Raising the Stakes: Inequality and Testing in the Russian Education System

Michelle Jackson, *Department of Sociology, Stanford University*
Tatiana Khavenson and Tatiana Chirkina, *National Research University Higher School of Economics*

Sociologists have argued that high-stakes tests open the door to high levels of educational inequality at transition points: in a high-stakes testing regime, parents and students are able to focus all energy and resources on test preparation, thus enhancing pre-existing inequalities in academic performance. But arguments about a special role for high-stakes tests are often prosecuted without explicit comparisons to other types of tests and assessments, usually because information on other tests is not available. In this article, we analyze a unique dataset on a contemporary cohort of Russian students, for whom we have PISA and TIMSS scores, low-stakes test scores, and high-stakes test scores. We compare the role each test plays in mediating socioeconomic background inequalities at the important transitions in the Russian educational system: the transition to upper secondary education and the transition to university. We find evidence in favor of a special role for the high-stakes test at the transition to university, but we also find evidence that gives cause to question the standard assumption that high-stakes tests should be a primary focus for those concerned about inequality of educational opportunity.

Introduction

In many countries, educational transitions are prefaced by tests and examinations designed to select students for the subsequent level of study. The extent to which selection is determined by performance in tests/examinations varies significantly across countries, across transitions within countries, and over time. In some countries, these tests/examinations are "high-stakes," and scores are

In this article, we present analyses of data drawn from the "Trajectories in Education and Career" study (TrEC—http://trec.hse.ru/). The research was undertaken within the framework of the Basic Research Program at the National Research University Higher School of Economics (HSE) and supported by a subsidy from the Russian Academic Excellence Project "5-100." We thank Martin Carnoy, Isak Froumin, Prashant Loyalka, participants at the CIES conference (San Francisco, April 2019), and the anonymous reviewers for helpful advice on earlier versions of this article. *Direct correspondence to Michelle Jackson, Department of Sociology, Stanford University: mvjsoc@stanford.edu*

Social Forces 1–23
doi:10.1093/sf/soz113

the primary selection mechanism at the transition. In other countries, "low-stakes" tests and class grades simply aid the student decision-making process by providing evidence on the likely chances of academic success. Layered on top of country-specific school testing and grading regimes in more recent years are international performance tests such as PISA "Programme for International Student Assessment" and TIMSS "Trends in International Mathematics and Science Study". These tests are "no-stakes" for students, but potentially high-stakes for schools and countries, which may be judged deficient, if scores are poor.

Sociologists of educational inequality have long argued that high-stakes tests open the door to high levels of educational inequality at transition points (Alon 2009; Suen and Wu 2006; see Heubert and Hauser 1999, for a discussion). To a much greater extent than for no- and low-stakes assessments, high-stakes tests allow for purposive action on the part of parents and students. If a test is given the power to determine future educational trajectories, parents and students can focus all effort and resources on that test in the full knowledge that the investment will pay off (e.g. Alon and Tienda 2007; Alon 2009; Buchmann et al. 2010). As a consequence, inequality of educational opportunity might increase relative to a system in which no high-stakes test exists.[1] At the other extreme, a no-stakes test is assumed to be inconsequential for student decisions and selection at educational transitions, because students, parents, and educational administrators are ignorant of any individual student's performance on these tests. Any inequality that exists in no-stakes test scores is reasonably assumed to result from social background differences in academic ability, socialization, and schooling.

Despite the distinct mechanisms proposed to be at play with respect to performance on high-, low-, and no-stakes tests, much existing educational transitions research has taken a relatively casual approach to performance measurement. Student performance is alternatively operationalized through country-specific examination results or grades, survey-specific test scores, or PISA/TIMSS results. These different operationalizations are justified either on the basis of theoretical equivalence—upon the assumption that all operationalizations measure the same underlying concept—or on the basis of pragmatism, because alternative measures are not available.

This standard treatment of performance measures raises two questions. First, can arguments about the special role of high-stakes tests in creating inequality at educational transitions be sustained when we compare the contribution of high-stakes tests with the contribution of other types of tests? And second, is our understanding of the mechanisms underlying educational inequality actually changed if we use one measure of performance rather than another?

In this article, we use a unique dataset that links educational panel data for a sample of Russian students to their PISA and TIMSS results, and their final examination results. Using this dataset, we evaluate the extent to which different measures affect our estimates of the contribution of performance inequalities to inequalities at educational transitions. A further advantage of this dataset is that we can observe the progress of a single cohort of students traveling all the

way through the education system. We can examine inequalities in educational opportunity at different educational levels, and further, we can assess the degree of within-level, or qualitative, differentiation in the system. This makes it possible to provide a comprehensive assessment of inequality of educational opportunity across the whole educational system in contemporary Russia.

In the following sections, we describe previous work on the role of high-stakes and low-stakes tests in generating educational inequality. We then briefly outline the structure of the Russian educational system before moving on to describe our data and present results on testing and inequality of educational opportunity.

High-Stakes Versus Low-Stakes Versus No-Stakes Testing

The definition of "high-stakes" tests is inconsistent across the literature in educational sociology. Here, we define high-stakes tests as tests (and assessments) that are designed to play a significant role in determining a student's progression through the educational system. In some systems, high-stakes tests determine whether or not a student is permitted to enter the next level in the educational system; in others, high-stakes tests determine to which track a student is assigned.[2] By design, test performance constrains students' educational options, although in many countries students are still given some freedom to determine their educational destinies.[3] In high-stakes testing regimes, we would therefore expect test-score performance to be important in assigning students to levels/tracks. This might then be expected to have knock-on effects for educational inequality.

In addition to high-stakes tests, most educational systems preserve a role for "low-stakes" tests and grades. These assessments are administered within the educational system—to evaluate student learning, and to provide feedback, for example—but unlike tests administered in international assessments, these in-class tests still have the potential to alter student outcomes, albeit through their effect on student and teacher expectations. In the most stratified systems, each transition is associated with a high-stakes test, while the less stratified systems might have a high-stakes test at one educational transition, but a low-stakes test at other transitions.

In recent decades, the testing regimes of many countries have been supplemented by international tests, most notably PISA and TIMSS.[4] These tests allow for comparisons across countries, for evaluations of the effectiveness of different educational arrangements, and for comparisons over time, but they are not consequential for any individual's progress through the educational system (Ramirez et al. 2018; Kamens and McNeely 2010). Because the individual-level results of these tests are not known to students, teachers, or educational institutions, there is no possibility that the scores could influence educational decisions directly. In the literature, international assessments are often described as "low-stakes," but as these assessments are rather pristine examples of tests without consequences for students, we label these tests as "no-stakes" to emphasize their distinctiveness. In general, no-stakes international assessment tests have not been seen as crucial determinants of inequalities in educational

attainment. Although some studies have used PISA scores in assessments of inequality at educational transitions (e.g. Holm and Meier-Jaeger 2013), these studies work upon the assumption that the scores measure underlying academic ability, and that academic ability influences educational decisions, rather than the assumption that PISA scores will influence educational decisions directly.

In this article, we explicitly compare the contribution of high-stakes, low-stakes, and no-stakes tests to socioeconomic background inequality at educational transitions in Russia. Why might the comparison of tests shed light on the inequality-generating process? To address this question, we can consider two potential scenarios relating test scores to educational transitions.

First, it may be that there are pre-existing high levels of inequality in academic ability in a given country, and that once test scores are given the power to determine educational trajectories, these test-score inequalities are straightforwardly translated into inequalities in educational outcomes. It is well-known that there are substantial differences in academic ability among socioeconomic groups, so there is good reason to suppose that even absent other inequality-generating processes, inequalities in educational attainment would be found if academic-ability gaps determined educational outcomes (Reardon 2011; see Van de Werfhorst and Mijs 2010 for a review). Second, it may be that existing inequalities in academic ability are exaggerated in the context of a high-stakes testing regime. Perhaps students from high-socioeconomic backgrounds build on their advantages in academic ability by participating in tutoring, for example, and thereby further increase their advantage over students from disadvantaged backgrounds (e.g. Alon 2009; Loyalka and Zakharov 2016).

Both scenarios might be treated as specific examples of the more general phenomenon of socioeconomically advantaged families acting to secure their children's interests However, which scenario is in play does have implications for our understanding of the mechanisms producing inequality. The first scenario suggests that families are engaged primarily in *achievement-directed* investments, that is, investments directed toward developing general academic capacities of the type measured in achievement tests. Such investments would include good nutrition, healthcare, well-resourced schools, and extra support for learning (see Lee and Burkham 2002, Duncan and Magnuson 2005 for useful overviews). Achievement-directed investments will pay off in the event that achievement *per se* comes to determine a child's lifecourse. The second scenario is instead consistent with *test-directed* investments. Such investments are directed toward securing strong performance on a given high-stakes test, and under any rational model would be predicted to be precisely targeted. Rather than investing in the development of general academic capacities, families here focus on maximizing performance on a single test (or set of tests), either via careful selection of schools or via private tutoring. Under this scenario, establishing a high-stakes test would be consequential because it would create inequalities that would otherwise not exist in the context of a low-stakes test regime. It is for this reason that sociologists have expressed particular concern about the effects of raising the stakes of a test.

In the absence of a randomized controlled trial, there are two types of evidence that would allow us to examine whether the achievement-directed or test-directed investment scenario better characterizes how educational trajectories unfold within a given high-stakes system. First, a quasi-experimental study that compared inequality of educational opportunity (IEO) before and after the implementation of a high-stakes test would provide evidence on the causal effects of high-stakes tests on IEO.[5] Second, if data amenable to quasi-experimental analysis are not available, descriptive evidence might be exploited instead; this evidence can be interpreted in the light of theory to provide a plausible baseline account of the effect of high-stakes tests on IEO. This study, in common with other work in this field, employs the latter approach.

We exploit a longitudinal dataset that allows us to compare the extent to which the different types of tests can account for socioeconomic inequalities in educational opportunity in Russia. We decompose socioeconomic inequalities at different transitions to determine the contribution of high-, low-, and no-stakes tests: the total IEO is decomposed into a set of indirect effects of socioeconomic status for each of the tests and a direct effect.[6] Our analyses leverage variation across tests and transitions to address three key questions related to our central interest in whether sociologists of IEO should treat high-stakes tests as somehow "special."

We ask, first, how much of the observed inequality of educational opportunity can be accounted for by inequalities in performance on high-stakes tests. The comparison of the contribution of different tests to total IEO will establish whether or not high-stakes tests stand out, and thus whether or not test-directed investments should be targeted by those wishing to reduce IEO in the context of a high-stakes system.

Second, is there variation across educational transitions in the contribution of high, low-, and no-stakes tests to IEO? Substantial variation *across* levels in the contribution of high-stakes tests relative to other tests would support claims of "specialness," while substantial variation *within* levels (horizontal differentiation) would raise questions about any straightforward account of the role of high-stakes tests in generating IEO.[7]

And third, is there variation by gender in the contribution of high-, low-, and no-stakes tests to IEO? It is well-known that family investments vary across male and female children (see Buchmann et al. 2008 for a review of gender inequality in education). These inequalities are likely to be of particular consequence for test-directed investments, which are highly instrumental and directed toward a specific purpose rather than part of standard socialization practices. However, as we highlight below, an additional complication in the Russian case is that male-typed fields are more common in vocational than academic upper secondary and tertiary education, which might counteract parental concerns that females are less worthy of investment. How these competing factors unfold will have consequences for our conclusions with respect to the stakes of high-stakes tests. Indeed, if we see substantial gender variation in the contribution of different tests to IEO, this would provide useful evidence that the stakes attached to a high-stakes test are context-dependent and contingent on other factors.

The Russian Educational System

The Russian educational system has much in common with contemporary European stratified systems, although it diverges from the traditional Germanic model in allowing tracking only in upper secondary education. Students are required to study general (i.e. academic) education up until the end of lower secondary education (grade 9, age 14/15). They are then sorted into academic or vocational tracks for grades 10–11 in academic schools and in vocational schools for 2–4 years (depending on the vocational track). Almost 70% of all students attend academic upper secondary education, and just over 60% of students will attend university (full descriptive statistics can be found in the appendix, table A1). There is no formal test that regulates entry into upper secondary tracks, but students receive State Final Assessment (GIA) grades just prior to the transition and these grades may therefore play a role both in setting student expectations and in determining the type of counseling that teachers will provide. As there is a strong emphasis on the counseling-out of weak students, the tests at the first transition might be seen as low-stakes assessments, but they surely are far from being no-stakes. At the end of the 11th grade, all students in the upper secondary academic track must take the Unified State Examination (USE) in order to graduate. This is both an end of high school test required for graduation and the entrance exam for higher education institutions.[8]

The USE fits the standard criteria for a high-stakes test: it is required for both high school graduation and entry into university. Further, the USE is the foundation of qualitative differentiation within the Russian university system, as different universities use different USE-score cutoffs when admitting students. There are USE tests for each of the 14 subjects offered in secondary school, but it is the USE scores in Russian language and mathematics that are most important, as it is necessary to pass these tests to receive an academic secondary school diploma (the other subject tests are optional). Most universities require scores in three or four subjects, and universities have the right to set their own minimum required scores as long as these scores exceed the minimum score for matriculation, set by the Ministry of Education.

Despite the Soviet emphasis on equal access to education, and despite almost universal access to high school education, all research confirms that there were persistent social background inequalities in educational attainment in Russia over the past half century (e.g. Konstantinovskiy 2012; Kosyakova et al. 2016). Previous work has shown significant social background inequalities at all levels of education, with evidence of increasing inequality in schooling after the transition (Gerber and Hout 1995; Gerber 2000, 2007). Access to the upper secondary academic track and selective university—the most prestigious path through the education system—is strongly related to social class background (Gerber 2007; Prakhov 2016). On the basis of existing research it would not, therefore, be surprising to find substantial social background inequalities in access to the more prestigious academic tracks. Our main interest, though, is to compare the size of estimated indirect effects when high-stakes, low-stakes, and no-stakes tests are used as measures of performance. We thereby aim to assess

the extent to which raising the stakes of a test is associated with a larger indirect effect.

Data and Methods

Data

We use data drawn from the Russian panel study, "Trajectories in Education and Career" (TrEC). This study followed a representative sample of Russian students from 8th grade up through university and beyond. The TIMSS 2011 survey was the first wave of the study, and in the second wave, the same sample of students took part in PISA 2012. The third wave took place in fall 2013, from which we obtain information about secondary track placement, with follow-ups allowing us to record university attendance. Because the first wave of the panel study was the TIMSS survey, the sample was defined according to TIMSS procedures. Panel retention rates are high, with 90% of the original sample participating in the PISA wave, and 85% participating in the third wave. Attrition in the later waves appears to be more likely for students attending vocational schools (Malik 2019), but given that our analysis of later-wave data excludes such students by design, selective attrition is unlikely to bias the results.

We use three types of variables in our analysis: student test performance, family socioeconomic status, and educational attainment. The variables are defined as follows:

Performance: As described above, we compare the contribution of high-, low-, and no-stakes tests to educational inequality at the two transitions. In all, we have scores from seven different tests:

- High-stakes: USE scores in math and Russian (11th grade).
- Low-stakes: GIA grades in math and Russian (9th grade).
- No-stakes: TIMSS (8th grade) and PISA (9th grade). Separate scores for math (TIMSS and PISA) and reading (PISA).[9]

The different types of tests differ from one another aside from any "stakes" attached to them. The high- and low-stakes tests are designed to measure skills and knowledge taught within the school curriculum. Among the no-stakes tests, TIMSS is also curriculum-based, while PISA aims to capture, "what [students] can do with what they know" (OECD 2016: 25; Van de Werfhorst et al. 2010). Further, the tests are administered at different ages. Students in our sample first take TIMSS, followed by PISA and the low-stakes GIA tests, and finally a smaller number of the initial respondents take the USE (age 17/18).

Family socioeconomic status: We use a continuous measure of family socioeconomic status to assess inequalities at the transitions. Socioeconomic status is measured using the PISA ISEI scale (OECD 2014), with the family status determined by whichever parent has the higher status.[10]

Educational attainment: We assess inequalities at the two important transitions in the Russian educational system. First, we examine inequality in access to the academic track. We compare students who enter the academic track to those who enter the vocational track after lower secondary education. The vocational track encompasses two different types of vocational education, but given that Kosyakova et al. (2016) show similar patterns of inequality when comparing the academic track and the two vocational tracks, we combine the two vocational tracks here to preserve statistical power. At the second transition, we examine access to the university, conditional upon attending the academic upper secondary track. For those who attend university, we know whether they attended a selective or non-selective institution. In common with other research on the Russian system, we define selective universities as those that have average USE scores for enrolled students of higher than 70 (Dobryakova and Kuz'minov 2016); all other universities are coded as not selective. Note that the selectivity of each university is determined by reference to a standard set of selectivity codes that is published and therefore known to students; we do not calculate the average USE scores for each university from our sample of students.

Methods

We employ a decomposition analysis to establish the extent to which test score performance mediates the relationship between social origin and educational track. We use the method proposed by Karlson et al. (2012) to decompose socioeconomic inequalities in transition-taking into indirect and direct effects. An initial "reduced" model is fitted as follows:

$$logit(Y) = \vartheta + \delta * X + \rho * \tilde{Z} + \pi,$$

where Y is the educational transition (e.g. academic vs. vocational), ϑ is a constant, δ and ρ are coefficients, π is error, X is social origins, and $\tilde{Z}$ is the residual from a linear regression of test score performance on social origins. This model is subsequently compared to a "full" model:

$$logit(Y) = \alpha + \beta * X + \gamma * Z + \varepsilon,$$

where α is a constant, β and γ are coefficients, ε is error, X is social origins, and Z is test score performance. The direct effect of social origin on educational transition is then given by β/δ, while the indirect effect of social origin that operates through test score performance is the total effect of social origin on educational transition minus the direct effect. As many Z variables can be added as required.

In our analysis, we will decompose socioeconomic inequalities at the transitions with respect to the set of performance variables. An analysis that takes into account more than one measure of performance inevitably raises the specter of multicollinearity, that is, the effects on model estimation of including highly correlated variables. Multicollinearity may lead to large standard errors on

estimated effects, because there is insufficient power to precisely identify the separate effects of closely related variables (Goldberger 1991). In one sense, of course, multicollinearity is a feature rather than a bug of our analysis, in that our main research question implies a worry about social scientists assuming independent effects of high-stakes tests without asking whether these tests are correlated with other, lower-stakes, tests. But we take seriously the concern that we do not have enough power to adjudicate on the question of interest.

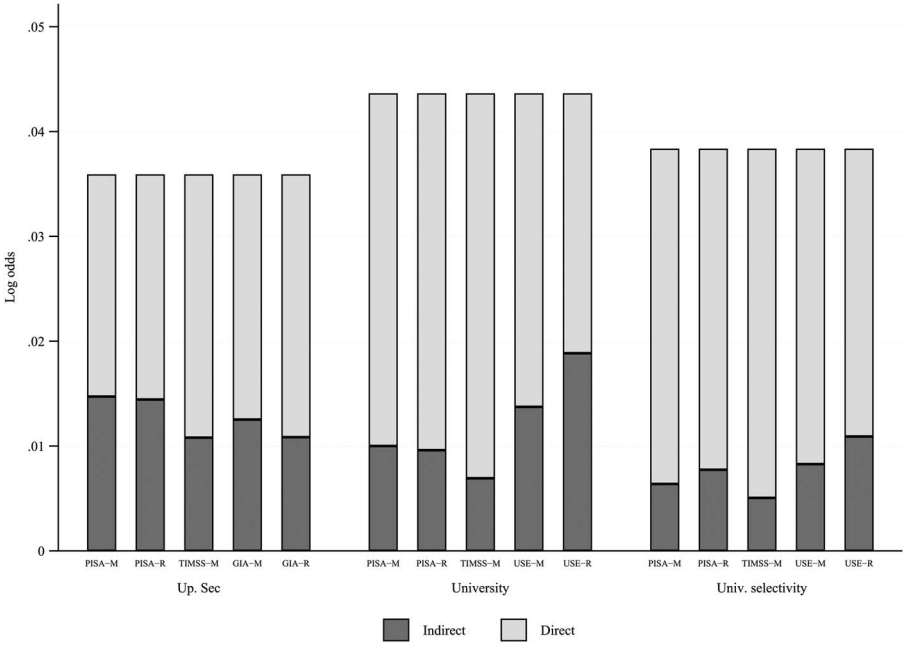
To assess the extent of multicollinearity, we examine the variance inflation factors (VIFs) in models predicting each of the educational transition outcomes as a function of all of the individual performance variables and family socioeconomic status (see online appendix, table A2). The VIFs range from 1.5 to 3.7, which implies that there is no serious multicollinearity problem in our analyses. Nevertheless, we take the precautionary step of including composite measures in analyses with more than one performance measure (i.e. the "net" effects analyses). We create composite measures for our no-stakes (PISA math, PISA reading, and TIMSS), low-stakes (GIA grades in math and Russian), and high-stakes tests (USE in math and Russian) by standardizing each variable and then taking the average score within each "stakes" group. Composite variables are an effective solution to the multicollinearity problem insofar as the variables to be combined represent similar theoretical constructs (O'Brien 2007: 683–684).

Results

We will examine socioeconomic inequalities at the two educational transitions, taking into account both the choice of track, and within the university track, the prestige of the university. Our analyses will examine the contribution of indirect effects to total inequality at the transitions, with a focus on three research questions in particular. First, to what extent can socioeconomic inequalities in high-stakes, low-stakes, and no-stakes test scores account for inequalities at each of the transitions? When we control for other tests, do high-stakes tests have more power in reproducing advantage, as measured through the indirect effects? Second, do we see across- and/or within-level variation in the explanatory power of high-, low-, and no-stakes tests? And third, are our conclusions altered when we consider other stratifying features of the Russian system? Although our research questions emphasize the indirect effects, we will also consider the size of the direct effects, that is, the effects of socioeconomic background that cannot be accounted for through performance on the tests. Examining the direct effects provides important context for our interpretation of the size of the indirect effects: sociologists and policy-makers have been concerned that high-stakes tests might exaggerate inequalities at educational transitions, and it is therefore essential to consider the magnitude of the estimated indirect effects for each type of test relative to the direct effects of social background.

We examine the contribution of the different tests to socioeconomic inequalities at the transition from lower to upper secondary academic/vocational education (9th grade), the transition to the university (11th grade), and for those

Figure 1. Estimated indirect and direct effects at the transition to upper secondary education, the transition to the university, and the selectivity of the university (for those who attend university)



who entered the university, the selectivity/prestige of the university. Given how common it is for sociologists to have access to but a partial set of the performance measures available to us here, we begin by presenting analyses that examine the estimated indirect effects for separate tests without controlling for other tests. Figure 1 therefore shows the indirect and direct effects of social background on each transition for each test separately (confidence intervals are not included in the figure, but we include full models with standard errors in the online appendix).[11]

As figure 1 shows, the effects of socioeconomic background are very similar across the transitions, with the log odds of making each transition increased by around 0.04 for each unit increase on the SEI scale. If we were to compare students originating in families at the 25th and 75th percentiles of SEI, we would obtain odds ratios of 2.59 for the lower to upper secondary transition, 2.98 for the transition to university, and 2.60 for the selectivity of the university.

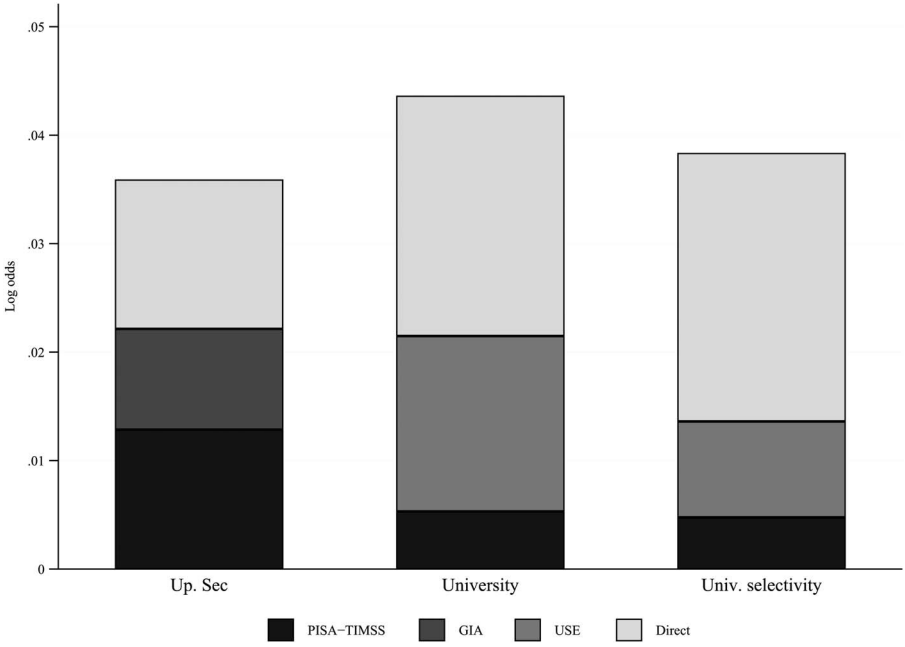
Turning to the contribution of the different performance measures to the total inequality at each transition, we may examine the size of the indirect and direct effects. The dark gray bars show the size of the indirect effects, which can be interpreted as the inequality that would be expected, all else equal, if only differences in test scores (or grades) were operating to produce inequality. The remainder—displayed in the light gray bars—is the direct effect

of socioeconomic background at each transition. It is clear that regardless of which performance measure is used, direct effects account for the majority of the inequality that we observe at each transition. At most, any single performance measure accounts for no more than 45% of inequality at the transition to upper secondary and the transition to university, and for no more than 30% of inequality in selective university attendance (conditional on attending university). As a consequence, it is also clear that if our interest was simply to assess the contribution of performance inequality to socioeconomic inequality at a given transition, with one or two notable exceptions our foremost conclusion would be very similar regardless of which performance measure was used; namely, that performance inequality does not account for the majority of inequality at any transition.

We turn now to consider the precise size of the estimated indirect effects across measures. At the transition to upper secondary education, figure 1 allows us to compare the no-stakes performance measures to the low-stakes GIA grades. Even if our general conclusions regarding the contribution of indirect effects to total inequality would be broadly the same across performance measures, there are some differences in the magnitude of these effects. The two PISA tests are particularly powerful mediators at this transition, accounting for a significantly higher level of total socioeconomic inequality than the other tests (the PISA reading and math tests account for 40% of the total inequality, while the other tests/grades account for around 30%). The GIA grades, which we categorize as "low-stakes" because the grades are known to students but do not play any formal role at this transition, mediate relatively little of the total effect of socioeconomic background.

We see more substantial differences among tests at the transition to the university, a transition that is formally linked to high-stakes tests. Here, we have measures of high-stakes and no-stakes tests, and at the extremes, we see that the no-stakes TIMSS scores produce the lowest estimated indirect effect (around 16% of total inequality) and that the high-stakes USE Russian scores produce the highest (at 43%).[12] If we had been employing a single measure of performance in our analysis, as is not uncommon in this field, we would here perhaps come to quite different conclusions if we happened to take either of these extreme cases. Further, we see that the two high-stakes tests (USE in math and Russian) account for a higher proportion of inequality than the three no-stakes tests, although it is only in the case of USE Russian that this difference is significant. The differences in the magnitude of indirect effects across high- and no-stakes tests might be taken as a first piece of evidence in support of the argument that high-stakes tests play a special role in the generation of inequalities in educational attainment. The same pattern is observed for qualitative differentiation at the university level: the high-stakes USE tests return higher indirect effects than the no-stakes PISA and TIMSS scores, although again, only some of these comparisons are significant.[13] Nevertheless, even when the high-stakes tests are used as measures, indirect effects account for less than a third of the inequality in access to selective universities.

Figure 2. The net contributions of the composite measures of performance, via indirect effects, and direct effects at the transition to upper secondary, the transition to the university, and the selectivity of the university (for those who attend university)



Bringing the results in figure 1 together, the fact that the absolute magnitude of indirect effects is greatest when high-stakes tests are used to measure performance provides support for the claim that high-stakes tests play a distinctive role in the transmission of inequality at educational transitions. But two additional findings must also be highlighted. First, even where high-stakes tests account for a higher proportion of the indirect effect of socioeconomic background than other types of tests, it is not clear that we would go too far astray if we had access only to low-stakes tests for analysis. And second, although high-stakes tests appear to be more important than low-stakes tests at the university level, the results caution against assuming that high-stakes tests play a similar role in determining *whether or not* a student attends university and *which* university the student attends. The USE is apparently more important in explaining inequalities in entry to university *per se* than in explaining entry to selective universities. A simple account of how the USE operates at the university level is therefore questionable, at least insofar as our interest is in the role of high-stakes tests in reproducing inequality.

Nothing that we have done so far speaks to whether high-stakes tests matter over and above other measures of performance, a question that our data leave us uniquely qualified to answer. In figure 2, therefore, we include all measures of performance, and calculate the size of the indirect and direct effects. To

preserve power and to guard against drawing the wrong conclusions because of multicollinearity, we create composite measures for the two PISA scores plus TIMSS, for the GIA grades in mathematics and Russian, and for the USE scores in mathematics and Russian. This allows us to compare the net indirect effects for the no-stakes tests, the low-stakes tests, and the high-stakes tests (i.e. the estimated indirect effects while controlling for the other scores). At the transition to the upper secondary level, we compare the indirect effects for the no-stakes PISA/TIMSS composite with those for the low-stakes GIA composite, and at the two university transitions, we switch the GIA composite for the high-stakes USE composite.[14]

It is clear that the estimates of direct effects are smaller when we employ the composite measures of performance in our analysis. In the analysis with single measures of performance, direct effects accounted for 55–85% of the estimated inequality at each transition. But when the composite measures are used, direct effects account for 40% of the inequality at the upper secondary transition, 50% of the inequality at the university transitions, and for 65% of the inequality in university selectivity. By implication, this means that the individual performance measures are far from perfectly correlated, and that additional information is obtained when all are included in the model.

When we examine the breakdown of the indirect effects across the different composite measures, we see a clear difference between the upper secondary transition and the two university transitions. The high-stakes USE composite accounts for just under 40% of the total inequality in university entrance, and for around a quarter of the inequality in access to selective universities. The no-stakes composite, on the other hand, accounts for under 15% of the total inequality for these outcomes. We do have evidence, then, that high-stakes USE tests capture a dimension of performance that is not measured in the no-stakes tests. But there are three conclusions that might also be drawn from figure 2 that cast the "stakes" of high-stakes tests in a more questionable light, at least as regards their role in generating educational inequality.

First, even if the high-stakes composites account for a substantial portion of socioeconomic inequality for the two university outcomes, the majority of inequality cannot be accounted for by these test scores. Even the 25–40% of inequality that is accounted for by high-stakes tests at the two university transitions takes on a different light when compared with the analysis for the transition to upper secondary, where the low-stakes GIA grades also account for around 25% of socioeconomic inequality. Second, the high-stakes USE scores are substantially more high-stakes for the university attendance outcome than for the university selectivity outcome. On the basis of these results, it would be wrong to assume that high-stakes tests have the same relationship with all forms of inequality in outcome arising at a given level. Finally, one advantage of our analysis is that we can evaluate socioeconomic inequality in educational attainment across the entire system. Although there is a temptation to focus on the importance of high-stakes tests and transitions when they are described as such, in absolute terms it is still the first transition that filters out the most students. From a whole system perspective, a larger number of low-SEI students

are pushed out of academic education by the low-stakes GIA grades than by the high-stakes USE scores.

The analyses to this point have examined socioeconomic inequalities at the transitions for the whole sample. We have argued that the stakes associated with high-stakes tests are considerable, with the USE accounting for a substantial portion of the inequality observed at the transition to university. Nevertheless, the evidence also indicates that sociologists of educational inequality must be mindful both of the inequality that can be attributed to low-stakes tests, and of the direct effects of socioeconomic background, in determining whether or not high-stakes tests are in some larger sense "special."

Gender

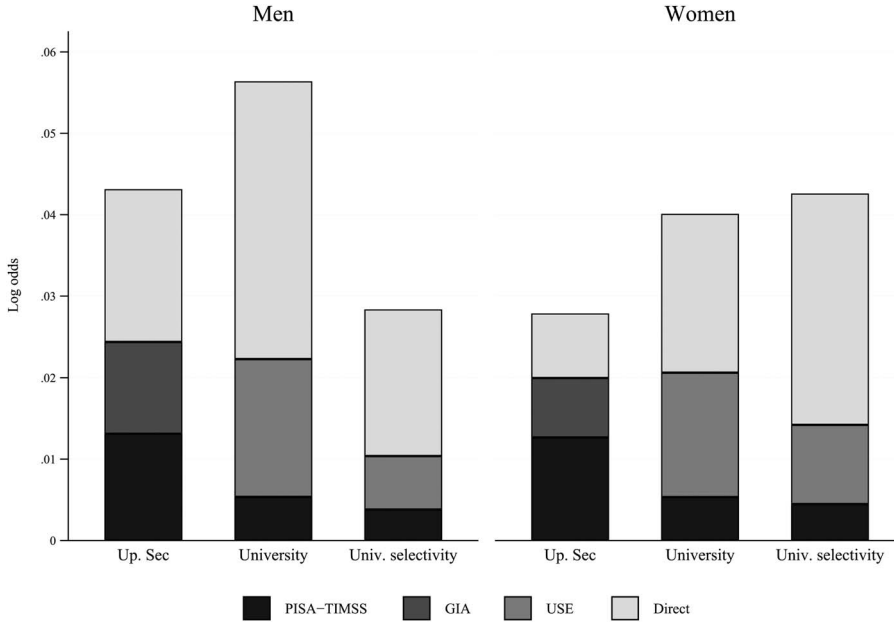
The Russian educational system is strongly gendered. Women are over two and a half times more likely to be found in the academic track than the vocational track, while for men this ratio is closer to one and a half (see table A1). Given our concerns that families might make different decisions with respect to investments in the education of males and females, we now examine socioeconomic inequalities at the transitions for men and women separately, with a focus on the role of high-stakes, low-stakes, and no-stakes tests at these transitions. If the contributions of the different stakes tests differ substantially between males and females, we might conclude that the stakes of high-stakes tests are higher for some groups than for others.

In figure 3, we show the breakdown of the indirect effects across the composite performance measures for men and women separately. Total socioeconomic inequality exhibits a different pattern for men and women across the transitions: inequality is highest for men at the transition to university, and lowest with respect to university selectivity, while socioeconomic inequality is low for women at the first transition, and higher for the two university outcomes. Do high-stakes tests also play a different role for men and women?

Examining the contribution of the different test scores to the total indirect effects, we see a similar pattern across men and women, and indeed, for both men and women we see the same pattern as previously. The high-stakes tests are clearly the most important component of the indirect effect at the transition to university, but they have significantly less influence when it comes to the selectivity of the university. Notably, for both men and women, the low-stakes GIA grades account for around the same amount of inequality at the transition to upper secondary as the high-stakes tests account for with respect to university selectivity. In sum, the analysis by gender gives no reason to question the basic conclusions reached on the basis of the full sample with respect to the role of high-, low-, and no-stakes tests in the Russian educational system: the stakes of high-stakes tests are invariant across genders.

Although the absolute size of the indirect effects is very similar across men and women, the same conclusions cannot be drawn about the absolute size of the direct effects. At the transition to upper secondary, the indirect effects are similar

Figure 3. The net contributions of the composite measures of performance, via indirect effects, and direct effects at the transition to upper secondary, the transition to the university, and the selectivity of the university (for those who attend university); men and women separately



across men and women, but the direct effects are much more substantial for men. For the university selectivity outcome, we now observe modest direct effects for men, but more substantial direct effects for women, with about two-thirds of total inequality coming from sources unrelated to the performance measures included here. In both cases, the higher level of inequality at each transition is observed because the direct effects are higher. The substantive meaning of this is clear: although the "stakes" of the tests are similar for men and women, insofar as inequalities differ across men and women, it is because students are making educational decisions on the basis of something uncorrelated with their test score performance.

At the upper secondary level, the most obvious explanation for the difference in the magnitude of direct effects is the highly gendered nature of the educational tracks on offer. Although women dominate upper secondary vocational fields in the arts, music, office administration, and textile work, men dominate the much larger fields of manufacturing, construction, and agriculture. An explanation rooted in gendered fields of vocational study is highly consistent with the pattern of the indirect and direct effects at this transition: test scores are relatively less important in explaining inequality for women than for men at the upper secondary level because low-SEI men are easily diverted from academic to vocational education, while low-SEI women have few good vocational options to pull them away from the academic path. As regards university selectivity, the

gender differences are again likely to be explained by the nature of the choices on offer, although in this case the constraint relates to geography. All of the highly selective universities in Russia are located in urban areas, and low-SEI women from rural areas are less likely to travel far from home to attend such institutions (see Hohlushkina 2001).

Discussion

In this article, we have examined socioeconomic inequalities at the key transition points in the Russian system, and have asked whether there is evidence that high-stakes tests play a special role in producing inequalities in transition-taking. In contrast to other work on high-stakes tests and educational inequality, we have explicitly compared the contribution of high-stakes tests alongside the contribution of low-stakes and no-stakes tests.

Sociologists of educational inequality might be concerned about high-stakes testing for two important reasons. First, students of differing socioeconomic background have different levels of measured academic ability, and when tests of this ability come to determine educational outcomes this ability inequality will be transferred to inequalities in transition-taking and attainment of qualifications. Second, students of differing socioeconomic background differ in their capacity to engage in purposive behavior with respect to the tests. While advantaged students can undertake test-oriented investment behaviors that will allow them to perform well—engaging tutors and test preparation services, for example—disadvantaged students have much less capacity to do the same. Thus, existing inequalities in academic ability between groups will be exaggerated in the context of high-stakes tests. Most of the discussion of high-stakes tests in sociology appears to be more concerned with inequalities in test-oriented investments than with pre-existing inequalities in academic ability.

Which scenario is most consistent with the evidence from Russia? In addressing this question, three findings should be highlighted. First, inequalities in the academic ability measured in no-stakes tests account for around a third of the inequality at the upper secondary transition, and just over a tenth of the inequality at the university transitions. If the academic ability measured in PISA and TIMSS were to be the sole basis of selection in the Russian system, all else equal we might expect to see smaller inequalities than we do at present. Second, from the analyses of net effects, we see that high-stakes tests do play a special role, over and above the academic ability measured in the other tests, at the transition to university. This finding leaves the door open for an explanation focused on purposive investments in high-stakes test performance. But a third important result of our analyses is the finding that regardless of which mechanisms are responsible for producing inequalities in tests of all stakes, a substantial amount of the inequality that we observe at the transitions is due to characteristics uncorrelated with performance (i.e., the direct effects). Indeed, the majority of total inequality at any transition is accounted for by factors other than high-stakes test performance. Further, where IEO differs across men and women, these

differences arise because of differences in the size of the direct effects. These findings suggest that any focus on achievement- and test-oriented investments should be accompanied by a detailed discussion of the "attainment-oriented" mechanisms that can account for that part of IEO that is *not* due to test-score performance.

One unexpected finding is the role that high-stakes tests play in relation to the university transition. When sociologists consider educational decisions, we recognize that students make decisions about both the level of education to pursue, and, within that level, about the type of education to pursue (see particularly, Lucas 2001; Alon 2009; Lucas and Byrne 2017). When we ask how high-stakes tests feed into these educational decisions, a baseline assumption might be that the tests will play a similar role in both the quantitative and qualitative aspects of educational choice, because test-oriented investments are presumably made in anticipation of a particular outcome. But one of the robust findings of this analysis has been that the high-stakes tests account for a much higher proportion of inequality with respect to the decision to attend university than with respect to the type of university to attend. This suggests, in turn, that our theories of educational inequality should be able to account for why test performance may matter more for "level" decisions than for "track" decisions (or vice versa, depending on context). It is easy to imagine a scenario, for example, in which a student of disadvantaged background may be pushed into attending university by strong test performance, but that the same student may attend a less selective university because of worries about cultural "fit" or cost. In many countries, including Russia, selective universities are based in only a small number of geographical locations, and the costs associated with moving to those locations may be prohibitive for poorer students. Geography may also be part of the explanation for why we see larger direct effects on selective university attendance for female students. Conversely, it is easy to imagine contexts in which university may be seen as a natural next step, regardless of test score performance, and in these contexts we would expect performance to matter much more for choices about track/selectivity than for choices about level (as appears to be the case in the U.S.; see Alon 2009; Alon and Tienda 2007).

It is important that our analyses do not speak to the value of high-stakes tests at educational transitions in all countries, or at all points in time. The role of high-stakes tests is likely to differ significantly across different institutional configurations. Where early transitions and tracking exert large selective effects on the types of students facing later transitions, we should expect to see a very different relationship between test score performance and educational decisions than where no selection on (observed and unobserved) student characteristics has previously occurred. Further, certain institutional characteristics will likely be associated with strong indirect effects, such as when governments mandate the strict use of high-stakes tests at educational transitions (e.g. Jackson and Buckner 2016). Our results do not, then, imply that high-stakes tests will never play a determinative role in the generation of inequality.

Another aspect that should be considered when tackling the importance of high-stakes tests is the nature of the tests themselves. One high-stakes test

is not the same as another, and there will be variation across countries and over time in the extent to which class-biased tests are employed. It is perhaps no accident that the Russian language USE is more powerful in explaining socioeconomic inequalities in access to university than the USE in mathematics, given that in common with other countries, tests of language and reading are more likely to be class-biased than tests of mathematical ability (Bourdieu and Passeron 1977). Similarly, the relatively weak explanatory power of TIMSS as compared to PISA in our analysis is consistent with a potential class bias in the PISA tests. In the Russian case, there is evidence that PISA scores are more highly correlated with family characteristics than with school characteristics, while the opposite is true of TIMSS scores (Carnoy et al. 2015). This highlights the importance of examining the details of the tests and the institutional structure within which they operate before assuming that high-stakes tests have negative consequences for inequality of educational opportunity.

Although our work cannot assess the causal effect of high-stakes tests on educational inequality, our analysis does speak to concerns that the introduction of high-stakes tests might be inequality-increasing, by presenting a plausible upper bound estimate of the contribution of high-stakes tests to IEO in Russia. We believe our estimate to be an upper bound because our analysis is still vulnerable to the time bias introduced by measuring performance on high-stakes tests closer to the transition than performance on low-stakes tests. And given the possibility that advantaged parents may circumvent the USE altogether—via foreign education, private university slots, and nonstandard paths—the total influence of performance on overall IEO is likely exaggerated. But an upper bound estimate may still be useful to policy-makers, particularly if those policy-makers aim to introduce high-stakes tests with the intention of reducing inequality of educational opportunity. If the primary intention is to reduce IEO, policy-makers might consider providing less advantaged students with access to the same type of out-of-school tuition programs that their advantaged peers now benefit from. More radical solutions might include lower test-score cutoffs for low-SEI students during the university admission process. Such policies would help to minimize any inequality-increasing effects of high-stakes tests while satisfying the political aim of giving greater rewards to those judged talented.

In considering the causal effect of high-stakes tests on educational inequality, one troubling alternative hypothesis might be that these tests guide student decisions in a way that other test scores would not, should the high-stakes tests be absent. At an extreme, perhaps students would not use performance-related characteristics to guide their educational decisions at all in the absence of the high-stakes tests, and the indirect effects would fall to zero across all measures. Although we agree that it is likely that high-stakes tests provide important guidance to students about their academic potential, this fact in and of itself does not imply that academic ability would fail to enter into educational decisions unless measured in a high-stakes test. Indeed, the evidence from the transition at the end of lower secondary education demonstrates that a substantial portion of inequality at this transition can be accounted for by differences in students'

performance as measured in low- and no-stakes tests. Further evidence from other countries should also lead us to be deeply skeptical that indirect effects would disappear were it not for the presence of a high-stakes test: the indirect effects estimated for Russia's high-stakes transition are in fact slightly lower than the indirect effects estimated on a sample of European countries representing a diverse range of testing regimes (Jackson and Jonsson 2013).

The comparison of indirect effects across high-, low-, and no-stakes tests is an essential step before high-stakes tests are assigned a special role in the inequality-generating process. Our analyses have shown that—in the Russian case—our broad assessment of the role that academic performance plays in educational decision-making would change little if we were to focus on the inequalities in performance captured by low- and no-stakes tests rather than on the inequalities that we observe in high-stakes tests. Although it is doubtless important to establish whether or not performance inequalities are magnified in a high-stakes context, as they indeed appear to be in Russia, insofar as we worry about the extent to which performance mediates inequalities in educational outcomes, it is pre-existing inequalities in academic ability, and the inequalities that enter in through the choices that students make at given transitions, that are potentially the greater cause for concern.

Notes

1. Discussions of high-stakes tests and inequality of educational opportunity are also found in work on the stratification of educational systems. The "stratification" of an educational system refers to its differentiation into tracks and levels (see Kerckhoff 1995). As highly stratified systems are very likely to use high-stakes tests and examinations to determine track placement, cross-national work on stratification and educational inequality frequently discusses mechanisms related to high-stakes tests in explaining why highly stratified systems are also highly unequal. The body of work on the relationship between high-stakes tests and educational inequality is therefore substantially greater than might be assumed on the basis of a cursory review of the "high-stakes testing" literature.
2. Note that in our review of the literature we treat the SAT in the United States as a high-stakes test, even if some authors explicitly disqualify the SAT from the high-stakes definition.
3. For example, in almost all countries it is possible to drop down to a less prestigious level of education, if the student chooses to do so. Private education may allow parents to buy direct access to a level or track for their child.
4. In some countries, such as the United States, an additional layer of testing exists to assess student progress, teachers, and schools (e.g. NCLB). These tests are again "no-stakes" for students, in that they play no role at educational transitions, but they may be important features of the educational landscape in the affected countries.

5. To our knowledge, there are no quasi-experimental studies of the independent effect of high-stakes tests on IEO. No doubt this can be largely attributed to a lack of appropriate data. But these studies are also missing from the literature because high-stakes tests are almost always introduced within a country as part of a larger package of reforms. Although there have been recent attempts to assess the causal effects of the Russian reform package (e.g. Francesconi et al., 2017), data and identification problems make it impossible to assess the independent effect of high-stakes testing on IEO.
6. Indirect and direct effects are also known as "primary" and "secondary" effects in the literature on IEO (Boudon 1974; Jackson 2013; cf. Morgan 2012).
7. One objection to our strategy here might be that the existence of a high-stakes test could perhaps create greater test score inequality across *all* performance tests within a country, no matter the stakes, and thus that our approach will provide an underestimate of the contribution of high-stakes tests to inequality at transitions. Research in fact shows the opposite relationship to be in place: countries with central examination systems (e.g. high-stakes tests) have lower levels of PISA test-score inequality than countries without such systems (Bol et al. 2014). Our analysis is therefore a conservative test of the hypothesis that the importance of high-stakes tests for educational inequality might be exaggerated.
8. Students who plan to attend a *College* after the academic track are not required to submit USE scores, and are simply required to pass the examination. Students in the vocational track are not required to take the USE, even if they later transfer to university. Although students who take a "hybrid" path still eventually gain access to a university, the evidence suggests that this path offers rather lower rewards than the standard path (see Yastrebov et al. 2018)
9. Note that TIMSS and PISA scores are provided as plausible values rather than as unique scores for each student on each test (PISA 2014). Following standard practice, we use *Stata*'s "pv" command to account for plausible values (Laukaityte and Wiberg 2017; Braun and von Davier 2017). When we create composites, we use the average of the plausible values for each of the tests. Although averaging plausible values can generate bias, we in fact see extremely similar results for analyses using plausible values and analyses using averages of the plausible values.
10. We considered using parental education as a measure of socioeconomic status, but as the average level of educational attainment is rather high in Russia, and overqualification is thus common, we settled upon SEI as a single omnibus measure of social background.
11. The KHB method does not allow us to compare the size of indirect effects across non-nested models. To calculate whether differences in the size of the indirect effects estimated using different performance measures are significant, we use bootstrapping with 1000 replications.
12. The TIMSS scores were the earliest performance measures, with the tests taken in 8th grade, so we cannot rule out that the small indirect effects

observed when TIMSS is used as the performance measure result from earlier measurement rather than weaker predictive power *per se*.

13. The indirect effect measured using USE math is not significantly higher than that estimated using the PISA reading and math measures.
14. We exclude the GIA grades from the university analyses because the indirect effects with respect to the GIA composite are approximately zero for both of the university outcomes.

Author Biographies

Michelle Jackson is an assistant professor in the Department of Sociology, Stanford University, and during the 2018–19 academic year was an Andrew W. Mellon Foundation Fellow at the Center for Advanced Study in the Behavioral Sciences, Stanford University. Her research focuses on educational inequality, social inequality and mobility, and statistical methods.

Tatiana Khavenson is a Research Fellow at the Institute of Education, Higher School of Economics, Moscow. Tatiana's research specializes in inequality in education, with a focus on the social structures that influence students' educational trajectories. Her recent publications include: Khavenson (2018) "Postsocialist Transformations, Everyday School Life, and Country Performance in PISA: Analysis of Curriculum Education Reform in Latvia and Estonia," in Chankseliani and Silova (eds.) *Comparing Post-Socialist Transformations: Purposes, Policies, and Practices in Education*. Oxford: Symposium Books.

Tatiana Chirkina is a Junior Research Fellow at the Institute of Education, Higher School of Economics, Moscow. Her research interests include educational inequality, social mobility and the factors associated with success for students of disadvantaged social background.

Supplementary Material

Supplementary material is available at *Social Forces* online, http://sf.oxfordjournals.org/.

References

Alon, S. 2009 The evolution of class inequality in higher education: Competition, exclusion, and adaptation. *American Sociological Review*, 74, 731–755.

Alon, S. and Tienda, M. 2007 Diversity, opportunity, and the shifting meritocracy in higher education. *American Sociological Review*, 72, 487–511.

Bol, T., Witschge, J., Van de Werfhorst, H.G. and Dronkers, J. 2014 Curricular tracking and central examinations: Counterbalancing the impact of social background on student achievement in 36 countries. *Social Forces*, 92, 1545–1572.

Boudon, R. 1974 *Education, Opportunity, and Social Inequality*. New York: Wiley.

Bourdieu, P. and Passeron, J.-C. 1977 *Reproduction*. London: Sage.

Braun, H. and von Davier, M. 2017 The use of test scores from large-scale assessment surveys: Psychometric and statistical considerations. *Large-scale Assessments in Education*, 5, 17.

Buchmann, C., Condron, D.J. and Roscigno, V.J. 2010 Shadow education, American style: Test preparation, the SAT and college enrollment. *Social Forces*, 89, 435–461.

Buchmann, C., DiPrete, T.A. and McDaniel, A. 2008 Gender inequalities in education. *Annual Review of Sociology*, 34, 319–337.

Carnoy, M., Khavenson, T. and Ivanova, A. 2015 Using TIMSS and PISA results to inform educational policy: A study of Russia and its neighbours. *Compare: A Journal of Comparative and International Education*, 45, 248–271.

Dobryakova, M.S. and Kuz'minov, YA.I. (eds.), 2016, Kachestvo priyema v rossiyskiye vuzy 2015. Moskva: ID VSHE.

Duncan, G.J. and Magnuson, K.A. 2005 Can family socioeconomic resources account for racial and ethnic test score gaps? *The Future of Children*, 15, 35–54.

Gerber, T.P. 2000 Educational stratification in contemporary Russia: Stability and change in the face of economic and institutional crisis. *Sociology of Education*, 73, 219–246.

Gerber, T.P. 2007 Russia: Stratification in postsecondary education since the Second World War. In Shavit, Y., Arum, R., Gamoran, A. (eds.), *Stratification in Higher Education: A Comparative Study*, pp. 294–321. Stanford, CA: Stanford University Press.

Gerber, T.P. and Hout, M. 1995 Educational stratification in Russia during the Soviet period. *The American Journal of Sociology*, 101, 611–660.

Goldberger, A.S. 1991 *A Course in Econometrics*. Cambridge, MA: Harvard University Press.

Heubert, J.P. and Hauser, R.M. 1999 *High Stakes: Testing for Tracking, Promotion, and Graduation*. Washington, DC: National Academies Press.

Hohlushkina, F. A. 2001. Выпускники учебных заведений среднего специального образования. In Г. А. Чередниченко (Ed.), Когда наступает время выбора (Устремления молодежи и первые шаги после окончания учебных заведений). СПб: Изд-во ПХГИ.

Holm, A. and Meier-Jaeger, M. 2013 Dentist, driver, or dropout? Family background and secondary education choices in Denmark. In Jackson, M. (ed.), *Determined to Succeed? Performance versus Choice in Educational Attainment*, pp. 228–252. Stanford: Stanford University Press.

Jackson, M. 2013 *Determined to Succeed? Performance versus Choice in Educational Attainment*. Stanford: Stanford University Press.

Jackson, M. and Buckner, E. 2016 Opportunity without equity: Educational inequality and constitutional protections in Egypt. *Sociological Science*, 3, 730–756.

Jackson, M. and Jonsson, J.O. 2013 Why does inequality of educational opportunity vary across countries? Primary and secondary effects in comparative context. In Jackson, M. (ed.), *Determined to Succeed? Performance versus Choice in Educational Attainment*, pp. 306–337. Stanford: Stanford University Press.

Kamens, D. H., and C. L. McNeely. 2010. Globalization and the growth of international educational testing and national assessment. *Comparative Education Review*, 54(1):5–25.

Karlson, K. B., A. Holm, and R. Breen. 2012. Comparing regression coefficients between same-sample nested models using logit and probit: A new method. *Sociological Methodology*, 42(1): 286–313.

Kerckhoff, A.C. 1995 Institutional arrangements and stratification processes in industrial societies. *Annual Review of Sociology*, 15, 323–347.

Konstantinovskiy, D.L. 2012 Social inequality and access to higher education in Russia. *European Journal of Education*, 47, 9–24.

Kosyakova, Y., Yastrebov, G., Yanbarisova, D. and Kurakin, D. 2016 The reproduction of social inequality within the Russian educational system. In Blossfeld, H.-P., Buchholz, S., Skopek, J., Triventi, M. (eds.), *Models of Secondary Education and Social Inequality: An International Comparison*, pp. 323–342. Cheltenham, UK: Edward Elgar.

Laukaityte, I. and Wiberg, M. 2017 Using plausible values in secondary analysis in large-scale assessments. *Communications in Statistics-Theory and Methods*, 46, 11341–11357.

Lee, V.E. and Burkam, D.T. 2002 *Inequality at the Starting Gate: Social Background Differences in Achievement as Children begin School, 20036*. 1660 L Street, NW, Suite 1200, Washington, DC: Economic Policy Institute.

Loyalka, P. and Zakharov, A. 2016 Does shadow education help students prepare for college? Evidence from Russia. *International Journal of Educational Development*, 49, 22–30.

Lucas, S.R. 2001 Effectively maintained inequality: Educational transitions and social background. *The American Journal of Sociology*, 106, 1642–1690.

Lucas, S. R., and D. Byrne. 2017. Seven principles for assessing effectively maintained inequality. *American Behavioral Scientist*, 61(1):132–60.

Malik, V. 2019 The Russian panel study 'Trajectories in Education and Careers'. *Longitudinal and Life Course Studies*, 10, 125–144.

Morgan, S. 2012 Models of college entry in the United States and the challenges of estimating primary and secondary effects. *Sociological Methods & Research*, 41, 17–56.

O'Brien, R.M. 2007 A caution regarding rules of thumb for variance inflation factors. *Quality and Quantity*, 41, 673–690.

OECD 2014 *PISA 2012 Technical Report - OECD*. Paris: OECD Publishing.

OECD 2016 *PISA 2015 Results (Volume I). Excellence and Equity in Education*. Paris: OECD Publishing.

Prakhov, I. 2016. The barriers of access to selective universities in Russia. *Higher Education Quarterly*, 70(2):170–99.

Ramirez, F.O., Schofer, E., and Meyer, J. W. 2018. International Tests, National Assessments, and Educational Development (1970–2012). *Comparative Education Review*, 62, 344–364.

Reardon, S.F. 2011 The widening academic achievement gap between the rich and the poor: New evidence and possible explanations. In Murnane, R.J., Duncan, G.J. (eds.), *Whither Opportunity? Rising Inequality and the Uncertain Life Chances of Low-Income Children*, pp. 91–116. New York City, NY: Russell Sage Foundation Press.

Francesconi, M., Slonimczyk, F., and Yurko, A.V. 2017. Moving on Up for High School Graduates in Russia: The Consequences of the Unified State Exam Reform (SSRN Scholarly Paper No. ID 2961757). Retrieved from Social Science Research Network website: https://papers.ssrn.com/abstract=2961757.

Suen, H.K. and Yu, L. 2006 Chronic consequences of high-stakes testing? Lessons from the Chinese civil service exam. *Comparative Education Review*, 50, 46–65.

Van de, W., Herman, G. and Mijs, J.J.B. 2010 Achievement inequality and the institutional structure of educational systems: A comparative perspective. *Annual Review of Sociology*, 36, 407–428.

Yastrebov, G., Kosyakova, Y. and Kurakin, D. 2018 Slipping past the test: Heterogeneous effects of social background in the context of inconsistent selection mechanisms in higher education. *Sociology of Education*, 91, 224–241.