Studies in Educational Evaluation 59 (2018) 141–149

Contents lists available at ScienceDirect

Studies in Educational Evaluation

journal homepage: www.elsevier.com/locate/stueduc

Gender Bias in teachers' grading: What is in the grade☆

Tomáš Protivínský, Daniel Münich*

CERGE-EI, a joint workplace of Charles University in Prague and the Economics Institute of the Czech Academy of Sciences, Politickych veznu 7, 111 21 Prague, Czech Republic

ARTICLE INFO

Keywords:
Educational economics
Gender gap
Discrimination
Grading
The Czech Republic
Academic achievement
Educational assessment
Performance of educational system

ABSTRACT

It is a common empirical finding across countries that, on average, girls outperform boys in languages, whereas boys outperform girls in mathematics. We enrich the existing empirical evidence by combining admission test scores and teachers' grading of 15-year-old pupils' performance in mathematics and their native language in the Czech Republic, and furthermore, we investigate possible gender bias in teachers' grading. The directions of the gender differences in performance we estimate are consistent with international patterns and we document that teachers' grading is biased in favor of girls both in mathematics and in native language. The gender effect in grading is sizeable across the whole performance distribution and can be explained neither by the students' differing perceptions of stress at exams, nor by the students' attitudes toward the subject in question. The most plausible explanation is that the gender grading gap is due to gender difference in non-cognitive skills, such as in-class behavior and homework, confounding teachers' grades but not test scores. Since grades constitute the main feedback about students' academic performance and a crucial factor in decision-making about their future academic careers, biased grading may cause inefficiencies to the educational system and consequently can negatively affect future labor market careers.

1. Introduction

Various types of gender gap are the subject of intensifying empirical research efforts worldwide. Our work is focused on gender bias in teachers' grading, which confounds the quantitative indicators used to identify actual (real) gender gaps in pupils' and students' skills achievement. Two different types of gender gap can be found in the economics of education literature: gaps in achievement and relative gaps in grades, awarded by teachers at their discretion, in contrast to test scores. The term "gender gap" (sometimes also called "gender achievement gap"), usually referring to the gap in performance between girls and boys, is commonly measured by the difference in their test scores. Several stylized patterns have been observed: girls outperform boys in languages, whereas boys typically score better in mathematics, although the latter difference is not as distinct and pronounced (cf. Machin & Pekkarinen, 2008; Cornwell, Mustard, & Van Parys, 2013; Guiso, Monte, Sapienza, & Zingales, 2008; Angelo, 2014). Machin and McNally (2005) document that the position of girls relative to boys has improved over the past few decades. Numerous authors have concluded that this gender gap originates from a gender-biased culture and environment rather than from innate biological differences in ability (Guiso et al., 2008; OECD, 2015). Relatively few authors, however, have explored gender gaps in grading as opposed to test-scores. In the Czech Republic, Straková, Potužníková, and Tomášek (2006) found that on average girls receive higher grades than boys in all measured subjects across the majority of types of education.

The gender grading gap then captures the extent to which a boy and a girl with the same level of skills (based on anonymous testing) obtain different grades via non-anonymous evaluation, typically by teachers. It constitutes a gender bias in grading when contrasted with other forms of evaluation. In most schooling systems, grades represent an established form of feedback to pupils and their parents about their performance and/or learning progress. Appropriate feedback can help pupils and parents to learn about the efficiency of their effort and foster their further progress. Grading is also an important factor in forming aspirations and making career-related decisions; Federičová (2016) shows that students rely heavily on grades when deciding whether to apply to selective high-schools. Symmetrically, incorrect or biased feedback can impede educational attainment and cause sub-optimal skills acquisition with life-long consequences. It can foster a mismatch between students' skills and the educational track they choose, and consequently lead to occupational mismatch on the labor market.

As Hinnerich, Höglin, and Johannesson (2011) put it, given the growing focus of contemporary societies on gender equality, it is

☆ The research was supported by the Czech Science Foundation (grant P402/12/G130).
* Corresponding author.
E-mail address: daniel.munich@cerge-ei.cz (D. Münich).

https://doi.org/10.1016/j.stueduc.2018.07.006
Received 13 October 2017; Received in revised form 13 July 2018; Accepted 16 July 2018

desirable to explore the nature and the causes of the gender grading gap in different settings and countries more intensively. So far, the gender grading gap has been investigated in India, Israel, Norway, Portugal, Sweden, the United Kingdom and the United States, although there have been substantial differences in the methodology used and to some extent also in the findings. The gender gap in grading 15-year-olds in the Czech Republic has been researched by Matějů and Smith (2014) and Matějů and Simonová (2013). We confirm their findings here using different, and to some extent better data, and further explore the possible role of differences in pupils' attitudes to subjects and in their perception of stress during exams.

In the following section we offer a thorough literature review; Section 3 then describes the key features of the Czech educational system related to the phenomena we are investigating and the data we use. Section 4 presents our quantitative findings about gender achievement and the grading gap, and in the final section we conclude, placing our findings within the relevant policy framework and discussing their implications for further research.

2. Literature review

A gender gap in grading occurs when a teacher gives students of different genders grades that systematically differ but not due to their performance; this can be caused by numerous factors including intentional or unintentional discrimination. The typical result is that boys on average receive worse grades than girls, despite their level of academic skills being the same. Lavy (2008) introduced a very explicit term for this phenomenon: "anti-male bias in grading". Other terms describing the same notion are gender grading gap, gender differences in teachers' grading or gender bias in teachers' grading and these are used interchangeably in the presented study.

The existing literature about the gender grading gap is neither vast, nor consistent. Research methodology in this area always relies on comparing two different evaluation schemes, i.e. anonymous and non-anonymous assessment, although there are differences in the particular types of assessment used. Falch and Naper (2013) propose a classification of assessments based on three dimensions: (i) anonymous vs. non-anonymous, (ii) high-stakes vs. low-stakes, (iii) one-day vs. assessment over time. For instance, teachers' grading represents non-anonymous, high-stakes assessment over time, whereas admission exams (e.g. to high schools or university studies) can be considered as anonymous, high-stakes, one-day assessment. Even though grading by teachers, whether one-off or over a period of time, is a prototypical example of non-anonymous assessment, some authors have also included other forms of evaluation. These differences in methodology can directly translate into differences in results (cf. Hanna & Linden, 2012).

Table 1 summarizes the existing findings about the gender grading gap. Most studies found a gender grading gap in favor of girls in some or all subjects, reporting that girls typically receive higher grades than boys with the same academic skills. Only two studies did not find any bias in grading – Hanna and Linden (2012) and Hinnerich et al. (2011). The latter investigated only one subject – Swedish as a native language, while the former used a very different methodology than all the other studies. As non-anonymous evaluation, Hanna and Linden (2012) asked teachers to evaluate an exam sheet with a particular gender, age and caste assigned. Falch and Naper (2013) suggest that the gender grading gap is caused by teacher-student interaction. This hypothesis is consistent with Hanna and Linden's finding that the gender grading gap does not occur at all when there is no teacher-student interaction. This disparity also suggests that personal ties between a teacher and a student might be important for the occurrence of the gender grading gap. The most distinct results are reported by Gibbons and Chevalier (2008) and Lavy and Sand (2015), who found bias against boys in some subjects and bias against girls in other subjects. These are the only cases when bias in grading against girls was identified in the empirical literature; the particular subjects concerned differ in these studies.

Two other differences between teachers' grading and anonymous evaluation that are documented in the empirical literature should be noted. The inflation of teachers' grading describes a common situation when pupils graded by their own teachers receive better grades compared to anonymous evaluation (Falch & Naper, 2013; Hinnerich et al., 2011). Lower variance of teachers' grading reflects the fact that teachers frequently under-assess high-achieving students and over-assess low-achieving students (Gibbons & Chevalier, 2008).

Several authors provide various possible explanations for why grades could be biased in favor of one gender (Angelo, 2014; Falch & Naper, 2013; Lavy, 2008): (i) different intensity of competition during evaluations, (ii) teacher-student interaction, (iii) non-cognitive skills reflected in teachers' grading. Lavy (2008) offers several other possible explanations, although his data does not confirm them – these are that anonymous and non-anonymous assessment might emphasize different parts of curriculum, hence measure different skills, interaction between teacher's grading bias and student's ability (in combination with lower variance of teachers' grading), statistical discrimination against boys or different timing of exams. Below, we briefly discuss the three aforementioned explanations and the empirical evidence in support of them.

2.1. Competitive environment

The majority of research compares two different evaluations related to forms of assessment that possibly differ in terms of the intensity of competitiveness and stress to which they subject the students. Several studies about performance under stress suggest that men outperform women, even if they would perform similarly in a non-competitive environment (in an experimental setting: Gneezy, Niederle, & Rustichini, 2003; in a real-world setting: Ors, Palomino, & Peyrache,

Table 1
An overview of the existing research about the gender grading gap (studies are listed in alphabetical order, including the findings of this study).

Author	Country	Age	Finding
Angelo (2014)	Portugal	Grade 12	Bias against boys in math and Portuguese
Bonesrønning (2008)	Norway	Grade 9	Bias against boys in math
Cornwell et al. (2013)	US	Kindergarten - grade 5	Bias against boys in reading, math and science
Emanuelsson and Fischbein (1986)	Sweden	Grade 6, 8, 9	Bias against boys in Swedish, math and English
Falch and Naper (2013)	Norway	Grade 10	Bias against boys in math and English, small bias also in Norwegian
Gibbons and Chevalier (2008)	UK	Age 14	Bias against boys in mathematics and science and against girls in English
Hanna and Linden (2012)	India	Age 7–14	No evidence of bias in exam comprising math, language skills and art section
Hinnerich et al. (2011)	Sweden	Grade 10–12 (high school)	No evidence of bias in Swedish
Lavy (2008)	Israel	Grade 10–12	Bias against boys in chemistry, computer science, math, physics, bible studies, biology, English, history, literature
Lavy and Sand (2015)	Israel	Grade 5–6	Bias against boys in English and against girls in math
Lindahl (2007a)	Sweden	Grade 9	Bias against boys in math, English and Swedish
Matějů and Smith (2014)	Czech Republic	Grade 9	Bias against boys in math and Czech language
Münich & Protivínský (2018)	Czech Republic	Grade 9	Bias against boys in math and Czech language