I Z A Institute
of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 12338

Math Scores in High Stakes Grades

Giorgio Brunello
David Kiss

MAY 2019

DISCUSSION PAPER SERIES

IZA DP No. 12338

Math Scores in High Stakes Grades

Giorgio Brunello
University of Padova and IZA

David Kiss
Leibniz University Hannover

MAY 2019

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany
Phone: +49-228-3894-0
Email: publications@iza.org
www.iza.org

ABSTRACT

Math Scores in High Stakes Grades*

We investigate whether tests taken during a high stakes grade by German primary and secondary students produce higher math scores than in lower stakes grades. We identify a high stakes grade with the final grade of primary or secondary school, because good performance in that grade can affect future opportunities. Our difference-in-differences estimates show that high stakes increase math scores on average by 0.17 to 0.23 standard deviations, a sizeable effect.

JEL Classification: I26, J24, D91

Keywords: high stakes testing, student motivation, achievement, (perceived) returns to education

Corresponding author:
Giorgio Brunello
Department of Economics and Management "M. Fanno"
University of Padova
via del Santo 33
35123 Padova
Italy
E-mail: giorgio.brunello@unipd.it

* We wish to thank Murat Kirdar, Lorenzo Rocco and seminar participants at the University of Padova and Bogazici University for very valuable comments and suggestions. This project is funded by the German Research Foundation (DFG), Priority Programme SPP1646, grant number KI 1948/1-1, which is gratefully acknowledged. It uses data from the National Educational Panel Study (10.5157/NEPS:SC2:7.0.0 and 10.5157/NEPS:SC3:8.0.0), that were collected as part of the Framework Program for the Promotion of Empirical Educational Research funded by the German Federal Ministry of Education and Research (BMBF). Since 2014, NEPS is managed by the Leibniz Institute for Educational Trajectories (LIfBi) at the University of Bamberg in cooperation with a nationwide network.

1 Introduction

It is well recognized that student outcomes depend on the effort of students, parents, peers, and teachers, and that these agents respond to incentives - including school vouchers (Rouse 1998, Behrman et al. 2016), performance-based teacher compensation (Glewwe et al. 2010, Duflo et al. 2012), school accountability systems (Figlio and Loeb 2011), central examinations (Jürges et al. 2005), cash rewards (Angrist and Lavy 2009, Dearden et al. 2009), and social norms in the class (Bursztyn and Jensen 2015).

There is also a growing awareness that school test scores reflect not only ability, knowledge and intelligence but also personality traits, motivation and incentives. Test takers may not exert maximal effort. When tests are low stakes, as in the OECD PISA project, some individuals try harder than others (see Duckworth et al. 2011). Scores can also be improved by rewarding students (see Borghans et al. 2008, Segal 2012).

We add to this literature by studying the effect of high stakes conditions on math test scores in the specific context of German primary and secondary schools. In Germany, students at the end of primary education are tracked into different types of secondary education (or tracks), some leading to college education and some others ending up in vocational training. Since selection into tracks often depends on school performance in the final grade of primary school, tests taken in that grade are likely to be affected by the higher incentives to perform well than in previous grades. In a similar fashion, students in the final grade of secondary education have much more at stake than students in earlier grades, because their school results in that grade can affect the transition to the labor market and access to vocational training and higher education.

We exploit two institutional features of the German system: first, the final grade of primary education varies across states (Lander) for students of the same age. In most German states, primary school ends after the fourth grade. In Brandenburg, however, primary education lasts six years. Stakes in the fourth grade are therefore much lower in Brandenburg than elsewhere. Second, the ninth grade is the final grade in lower-tier secondary schools (Hauptschule) but not in middle-tier secondary schools (Realschule), which end instead in the tenth grade. Conditional on age and grade, students enrolled in the former type of schools who take the test in the ninth grade have higher stakes than students enrolled in the latter type.

We use data on repeated assessments of students across different German states and a difference-in-differences technique to evaluate whether high stakes grades in primary and secondary schools affect test scores. We find that math scores in primary and secondary education are 0.23 and 0.17 standard deviations higher when tests are taken in a high stakes grade than when they are not. These are substantial effects, given that, as a rule of thumb, student scores typically increase by 0.25 to 0.33 standard deviations during a school year (Wößmann 2016).

While there are a few studies that find a positive association between student outcomes and final exam periods (Winfield 1990, Frederiksen 1994, Bishop 1998, Neill and Gayler 2001), we are aware of only two papers (Jacob 2005, Federivcova and Munich 2017) that estimate the average treatment effects of high stakes on test scores by using a difference-in-differences framework similar to ours. Both papers find that high stakes testing boosts math achievement.

On the one hand, Jacob (2005) compares the mean math achievement levels of students living in US districts where high stakes testing was on (Chicago) to mid-west districts where stakes were lower, and finds positive effects of high stakes testing (around 0.3 s.d.). His estimates are larger than ours, probably because they capture the combined impact of high stakes on students, teachers and schools, that could be sanctioned in the case of poor performance. On the other hand, Federivcova and Munich (2017) use TIMMS data for Slovak primary students and report that admission exams near the end of primary education, when there is much at stake, increase student math scores by 0.2 standard deviations, similar to our findings. Compared to this study, our investigation explicitly tests for the presence of parallel trends and uses longitudinal rather than repeated cross section data.

The paper proceeds as follows. The next section introduces the data. Section 3 develops the empirical strategy by discussing the relationship between stakes and timing of assessments, defining the treatment and control groups, and introducing the DiD (difference-in-differences) framework. Results are discussed in Section 4. Section 5 concludes.

2 Data

Our data on math test scores are drawn from the National Educational Panel Study (NEPS), Starting Cohorts 2 and 3, a representative panel of German students. Students were interviewed and tested for the first time in the falls of 2010, upon enrollment in primary education (Starting Cohort 2) or at the beginning of secondary school (Starting Cohort 3). In either case, educational progress was tracked using follow-up assessments.[1] Even though panel participants in NEPS were interviewed each year, math scores were assessed irregularly, as shown in Table 1: primary school students were tested in grades 1 ($t = 1$), 2 ($t = 2$), and 4 ($t = 3$), and secondary school students were assessed in grades 5 ($t = 1$), 7 ($t = 2$) and 9 ($t = 3$).

We standardize math scores by grade using all the students in the original data. We then select our working sample as follows: for reasons given in Section 3, the primary school sample only includes students from Brandenburg, Bavaria, and North Rhine-Westphalia. Our secondary school sample comprises all students enrolled either in the lower track (Hauptschule) or the middle track (Realschule).

Table 2a presents, for the primary school students in the working sample, the summary statistics of the variables used in this study, by grade. Average math scores range from 0.00 to 0.07, average age ranges from 7.80 to 10.80, and average math scores in the initial test (grade 1) are close to 0. For the majority of students, the highest educational attainment of parents is tertiary education (61%), followed by vocational training (23%). There are 2760 students in the first grade and 2458

[1] See Bela et al. (2012) for details and Blossfeld et al. (2011) for a general introduction to NEPS.

students in the fourth grade.

As shown by Table 2b, students in the secondary school sample have below-average math skills. The reason is that our working sample includes only those enrolled in the lower (Hauptschule) and middle (Realschule) tracks, thereby excluding the upper and higher performing track (Gymnasium). The effects of this selection show up also in the average education of parents, about 60 percent of whom have completed at most vocational training. The number of observations is equal to 1906 in grade 5, 2232 in grade 7 and 1675 in grade 9. The peak in grade 7 is due to the fact that new students were added to the panel. There is also substantial attrition between grades 7 and 9. The observed increase in the initial achievement (from -.46 to -.43) suggests that this attrition concerns mainly lower ability students. At the same time, however, average math achievement is broadly unchanged.

3 Empirical Strategy

In Germany, teachers at the end of primary school have to inform parents with a written recommendation about the secondary school track they consider most suitable to each student (see Kultusministerkonferenz 2010). In the majority of states, however, parents can ultimately decide where to enroll their children, the most common case being the choice of the upper-tier track (Gymnasium) even when this track is considered too challenging by teachers.

In a few states, however, parental choice is limited. In Baden-Wuerttemberg, Saxony, and Saxony-Anhalt, students have to pass a central admission exam in order to access the upper-tier secondary school track. In Bavaria, Brandenburg, North Rhine-Westphalia and Thuringia, students must instead attend "probationary classes", which are usually formed towards the end of primary education and taught by both a primary and a secondary teacher. After two or three school days, these teachers decide whether candidate students are suited to the preferred secondary school track. In these states, school performance in the final grade is a crucial factor affecting future education. Therefore, the final grade can be considered as a high stakes.

A key feature exploited in this paper is that - although selection rules are similar across the last four states - the critical final grade is the fourth in Bavaria, North Rhine-Westphalia and Thuringia and the sixth in Brandenburg. Therefore, students in the fourth grade in Brandenburg face lower stakes than students in the same grade in the other three states. We compare the math test scores of fourth graders across these states by assigning Brandenburg students to the control group ($D = 0$) and the students in Bavaria and North Rhine-Westphalia to the treatment group ($D = 1$).[2] As shown in Table 2a, 97 and 3 percent of the working sample belong to treatment and control, respectively.

A similar approach can be applied to secondary school students enrolled in the lower and middle track. Since the final school grade is the ninth in the lower track and the tenth in the middle track, tests taken during the ninth grade occur in a high stakes environment for the former and in a lower stakes one for the latter. Table 2b shows that around a third of the secondary school students in the sample belongs to the treatment group ($D = 1$) and the rest is assigned to the control group ($D = 0$). Table 3 summarizes our treatment and control group definitions.

3.1 Difference-in-Differences Model

We estimate the following empirical model:

$$\begin{aligned} mathscore_{it} \; = \; & \alpha + \lambda_2 \cdot t_2 + \lambda_3 \cdot t_3 + \\ & \beta \cdot D_i + \delta_2 \cdot (t_2 \times D_i) + \delta_3 \cdot (t_3 \times D_i) + \varepsilon_{it} \end{aligned} \tag{1}$$

where the standardized math score of student i measured in period $t \in \{1, 2, 3\}$ is regressed on two time fixed effects t_2 and t_3, the treatment dummy D and the interactions of treatment with time effects. Standard errors are clustered at the school level.

[2] We omit students in Thuringia because their pre-treatment trends in math scores deviate from the trends in the other three states.

We estimate two specifications of (1): the baseline, which adds to (1) as additional regressors individual age, gender and parental education, and controls for unobserved ability using initial math scores; an alternative specification which controls for unobserved ability using student fixed effects. Compared to the baseline, the second specification considers only students who participated in all three assessments.

During the first and second assessment, stakes are low for both treatment and control groups. If the parallel trends assumption holds, the estimated parameter δ_2 should be statistically equal to zero. As students move from the second to the third test, however, the stakes become high for the treatment and remain low for the control group. Therefore, parameter δ_3 is the average treatment effect of a high stakes school grade on math test scores.[3]

4 Results

Main Findings

Table 4 presents our main findings. The first two columns refer to primary school students and the latter two to secondary school students. Columns (1) and (3) are for the baseline specification, and columns (2) and (4) for the specification with student fixed effects.

We find that the estimated parameter $\hat{\delta}_2$ is always very small and statistically not different from zero, suggesting that the parallel trends assumption required by the difference-in-differences model is satisfied. Conversely, estimated parameter $\hat{\delta}_3$ is positive and statistically different from zero. The estimated effect ranges from 0.17 to 0.23 standard deviations, a sizeable impact given that one additional year of school is estimated to increase average scores by one quarter to one third of a standard deviation (Wößmann 2016).

Table 5 provides estimates by gender. We find that high stakes effects are larger

[3] Using the potential outcomes notation, the counterfactual math score of treated students in period $t = 2$ is assumed to be $E(mathscore^0 \mid D = 1, t = 2) := \alpha + \beta + \lambda_2$. Hence, $\delta_2 = E(mathscore^1 \mid D = 1, t = 2) - E(mathscore^0 \mid D = 1, t = 2)$ tests for parallel trends. On the other hand, parameter $\delta_3 = E(mathscore^1 - mathscore^0 \mid D = 1, t = 3)$ captures the high stakes effect.

(and more precisely estimated) for girls than for boys, although the difference is not statistically significant at the conventional levels of confidence. A reason why girls may be more affected by high stakes than boys is that girls take exams in high stakes periods more seriously than boys. As shown by Xu (2006) and Wagner et al. (2008), for example, girls spend more time on homework than boys. Another reason is that, since girls have lower average math scores than boys, they may have more room "to catch up" in tests taken in high stakes grades.

In Table 6 we compare students whose parents have completed at most vocational training ("≤voc.") with students having parents with higher education (either upper-secondary school or tertiary education (">voc."). We find that high stakes effects are larger for the less privileged, although the difference with the more privileged is not statistically different from zero.

This finding could be due to the fact that, when parents are well-educated, their supervision and monitoring efforts are high in general, and therefore less affected by changes in stakes. When parents are less-educated, however, they may intensify their monitoring efforts as stakes increase. Alternatively, the marginal returns to effort in high stake grades may be higher for the less privileged, who do not have access to the networks available to the privileged (Brunello et al. 2009).

Mechanisms

What are the mechanisms explaining our findings? Candidates include measures of effort by students, teachers and parents. Our data include information on parental monitoring, private tutoring, teaching methods and teacher satisfaction with student performance. We add these variables as controls in (1) but find that they have a negligible impact on both the size and the precision of $\hat{\delta}_3$, perhaps because of the large number of missing values - often above 50 percent of the sample.

A potential mechanism is that students in high stakes periods are more likely to be taught by higher quality teachers. To investigate this, we use periods $t = 1, 2$ - when stakes are low for both treatment and control students - and a value added specification to estimated teacher fixed effects. We merge these effects to our working sample, which contains also period $t = 3$. By so doing, however, we lose a substantial number of observations, because two thirds of ninth grade secondary school teachers do not appear in the fifth grade (the situation being even worse for the primary schools of Bavaria and Brandenburg). With these limitations in mind, we find that adding teacher quality to (1) does not affect the key coefficients, suggesting that it is not driving our findings.

5 Conclusions

Using a difference-in-differences technique, we have compared German primary and secondary school students taking a common math test in the same grade but with different stakes. We have exploited the fact that the high stakes final grade of primary or secondary education differs across German states or school tracks. We have found that taking the test in a high stakes grade has statistically significant and sizeable effects on math test scores. In particular, we have estimated that high stakes increase test scores by .23 and .17 standard deviations in primary and secondary schools, respectively. This gain is equivalent to 0.6 additional years of schooling, and is higher for girls and for students with a less privileged parental background.[4]

These results suggest that students who may perform poorly when adequate motivation is missing (as in low stakes tests) could perform significantly better in the presence of stronger incentives. They confirm that caution should be exercised when comparing or even ranking countries using average low-stakes assessments (see Gneezy et al. 2017). These ranks may change when, in some countries, tests are taken during high stakes periods. It would be also important to know whether the gains associated to high stakes are temporary or permanent. As our data do not

[4] We have also estimated the effect of high stakes on reading test scores, which can be done in our data only for secondary school students. Consistent with Jacob (2005), we have found smaller (and, in our case, statistically insignificant) high stakes effects than for math scores. Jacob (2005) notes that "education evaluations... show larger effects in math than reading, presumably because reading achievement is determined by a host of family and other non-school factors while math achievement is determined largely by school." (p.771)

permit to separate these effects, this important question, as well as a more detailed investigation of mechanisms, must be left to future research.

References

Angrist, Joshua, and Victor Lavy. 2009. "The effects of high stakes high school achievement awards: Evidence from a randomized trial." *American Economic Review*, 99(4): 1384–1414.

Behrman, Jere, Michela Tincani, Petra Todd, and Kenneth Wolpin. 2016. "Teacher quality in public and private schools under a voucher system: The case of Chile." *Journal of Labor Economics*, 34(2): 319–362.

Bela, Daniel, Sebastian Pink, and Jan Skopek. 2012. "Data manual. Starting Cohort 3 – From lower to upper secondary school. NEPS SC3 1.0.0." NEPS Research Data Paper, University of Bamberg.

Bishop, John. 1998. "Do curriculum-based external exit exam systems enhance student achievement?" Working Paper.

Blossfeld, Hans-Peter, Hans-Günther Rossbach, and Jutta von Maurice. 2011. *Education as a Lifelong Process – The German National Educational Panel Study (NEPS).* Zeitschrift für Erziehungswissenschaft – Sonderheft 14. Wiesbaden: Springer.

Borghans, Lex, Angela Lee Duckworth, James Heckman, and Bas Ter Weel. 2008. "The economics and psychology of personality traits." *Journal of human Resources*, 43(4): 972–1059.

Brunello, Giorgio, Margherita Fort, and Guglielmo Weber. 2009. "Changes in compulsory schooling, education and the distribution of wages in Europe." *The Economic Journal*, 119(536): 516–539.

Bursztyn, Leonardo, and Robert Jensen. 2015. "How does peer pressure affect educational investments?" *Quarterly Journal of Economics*, 130(3): 1329–1367.

Dearden, Lorraine, Carl Emmerson, Christine Frayne, and Costas Meghir. 2009. "Conditional cash transfers and school dropout rates." *Journal of Human Resources*, 44(4): 827–857.

Duckworth, Angela, Patrick Quinn, Donald Lynam, Rolf Loeber, and Magda Stouthamer-Loeber. 2011. "Role of test motivation in intelligence testing." *Proceedings of the National Academy of Sciences*, 108(19): 7716–7720.

Duflo, Esther, Rema Hanna, and Stephen Ryan. 2012. "Incentives work: Getting teachers to come to school." *American Economic Review*, 102(4): 1241–78.

Federivcova, Miroslava, and Daniel Munich. 2017. "The impact of high-stakes school admission exams on study achievements: quasi-experimental evidence from Slovakia." *Journal of Population Economics*, 30(4): 1069–1092.

Figlio, David, and Susanna Loeb. 2011. "School accountability." Vol. 3 of *The Handbook of the Economics of Education*, ed. by Eric Hanushek and Finis Welch, Chapter 8, 383–421. Elsevier.

Frederiksen, Norman. 1994. "The influence of minimum competency tests on teaching and learning." Working Paper.

Glewwe, Paul, Nauman Ilias, and Michael Kremer. 2010. "Teacher incentives." *American Economic Journal: Applied Economics*, 2(3): 205–27.

Gneezy, Uri, John List, Jeffrey Livingston, Sally Sadoff, Xiangdong Qin, and Yang Xu. 2017. "Measuring success in education: the role of effort on the test itself." NBER Working Paper 24004, Cambridge: National Bureau of Economic Research.

Jacob, Brian. 2005. "Accountability, incentives and behavior: The impact of high-stakes testing in the Chicago Public Schools." *Journal of Public Economics*, 89(5-6): 761–796.

Jürges, Hendrik, Kerstin Schneider, and Felix Büchel. 2005. "The effect of central exit examinations on student achievement: Quasi-experimental evidence from TIMSS Germany." *Journal of the European Economic Association*, 3(5): 1134–1155.

Kultusministerkonferenz (Ed.). 2010. *Übergang von der Grundschule in Schulen des Sekundarbereichs I und Förderung, Beobachtung und Orientierung in den Jahrgangsstufen 5 und 6 (sog. Orientierungsstufe).* Bonn: Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland.

Neill, Monty, and Keith Gayler. 2001. "Do high-stakes graduation tests improve learning outcomes? Using state-level NAEP data to evaluate the effects of mandatory graduation tests." *Raising Standards or Raising Barriers*, 107–126.

OECD (Ed.). 1999. *Classifying educational programmes: Manual for ISCED-97 implementation in OECD Countries.* Paris: OECD Publishing.

Rouse, Cecilia Elena. 1998. "Private school vouchers and student achievement: An evaluation of the Milwaukee Parental Choice Program." *Quarterly Journal of Economics*, 113(2): 553–602.

Segal, Carmit. 2012. "Working when no one is watching: Motivation, test scores, and economic success." *Management Science*, 58(8): 1438–1457.

Wagner, Petra, Barbara Schober, and Christiane Spiel. 2008. "Time students spend working at home for school." *Learning and Instruction*, 18(4): 309–320.

Winfield, Linda. 1990. "School competency testing reforms and student achievement: Exploring a national perspective." *Educational Evaluation and Policy Analysis*, 12(2): 157–173.

Wößmann, Ludger. 2016. "The importance of school systems: Evidence from international differences in student achievement." *Journal of Economic Perspectives*, 30(3): 3–32.

Xu, Jianzhong. 2006. "Gender and homework management reported by high school students." *Educational Psychology*, 26(1): 73–91.

Tables

Table 1: Timing of math skill assessments in NEPS

t	$= 1$	$= 2$	$= 3$
Primary sample	1st grade	2nd grade	4th grade
Secondary sample	5th grade	7th grade	9th grade
Stakes	low (always)	low (always)	high (sometimes)

Table 2a: Descriptive statistics (German primary school students)

Grade	1st grade		2nd grade		4th grade	
	mean	s.d.	mean	s.d.	mean	s.d.
Math achievement	.00	.98	.02	.98	.07	.95
Age	7.80	.68	8.80	.68	10.80	.68
Female	.51		.51		.51	
Initial math achievement	.00	.98	.01	.98	.03	.97
Highest parental education:						
lower- or middle-secondary	.04		.04		.04	
upper-secondary	.12		.12		.12	
vocational training	.23		.23		.23	
tertiary	.61		.61		.61	
% missing parental educ.	9%		9%		8%	
N(students)	2760		2672		2458	
% in treatment group	97%		97%		97%	

Data source: NEPS Starting Cohort 2. Primary students who were assessed during the first, second, and fourth grades. Standard deviations are not reported for binary random variables. Prior to restricting the original data to our working sample, math achievement was standardized to have mean 0 and standard deviation 1 within each grade. Highest parental education is classified using International Standard Classification of Education (ISCED) guidelines, see OECD (1999) for details.

Table 2b: Descriptive statistics (German secondary school students)

Grade	5th grade		7th grade		9th grade	
	mean	s.d.	mean	s.d.	mean	s.d.
Math achievement	−.52	.84	−.53	.88	−.52	.81
Age	11.19	.84	13.19	.85	15.19	.85
Female	.47		.46		.45	
Initial math achievement	−.52	.84	−.46	.84	−.43	.85
Highest parental education:						
lower- or middle-secondary	.07		.05		.05	
upper-secondary	.07		.06		.07	
vocational training	.60		.66		.64	
tertiary	.25		.23		.24	
% missing parental educ.	32%		41%		39%	
N(students)	1906		2232		1675	
% in treatment group	37%		32%		30%	

Data source: NEPS Starting Cohort 3. Lower-tier and middle-tier secondary students who were assessed in the fifth, seventh, and ninth grades. Standard deviations are not reported for binary random variables. Prior to restricting the original data to our working sample, math scores were standardized to have mean 0 and standard deviation 1 within each grade.

Table 3: Definition of treatment and control groups (by schooling level)

Sample	D	Federal state or school tier	$t = 3$
Primary	Treated $D = 1$	Bavaria and North Rhine-Westphalia	Final grade in primary school
	Control $D = 0$	Brandenburg	Primary school lasts two more years
Secondary	Treated $D = 1$	Lower-tier secondary school (Hauptschule)	Final grade in lower-sec. school
	Control $D = 0$	Middle-tier secondary school (Realschule)	Middle-sec. school lasts one more year

Table 4: High stakes effects on math scores. Primary and secondary education

	Primary school		Secondary school	
Specification:	baseline	FE	baseline	FE
$\hat{\delta}_2$	0.00	-0.03	-0.02	-0.05
	(0.085)	(0.091)	(0.033)	(0.055)
$\hat{\delta}_3$	0.23***	0.21***	0.17***	0.18***
	(0.070)	(0.080)	(0.051)	(0.067)
R^2	0.60	0.01	0.69	0.03
N	7890	6684	5813	3078

The first specification ("baseline") includes the control variables listed in Tables 2a and 2b plus a dummy variable for missing parental education. The second specification ("FE") adds student fixed effects and age to (1). Sample sizes are smaller in the fixed effect (FE) specifications because students must have participated in all three assessments.

Table 5: High stakes effects by gender

	Primary school		Secondary school	
Subgroup:	girls	boys	girls	boys
mean(*mathscore*)	−.13	.12	−.68	−.38
$\hat{\delta}_2$	-0.02	0.02	0.00	-0.04
	(0.092)	(0.101)	(0.042)	(0.043)
$\hat{\delta}_3$	0.25***	0.21	0.21***	0.15**
	(0.088)	(0.145)	(0.059)	(0.067)
R^2	0.60	0.60	0.67	0.69
N	4033	3857	2674	3139

All estimates include the control variables listed in Tables 2a and 2b plus a dummy variable for missing parental education. Mean math scores are computed at $t = 1$. For both the primary and secondary samples, we cannot reject the null hypothesis $\hat{\delta}_3^{girls} = \hat{\delta}_3^{boys}$. The p-value of the corresponding test statistic is .84 in the primary sample, and .42 in the secondary sample (based on SUEST).

Table 6: High stakes effects by parental education

	Primary school		Secondary school	
Subgroup:	≤voc.	> voc.	≤voc.	> voc.
mean(*mathscore*)	−.52	.21	−.64	−.23
$\hat{\delta}_2$	-0.03	0.02	-0.01	-0.07
	(0.214)	(0.084)	(0.034)	(0.070)
$\hat{\delta}_3$	0.30	0.19**	0.18***	0.12
	(0.217)	(0.088)	(0.048)	(0.107)
R^2	0.61	0.54	0.69	0.64
N	2324	5566	4176	1637

Students whose parents have at most completed vocational training ("≤voc."), and students whose parents have completed either upper-secondary school or tertiary education (">voc."). All estimates include the control variables listed in Tables 2a and 2b plus a dummy variable for missing parental education. For both the primary and secondary samples, we cannot reject the null hypothesis $\hat{\delta}_3^{\leq \text{voc.}} = \hat{\delta}_3^{> \text{voc}}$. The p-value of the corresponding test statistic is .67 in the primary sample, and .53 in the secondary sample (based on SUEST).