves students' performance on that exam. Among the various available strategies employed by 11th grade teachers, such homework assignments are the only strategies we were able to identify having a positive effect on students' test performance. The use of practice tests and web-based exercises do not appear to influence high-stakes exam scores or, in the case of frequently assigning website use for basic level students, may even have a negative effect on their test performance.

When we divide the sample into students who are in the basic track in both Russian and mathematics, we find that for students in the basic track, high amounts of B-type homework has a positive effect on USE performance, and likely also high amounts of C-type homework. However, for students in the advanced track, only more C-type homework affects USE performance. For both groups of students, the effects sizes of higher amounts of homework are larger than when we estimate the model for all students together. This is particularly true for advanced track students. For students in basic level courses, high amounts of B-type homework exercises could increase students' average USE scores by almost 0.2 SDs (Table 5), which would affect their college choices. Advanced track students with teachers that gave medium amounts of homework could increase their test scores by almost one-third of a standard deviation and those with teachers assigning high amounts of C-type homework exercises another 0.15 standard deviations (Table 6).

6. Conclusions

We can draw three important conclusions from the results of our study. The first is that specific teaching practices can raise high school students' outcomes on high-stakes exams. Specifically, targeting homework exercises of the appropriate level of difficulty at students of different levels of performance can increase student achievement on these exams. The impact of targeting is particularly notable for students in the advanced mathematics track where assigning more homework in the more difficult C-type items in mathematics and Russian has a positive and large effect on their test performance.

The second conclusion we draw from the study is that some teaching strategies are more effective than others. We show that the type and amount of test-specific homework teachers assign to students preparing for high-stakes exams positively affects student exam performance, and that doing more of the type of homework most relevant to the level of the class is also important. Students who are already in an advanced course benefit more from homework exercises that focus on more difficult problems and students at the basic level profit from additional exercises at the basic level and probably also at the advanced level. However, having students take practice tests—a popular method of test preparation in many countries (see, for example, Becker, 1990 for a review of the effects of outside of school tutoring on college entrance test scores in the United States) — has no significant effect on high-stakes exam performance whether students are in a basic or advanced track. The use of web-based exam preparation materials also does not have a positive effect on students' scores and may even have a negative effect on basic track students' test scores when they are assigned website use very frequently as a way to study for the USE test.

The third conclusion that we draw is that the *amount* of use of specific practices, namely targeted type of homework assignments, can have larger or smaller effects on improving students' high stakes test performance. For basic level students, we found that assigning high amounts of B-type or C-type test item homework has significant positive effects, whereas assigning medium of either type of homework does not. To the contrary, for advanced track students, assigning C-type rather than B-type of homework is important for improving test performance, but both medium and high amounts of C-type homework have significant positive effects on students' test scores. Indeed, increasing the amount of C-type homework from the low to the medium amount has about double the effect size on test scores as increasing from the medium to the high amount.

The wider implications of these estimates for high school students are also worth discussing. They support Bishop's earlier assertion that school exit exams intended as a summative evaluation of the educational system's curriculum focus teachers' and school efforts on teaching to that summative evaluation in specific ways.[13] Yet, beyond that, the results of our study tend to support his argument that successful teaching to these high-stakes tests focuses more on problem-solving strategies rather than rote-learning (Bishop, 1997). The fact that high amounts of homework aimed

[13] Many school exit exams also serve as college entrance tests, although Bishop argued that teachers have less of a stake in college entrance exam scores. In Russia, however, a very high percentage of students who take the USE intend to go to college.

at learning how to do USE B- and C-type questions has a positive effect but that more in class test taking does not suggests that good "teaching to the test" may not be so different from what many consider good teaching. Such homework is geared to learning how to interpret more complex problems. It is generally not memorization or rote learning.

References

Bishop, J. (1996). The impact of curriculum-based external examinations on school priorities and student learning. *International Journal of Education Research*, 23 (8): 653–752

Becker, B. (1990). Coaching for the Aptitude Test: Further Synthesis and Appraisal. *Review of Educational Research*, 60, 373 – 417.

Bishop, J. (1997). The Effect of National Standards and Curriculum-Based Exams on Achievement. *American Economic Review*, 87, 2: 260-264.

Bonesrønning, H., Falch, T., & Strøm, B. (2005). Teacher sorting, teacher quality, and student composition. *European Economic Review*, 49: 457–83.

Boyd, D., Grossman, P., Lankford, H., Loeb, S., & Wyckoff, J. (2006). How changes in entry requirements alter the teacher workforce and affect student achievement. *Education Finance and Policy*, 1(2): 176-216.

Carnoy, M., Chilisa, B., & Chisholm, L. (2012). The Low Achievement Trap: Comparing Schooling in Botswana and South Africa. Cape Town: HSRC Press.

Carnoy, M., Loyalka, P., Froumin, I., Dossani, R., Tilak J., & Rong W. (2013). Higher education in the global knowledge economy: Triumph of the BRICS? Stanford, CA: Stanford University Press.

Clotfelter, C.T., Ladd, H.F. & Vigdor, J.L. (2007). Teacher Credentials and Student Achievement: Longitudinal Analysis with Student Fixed Effects. *Economics of Education Review*, 26 (6): 673-682.

Clotfelter, C.T., Ladd, H.F. & Vigdor, J.L. (2010). Teacher Credentials and Student Achievement in High School: A Cross-Subject Analysis with Student Fixed Effects. *Journal of Human Resources*, 45(3): 655 – 682.

Darling-Hammond, L. (2009). Educational opportunity and alternative certification: New evidence and new questions. A SCOPE Policy Brief.

Dee, T. (2007). Teachers and the Gender Gaps in Student Achievement. *Journal of Human Resources*, 42 (3): 528-554.

Dee, T. & Cohodes, S. (2008). Out-of-Field Teaching and Student Achievement: Evidence from Matched-Pairs Comparisons. *Public Finance Review*, 36 (1): 7-32.

Ferguson, R. & Ladd, H. (1996). How and Why Money Matters: An Analysis of Alabama Schools. In H.F. Ladd, (ed.) Holding School Accountable. Washington, DC: Brookings Institution, 265-98.

Goldhaber, D. & Brewer, D. (1997). Evaluating the Effect of Teacher Degree Level on Educational Performance. In W. Fowler (Ed.), *Developments in School Finance 1996* (pp. 197-210). Washington, DC: NCES.

Goldhaber, D. D., & Brewer, D. J. (2000). Does teacher certification matter? High school teacher certification status and student achievement. *Education Evaluation and Policy Analysis*, 22: 129-45.

Hanushek, E.A., Kain, J.A. & Rivkin, S.G. (2005). Teachers, Schools, and Academic Achievement. *Econometrica*. 73 (March): 417-58.

Hill. H. Rowan, B., & Ball, D. (2005). Effects of Teachers' Mathematical Knowledge for Teaching on Student Achievement. *American Educational Research Journal*, 42. 2: 371-406.

Kane, T., McCaffrey, D., Miller, T., Staiger, D. (2013). Have We Identified Effective Teachers? Seattle, WA: Bill and Melinda Gates Foundation. http://www.metproject.org/downloads/MET_Validating_Using_Random_Assignment_Research_Paper.pdf Downloaded July 7, 2013.

Koshkin, P. (2011). Cheating on Russia's Countrywide Exam Has Further Fueled Debates Over Its Objectivity. National Research University Higher School of Economics Press Release, July 18. Moscow: HSE.

Kouzmina Y. & Tyumeneva Y. (2011). Chitatelskaya gramotnost' 15-letnih shkolnikov: znachimost' semeynih, individual'nih I schkol'nih characteristic (po dannim rossiyskoy viborki PISA-2009). *Voprosi obrazovania*, 3, 164-191.

Kukla-Acevedo, S. (2009). Do teacher characteristics matter? New results on the effects of teacher preparation on student achievement. *Economics of Education Review*, 28(1), 49–57.

Lankford, M., Loeb, S., & Wyckoff, J. (2002). Teacher Sorting and the Plight of Urban Schools. A Descriptive Analysis. *Educational Evaluation and Policy Analysis*. 24(1), 37–62.

Monk, D. H. (1994). Subject Area Preparation of Secondary Mathematics and Science Teachers and Student Achievement. *Economics of Education Review*. Vol. 13(2): 125-45.

Nye, B., Konstantopoulos, S, & Hedges, L.V. (2004). How Large are Teacher Effects? *Educational Evaluation and Policy Analysis*, 26, 237-257

Rice, J. (2003). Teacher quality: Understanding the effectiveness of teacher attributes. Washington, DC: Economic Policy Institute.

Rockoff, J. (2004). The impact of individual teachers on student achievement: Evidence from panel data. *American Economic Review*, 94, 247–252.

Sanders, W.L., & Rivers, J.C. (1996). Cumulative and residual effects of teachers on future student academic achievement. Research Progress Report, University of Tennessee Value-Added Research and Assessment Center.

Schwerdt, G., & Wuppermann, A. (2011). Is traditional teaching all that bad? A within-student between subject approach. *Economics of Education Review*, 30(2), 365–379.

van Klaveren, C. (2011). Lecturing style teaching and student performance. *Economics of Education Review*, 30(4), 729–739

Authors:

1. Andrey Zakharov (corresponding author). National Research University Higher School of Economics. International Laboratory for Educational Policy Research. Deputy Head. Email: ab.zakharov@gmail.com
2. Martin Carnoy. Stanford University. Vida Jacks Professor of Education. Email: carnoy@stanford.edu
3. Prashant Loyalka. Stanford University. Freeman Spogli Institute for International Studies. Center Research Fellow. Email: loyalka@stanford.edu