Journal of Public Economics 92 (2008) 2083–2105

Contents lists available at ScienceDirect

Journal of Public Economics

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

Do gender stereotypes reduce girls' or boys' human capital outcomes? Evidence from a natural experiment ☆

Victor Lavy *

Department of Economics, Hebrew University, Israel
Department of Economics, Royal Holloway, University of London, United Kingdom
NBER, United States
CEPR, United Kingdom

ARTICLE INFO

Article history:
Received 9 November 2006
Received in revised form 12 February 2008
Accepted 25 February 2008
Available online 12 March 2008

Keywords:
Gender stereotypes
Discrimination
Natural experiment

ABSTRACT

Schools and teachers are often said to be a source of stereotypes that harm girls. This paper tests for the existence of gender stereotyping and discrimination by public high-school teachers in Israel. It uses a natural experiment based on blind and non-blind scores that students receive on matriculation exams in their senior year. Using data on test results in several subjects in the humanities and sciences, I found, contrary to expectations, that male students face discrimination in each subject. These biases widen the female–male achievement difference because girls outperform boys in all subjects, except English, and at all levels of the curriculum. The bias is evident in all segments of the ability and performance distribution and is robust to various individual controls. Several explanations based on differential behavior between boys and girls are not supported empirically. However, the size of the difference is very sensitive to teachers' characteristics, suggesting that the bias against male students is the result of teachers', and not students', behavior.

1. Introduction

Schools and teachers are often said to be a source of stereotypes that harm girls. Bernard (1979), Dusek and Joseph (1983), Madon et al. (1998), and Tiedemann (2000) are only a few of the many scholars who have claimed that teachers occasionally rely on stereotypes in forming perceptions about their students.[1] Examples of such stereotypical perceptions are that boys excel in math and science and girls excel in other subjects, or that boys have more talent and that girls compensate by working hard (Deaux and LaFrance, 1998). Girls are then encouraged, on the basis of these stereotypes, to pursue traditionally female studies instead of mathematics, science, and other traditionally male subject areas, from as early as first grade (Carr et al., 1999) and women are steered toward certain occupations, as evidenced by studies of college students (Glick et al., 1995), PhD holding research students, (Rowsey, 1997) and others (Deaux and LaFrance, 1998). Another claim about stereotypes is that beliefs are manifested through teachers' evaluation of students. This claim is supported by evidence from a survey of 1st grade teachers (Fennema et al., 1990), the AAUW (1992) report, which surveyed girls from kindergarten to 12th grade, a survey of mothers and their 11–12 year old children

☆ I would like to thank Daron Acemoglu, Joshua Angrist, David Author, Abhijit Banerjee, Raquel Fernandez, Oded Galor, Esther Duflo, Guy Michaels, Sendhil Mullainathan, Analia Schlouser, Michael Steinberger for discussion and suggestions and seminar participants at CEPR conference, Hebrew University, MIT, Tel-Aviv University and the referees and editor of this journal for useful comments. Alex Levkov and Katherine Eyal provided excellent research assistance. I also thank the Israeli Education Ministry officials for assisting with the data. All errors are my own.

* Department of Economics, Hebrew University, Mt. Scopus, Jerusalem 91905, Israel.
E-mail address: msvictor@mscc.huji.ac.il.

[1] These conclusions result from studies with widely different samples, such as 240 male and female high-school teachers (Bernard, 1979), a meta study of 75 different studies (Dusek and Joseph, 1983), 2000 American 7th graders in public school math classes (Madon et al., 1998) and 600 German elementary school students (3rd and 4th graders and their parents) (Tiedemann, 2000).

0047-2727/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.jpubeco.2008.02.009

(Jacobs and Eccles, 1992), and others (Ben-Zvi Mayer et al., 1995; Hildebrand, 1996).[2] The bottom line of the literature on gender stereotypes is that they are partly responsible for one of the alleged forms of discrimination against women and that they may have far reaching implications and consequences regarding gender differences in human capital investment and outcomes. However, there is very little convincing evidence to date that substantiates these claims and this study attempts, using a unique empirical context, to fill some of this deficiency.

This paper tests for the existence of gender stereotyping and discrimination by public high-school teachers in Israel. It uses a natural experiment based on blind and non-blind scores that students receive on matriculation exams in their junior and senior years. The natural experiment arises from rules that are used in Israel to determine scores in matriculation subjects (these rules are explained in detail in Section 3.1). This testing protocol elicits two scores, a blind and a non-blind score, both of which are meant to measure the student's knowledge of the same material. Due to this testing method, we may safely assume that the blind score is free of any bias that might be caused by stereotyped discrimination on the part of the external examiner. The non-blind score, however, may reflect biases occasioned by teachers' gender stereotypes.

As long as the two scores are comparable, i.e., as long as they measure the same skills and cognitive achievements (grounds for this assumption are discussed further in Section 3.1), the blind score may be used as the counterfactual measure to the non-blind score, which may be affected ("treated") by stereotyping and discrimination. This identification framework is similar to that used by Goldin and Rouse (2000) and by Blank (1991). I use the difference between the boys' and girls' gaps between the blind and the non-blind test scores as a measure of a potential gender bias.

Using data on all matriculation scores of several cohorts of high-school seniors in Israel, I applied this natural-experiment framework to test for the existence of such a gender bias in nine subjects—four in the humanities (English, history, literature and biblical studies), mathematics, and four in science (biology, chemistry, computer science, and physics). The distributions of the blind and non-blind scores in many of these subjects are very similar and, in many cases, are identical. The basic results of these quasi-experiments show that, contrary to expectations, the potential bias is against male students. The sign of this difference is the same in all nine subjects examined and in all tests in cases where there is more than one exam per subject. The extent of the potential bias varies by subject and test, ranging from 0.05 to 0.25 of the standard deviation of the blind-score distribution. This gap against male students, on average, doubles the gender-score difference because female students outperform male students on state external exams in all subjects except English. The results are not sensitive to various student-level controls because the identification strategy is based on differences-in-differences at the student level, for which reason individual fixed effects are assumed away. In some subjects the difference is largest for low achievers and in some subjects it is actually largest for the most proficient male students.

The basic results withstand several specification checks. Overall, they do not support the hypotheses that the gender difference in the non-blind score reflects statistical discrimination against male students. For example, limiting the sample to schools where boys outperform girls on average, overall or in specific subjects, leaves the results basically unchanged. The variance in performance of boys is higher on average and in every subject than that of girls, suggesting that statistical discrimination against boys may also occur due to "noisier" signals in boys' test scores. The data, however, do not support this interpretation because the gender potential bias is not different in schools where girls demonstrate more variability in performance on average. I also examined the possibility that the results mainly reflect the effect of the specific pattern in the timing of the exams where the state follows the school exam. Using data from a second chance state-level exam in English that was taken 4 months after the school-level exam led to almost identical estimates, suggesting that the short time interval between the state and school exams cannot explain our basic results.

An interesting and obvious question in this context is whether this estimated gender difference represents students' behavior or teachers' behavior. An example of students' behavior that may explain this potential bias is differential pattern of mean reversion by gender, e.g., due to a time-varying gender difference in knowledge or due to girls not performing as well in the state exams because they may represent a more 'pressured environment' for girls. The evidence suggests, if anything, stronger mean reversion for girls than for boys, namely girls tend to improve their scores more if they perform below average at the school exam (either relative to the class mean, the class mean by gender or the prior-self in earlier exams). Another possibility is that the state and school exams do not share the same content and/or do not measure the same skills. For example, some systematic gender differences in within-class behavior (discipline, attitude, absenteeism) may end up impacting grading at the school level and not at the state level. Unfortunately, the data do not include information that would help in directly addressing this source of concern but various pieces of indirect evidence do not support this hypothesis either. For example, the basic evidence is relatively robust to controlling for lagged school scores for other exams in the same or other subjects.

The paper also examines explanations based on empirical insights gained from experiments in social psychology. One such important insight is that a "stereotype threat"–the threat of being perceived as a negative stereotype or the fear of a poor performance that would confirm the stereotype–may be powerful enough to shape the intellectual performance and academic identities of entire groups of people. The implication in our context is that the difference in favor of girls of the gap between the non-blind and blind-test scores reflects the inferior performance of girls in non-blind tests because it involves a stereotype threat and superior performance in blind tests that conceal their gender. The evidence provided in this paper does not allow us to state that this mechanism explains the negative male difference. Instead, the data support the hypothesis that teachers' behavior is responsible for this difference. I show that the gender potential bias in various subjects is very sensitive to teachers' characteristics

[2] Distortions in teachers' evaluation of their students were also discussed recently in another context, that of test-based accountability systems in education. For example, Jacob and Levitt (2003) have shown that teachers cheat on standardized tests (using data from Chicago Public Schools and students from grades 3 to 8), and that the frequency of cheating appears to respond strongly as incentive for high test scores increase even mildly.