

IFN Working Paper No. 1263, 2019

Gender Grading Bias in Junior High School Mathematics

Petter Berg, Ola Palmgren and Björn Tyrefors

Gender Grading Bias in Junior High School Mathematics

Petter Berg^a, Ola Palmgren^b and Björn Tyrefors^{c*}

February 13, 2019

Affiliations

^a Petter Berg and ^b Ola Palmgren

Department of Economics, Stockholm University, SE-10691 Stockholm

Tel.: +46(0)8-16 20 00

^{c*} **Corresponding author:** Björn Tyrefors

Research Institute of Industrial Economics (IFN), Box 55665, 102 15 Stockholm

E-mail: bjorn.tyrefors@ifn.se

Tel.: +46(0)8-665 4500

Abstract

Admission to high school in Sweden is based on the final grades from junior high. This paper compares students' final mathematics grade with new data from a high school introductory test score in mathematics. Both the grades and the test are based on the same syllabus, but teachers enjoy great discretion when deciding final grades, as the grades are not externally evaluated. The results show a substantial grading difference, consistent with grading bias against boys.

Keywords: Educational economics; Gender; Grading bias; Mathematics

Classification codes: I20; I24; I28; J16

1 Introduction

Though they have received increasing attention in economics, studies on biased grading are still few. Interest is motivated by the growing gender gap in educational attainment. Previous studies on gender grading bias have found bias against boys.¹ A related strand of literature study how having a teacher of the same type affects students' performance.² In this paper, we compare final junior high school mathematics grades to high school introductory test scores, both based on the very same syllabus.

We find that boys are given lower final grades but that the tests taken about three months later show that boys outperform girls. While this is consistent with bias against boys in the final grading, we cannot be certain that this difference is discriminatory. Even though both tests are based on the same syllabus, it may be possible that different skills still are measured. However, the well-known evaluation in mathematics, Trends in International Mathematics and Science Study (TIMSS), revealed that in 2015, boys outperformed girls in Sweden. Additionally, the math evaluation in PISA 2015 indicates that boys and girls are roughly on par (OECD, 2019). As far as these evaluations measure objective math knowledge, our results point to a grading bias. Nevertheless, boys could be more motivated in a low-stakes test. However, this is at odds with prior research that shows girls exert more effort than boys (see DeMars et al. , 2013). Lastly, our results are also related to studies indicating the problems with being graded by one's own teacher as grades seem to unfairly benefit certain groups (Hinnerich and Vlachos, 2017). As final grades

¹ E.g. see Lavy (2008), Hinnerich et al. (2011), Hinnerich and Vlachos (2013), Lindahl (2007), Feld et al. (2016).

² See Dee (2005) and Cornwell et al. (2013).

are the sole selection mechanism to higher education, this is crucial from an economic efficiency point of view.

2 Material and methods: institutional background, data and empirical design

2.1 Institutional background

The final mathematics grade in junior high school is given after nine years of compulsory schooling. The vast majority of Swedish youth enroll in high school education. Students' achievements in different subjects are graded on a 7-tiered scale from A to F, where F indicates failing. To calculate a grade point average (GPA), the main selection mechanism for students going from compulsory school to high school, the grades are translated into a cardinal scale with 0 for an F, 10 for an E and then incrementally by 2.5 to 20 for an A. Final grades are based on absolute knowledge criterion and Mathematics, Swedish and English have nationally stipulated prerequisites for each grade. Grades are based on the level of knowledge, and must not reflect participation or ambition. In practice, however, teachers enjoy great discretion when deciding grades. Because grades are not externally evaluated, teachers could base their grades on anything they observe.

2.2 Data

We study the test takers of the introductory mathematics test when starting high school in the Stockholm municipality for the years 2012-2017. While all public high schools must conduct the math test, voucher schools do so voluntarily. The goal is to evaluate the sending (junior high) school's quality given the syllabus. Moreover, the test is supposed to be an instrument to help the

receiving (high) schools take action (Education Department, 2014). The test is developed by experts on mathematics grading at Stockholm University based on the syllabus from grade 9 and consists of 35 questions. It is graded at the receiving schools, with no name removal before grading. However, strict instructions are given on a point-by-point basis.

Table 1 shows the number and share of participating schools and. If fewer than 10 students took the test at a school, data are excluded. Hence, we study approximately 20 participating schools, or 90% of public high schools. Another source of attrition is students being absent on the test day. Therefore, at the individual level, we capture approximately 80% of students in Stockholm public high schools.

Table 1. Number and share of participating public high schools and students

Year	2012	2013	2014	2015	2016	2017	N
No. of participating schools	21	20	20	21	19	22	123
Share of participating schools	91%	87%	87%	91%	86%	92%	
No. of participating students	3497	2571	3468	3613	3643	4407	21199
Share of participating students	75%	57%	75%	81%	81%	88%	

Because participation is voluntary for voucher schools, there is a lower participation rate here.³ Almost 100% of sending schools are represented, but with a lower student participation rate as all junior high school graduates do not immediately continue onto high school and also may choose to attend high schools outside Stockholm municipality.⁴

³ The participation rate is approximately 40 %. However, results are very similar with or without considering students at voucher schools.

⁴ Data can be made available upon request.

Table 2 summarizes the final grades and test scores. We see that girls on average have a grade of 15.1 compared to 14.8 for boys, a difference of approximately 0.3 units but that the relationship is the opposite for the test scores, whereby boys outperform girls by approximately 1.6 points. Compared to the overall mean, this is an approximately 7% difference in favor of boys.

Table 2. Summary statistics

	<i>All</i>				<i>Girls</i>				<i>Boys</i>			
	Mean	SD	Min	Max	Mean	SD	Min	Max	Mean	SD	Min	Max
Grade	14.97	3.71	10	20	15.11	3.69	10	20	14.82	3.73	10	20
Score	23.27	10.47	0	43	22.54	10.36	0	43	24.15	10.52	0	43

2.3 Empirical design

We follow the standard methodological approach used in previous literature⁵ by postulating a difference-in-difference equation as

$$Grade_i - Testscore_i = \alpha + \beta Boy_i + u_i \quad (1)$$

The equation difference out unobserved ability and solve the omitted variable problem, as long as ability has a similar effect on both the final grade and test score. We argue that this assumption is reasonable as both the grade and test is based on the same syllabus. Another

⁵ A discussion is found in Lavy (2008).

concern is that students' names are not removed in the introductory test. However, as pointed out in Lavy (2008), detecting biased gender grading seems unrelated to anonymity but instead hinges on the external grader having no personal relation with the student. Lastly, we also want to highlight the fact that the grading of high school mathematics tests based on the syllabus in Sweden has been found to have high reliability and that test scores have been used as reference point for ability when comparing final grades (Vlachos, 2018). Hence, we argue that β measures bias. We standardize the test scores for comparison and hence β is interpreted as the share of a standard deviation of the test scores. A negative value for β indicates that boys suffer from grading bias. We can naturally add control variables, fixed effects for year and sending school and lastly the sending school characteristics, as they change annually.

3 Results

3.1 Main results

In column 1, we estimate equation 1 and find a significant negative bias effect of 23% of a standard deviation against boys. A rough calculation comparing the grade distribution to the test distribution shows approximately 40% of male students of an average grade level. Thus, nearly every third boy receives a lower grade on average compared to the test score.⁶

Interestingly, this is approximately double the size that Lindahl (2007) estimates when comparing final grades to internally graded national tests. This is to be expected as Hinnerich and Vlachos (2013) show that national test scores are manipulated when they are internally

⁶ Calculation is available upon request.

graded. Thus, it is reasonable that the size of the bias effect when comparing final grades to an externally graded test should identify a larger bias. Lastly, when adding the fixed effects for year and sending school as well as the sending school characteristics, the results remain unchanged.

We have also run more specifications based on subgroups where little evidence of heterogeneous effects is found. Estimating per quintile of the test distribution yields a negative and substantial bias effect in all quintiles and separate models for voucher and public schools, over and above the median for the school characteristics yields estimates from -0.251 to -0.226.

Table 3. Gender grading bias effects

	-1	-2	-3	-4	-5	-6
Boy	-0.232*** (-0.007)	-0.233*** -0.007	-0.237*** -0.007	-0.238*** -0.009	-0.239*** -0.010	-0.238*** -0.011
Share of female teachers				0.255 -0.137		0.0455 -0.18
Parents' education					0.115 -0.0986	0.133 -0.0988
Share of immigrants					0.387 -0.215	0.334 -0.22
Share of boys					0.118 -0.11	0.127 -0.11
Year fixed effects	No	Yes	Yes	Yes	Yes	Yes
Sending school fixed effects	No	No	Yes	Yes	Yes	Yes
N	33486	33486	33486	19130	12572	12364
R ²	0.036	0.044	0.114	0.102	0.1	0.098

4 Conclusion

We find evidence of substantial grading bias against boys in mathematics using new diagnostic test data. The magnitude of this bias can be rationalized by the Swedish decentralized grading system whereby a grader enjoys great discretion when deciding grades. The magnitude of the bias is also consistent with findings in previous studies. This study speaks directly to the

problem of local grading with no external evaluation even though grades are stipulated by law to represent absolute knowledge. As final grades determine entrance into high school programs, this is critical with respect to efficiency and fairness. A complementary element of external grading when determining the final grade should be welfare enhancing.

Acknowledgment

We thank the Education department and Sweco for data provision. Tyrefors thanks Jan Wallanders och Tom Hedelius stiftelse and the Marianne and Marcus Wallenberg Foundation for financial support. We would like to recognize initial work on this paper by Palmgren and Berg under the supervision on Tyrefors.

Funding

This research did not receive any specific funding from the public, commercial or not-for-profit sectors.

Conflicts of interest

None.

References

- Cornwell, C., Mustard, D.B., Van Parys, J., 2013. Noncognitive skills and the gender disparities in test scores and teacher assessments: evidence from primary school. *J. Hum. Resour.* 48, 236–264. <https://doi.org/10.3368/jhr.48.1.236>.
- Dee, T.S., 2005. A teacher like me: does race, ethnicity, or gender matter? *Am. Econ. Rev.* 95, 158–165. <https://doi.org/10.1257/000282805774670446>.
- DeMars, C.E., Bashkov, B.M., Socha, A.B., 2013. The role of gender in test-taking motivation under low-stakes conditions. *Res. Pract. Assess.* 8, 69–82.
- Education Department, 2014. Resultat från Stockholmsprovet i Matematik, Genomfört i Gymnasieskolor Höstterminen 2013. (Dnr 1.2.3-129/2014) Stockholm Municipality, Utbildningsförvaltningen.
- Feld, J., Salamanca, N., Hamermesh, D.S., 2016. Endophilia or exophobia: beyond discrimination. *Econ. J.* 126, 1503–1527. <https://doi.org/10.1111/eoj.12289>.
- Hinnerich, B.T., Höglin, E., Johannesson, M., 2011. Are boys discriminated in Swedish high schools? *Econ. Educ. Rev.* 30, 682–690. <https://doi.org/10.1016/j.econedurev.2011.02.007>.
- Hinnerich, B.T., Vlachos, J., 2013. Lika för Alla? Omrättning av Nationella Prov i Grundskolan och Gymnasieskolan under tre år. Resultatbilaga, Dnr U2009/4877/G: Skolinspektionen.
- Hinnerich, B.T., Vlachos, J., 2017. The impact of upper-secondary voucher school attendance on student achievement. Swedish evidence using external and internal evaluations. *Lab. Econ.* 47, 1-14. <https://doi.org/10.1016/j.labeco.2017.03.009>

Lavy, V., 2008. Do gender stereotypes reduce girls' or boys' human capital outcomes? Evidence from a natural experiment. *J. Public Econ.* 92, 2083–2105.

<https://doi.org/10.1016/j.jpubeco.2008.02.009>.

Lindahl, E., 2007. Spelar Lika kön och Etnisk Bakgrund på Lärare och Elever Roll för Provresultat och Slutbetyg? IFAU Rapport 2007, Uppsala, p. 23.

OECD, 2019. Mathematics Performance (PISA) (indicator). doi: 10.1787/04711c74-en.

Vlachos, J., 2018. Trust-Based Evaluation in a Market-Oriented School System, IFN Working Paper No 1217.