Cognitive Ability Stereotyping and Gender Discrimination in Schooling Outcomes: Evidence from a Natural Experiment*

Victor Lavy**

The Hebrew University of Jerusalem and CEPR

November 2003

Incomplete and Preliminary Draft

.

* I would like to thank Joshua Angrist, David Author, Esther Duflo, Mike Greenstone, Sendhil Mullainathan, Analia Schlouser and Michael Steinberger for helpful discussions and suggestions, and also thank Education Ministry officials for assisting me with the data and Alex Levkov for his excellent research assistance. All errors are my own.
** Email: msvictor@mscc.huji.ac.il, and vlavy@mit.edu.

Cognitive Ability Stereotyping and Gender Discrimination in Schooling Outcomes: Evidence from a Natural Experiment

Abstract

Gender differences in productivity-altering characteristics, such as investment in human capital, have not been studied in the context of sex-biased practices. This lack of attention is in contrast to claims that stereotyping and impartiality are the cause of the skewed patterns of human-capital investment and occupational choice against women. A common belief is that schools and teachers are major sources of stereotypes about gender differences in cognitive ability. 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 the matriculation exams that they take during their senior year in high school. 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 of the subjects. These biases widen the female–male achievement gap because girls outperform boys in all subjects, except English, and at all levels of the curriculum. The bias is greater for underachieving male students than for more proficient ones and is resilient to the influence of open and known information about students' ability and economic incentives that sanction teachers who incur large measured differences between blind and non-blind scores. In mathematics, male teachers—one-third of all teachers of this subject—account for all of the bias against male students.

1. Introduction

Gender-biased practices have been the focus of many studies. The empirical literature on gender discrimination has focused mainly on disparities in labor-market outcomes between males and females, foremost in earnings and hiring, given differences in observable productivity-altering characteristics. However, gender differences in productivity-altering characteristics, such as the form of human capital and occupational choice, have not been studied in this context of gender-biased practices. This lack of attention is surprising in view of the persistence of substantial male–female differences in educational and occupational choices and the expression of considerable concern about the dearth of female scientists and engineers in the United States and Europe. Women continue to avoid college majors and occupations that entail moderate amounts of coursework in mathematics, even though they match or surpass men's performance in high-school math and science courses.[1] Gender-related differences in career choices also persist, especially in the fields of physical science, mathematics, and engineering, where women hold only about 10 percent of jobs.[2] In 1995, women constituted about 46 percent of the U.S. labor force but only 22 percent of scientists and engineers in the labor force (National Science Foundation, 1999). In the U.K., only 6 percent of engineers in the labor force are women.[3]

One common belief is that this skewed pattern of human-capital investment and occupational choice is determined, among other factors, by social expectations, especially stereotypes and the internalization of stereotypical perceptions.[4] One view that is claimed to be stereotyped, for example, is that boys excel in certain subject areas and girls excel in others. Most subscribers to this view also believe that boys perform better in two subjects in particular, math and science. It has been found that parents (especially mothers) often entertain stereotypic gender-role beliefs about their children's abilities such that they overestimate the ability of sons

[1] In the initial college years, women are 2.5 times more likely than men to leave fields of mathematics, engineering, and physical sciences (Crocker, Major, and Steele, 1997). Girls in the U.S also have lower enrollment rates than boys in advanced science and math courses in high school (Madigan, 1997). For additional related evidence, see Hyde, Fenneman, and Lamon (1990) and Lefevre, Kulak, and Heymans (1992).

[2] See, for example, Eccles (1994).

[3] See this and more related evidence in a report from the the Equal Opportunity Commission, 2003. In U.K higher-education institutions in year 2000/01, for example, the male/female student ratio was 1.6 in physical and mathematical sciences, 4.1 in computer science, and 5.4 in engineering and technology.

[4] Psychologists who have studied stereotypes have assumed traditionally that beliefs about social groups are a powerful determinant of attitudes and behaviors toward members of these groups (Gardner, 1994). Stereotyping is thought to promote prejudice, which promotes discrimination (Dovido et al., 1996). In other words, beliefs about members of a social group are assumed to arouse liking or disliking for the group, which, in turn, dictates behavior toward group members. Recent experimental evidence shows that stereotypes and prejudice do appear to be positively associated. However, there is little evidence about the link between stereotypes and discrimination. For example, research on the relationship between stereotypes and racial discrimination has found only a modest relationship between whites' stereotypes of blacks and measures of discrimination (Dovido et al., 1996). Thus, there is very little research that demonstrates that stereotypes cause discrimination.

and underestimate the ability of daughters in math and science (Eccles, Jacobs, and Harold, 1990).[5]

On the basis of this stereotype, it is claimed, girls are encouraged to pursue traditional female studies instead of mathematics, science, and other traditionally male subject areas, and that women are steered toward certain careers that do not correspond to their interests and wishes. Such gender stereotyping, it is claimed, is often reinforced by parents and teachers. Tiedemann (2000), for example, showed that teachers' perceptions are consistent with stereotypes of gender differences, i.e., teachers believe that boys have more talent and that girls compensate by working harder. Furthermore, this bias in teacher perceptions of their students' talent and effort in math is linked to the teachers' own category-based, gender-role stereotypic beliefs about the general distribution of math talent between boys and girls.[6] Dusek and Joseph (1983) and Madon et al. (1998) are among many scholars who have shown that teachers occasionally rely on stereotypes in forming perceptions about their students.[7]

In this study, I use a natural experiment to test whether teachers have a gender-biased perception of students' cognitive ability that leads to discrimination in end-of-high -school achievement measures. In particular, the study estimates the extent of gender discrimination in scores on matriculation exams taken by high-school seniors. Since high-school matriculation in Israel has a high rate of return in the labor market and provides access to post-secondary schooling, such discrimination may cause loss in earnings over one's lifetime.

The natural experiment arises from the rules that are used in Israel to determine scores in matriculation subjects. The final matriculation score in a given subject is the mean of two

[5] Other related studies have found that parents' perceptions of the value of math and its difficulty for their children influenced both the children's confidence in their ability and the likelihood that they would enroll in advance math classes.

[6] Several studies document how stereotypes affect teachers' behavior in the classroom. Teachers give boys more praise, criticism, encouragement, and permission for strategy use. Teachers often view boys' rebellious invented strategies as signs of a promising future in math and unconsciously control girls more than boys. They apply this control by encouraging strategy use or accepting their lack of participation (Hyde and Jaffee, 1998). Carr, Jessup, and Fuller (1999) argue that in teacher-student interactions girls and boys are influenced to develop different skills, knowledge, and motivation. For example, interaction of teachers with boys often increases their understating and self-concepts in math, unlike the outcome of such interactions between teachers and girls. Rebhorn and Miles (1999) found that teachers often call on males and praise them but give girls more time for easy tasks. Middleton and Spanias (1999) report that teachers reinforce learning helplessness among girls: when teachers encounter girls who do not want to succeed, they are liable to allow them to persist in their apathetic view of mathematics. According to the National Center of Education Statistics (1997), females are less likely than males to be advised, counseled, and encouraged to take mathematics courses. Fennema and Hart (1994) found that teachers tend to structure their classrooms in ways that favor male learning.

[7] Hallinan and Sorensen (1987) found that intra-class assignments to ability groupings in school were influenced more by pupil gender than by pupils' performance in math, with boys tending to be assigned to higher ability groups. Girls with high mathematical aptitude were less likely to be assigned to a high set than boys with similar aptitudes. Intra-class grouping was found to have no effect on pupil performance compared with whole-class teaching.

intermediate scores. The first is based on "national" exams that are "external" to the school because they are written and graded by an independent agency. The scoring process for these exams is anonymous; the external examiner is not told the student's name and gender. The second intermediate score is based on a school-level ("internal") exam that mimics the national exam in material and format but is scored by the student's own teacher, who of course knows the student's name and gender. 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,[8] 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 very similar to that used by Rouse and Goldin (2000) and by Blank (1991).

Using data on all matriculation scores of several cohorts of high-school seniors in Israel, I applied this natural experiment framework to test for gender discrimination in seven subjects—two in the humanities (English and Hebrew literature), mathematics, and four in science (biology, chemistry, computer science, and physics). Section 2 of this paper shows that the distributions of the blind and non-blind scores are very similar and, in many cases, are identical. The basic results of the experiment, presented in Section 4, show that, contrary to expectations, public-school teachers in Israel discriminate against male students. The sign of the bias is the same in all seven subjects examined and in all tests in cases where there is more than one exam per subject. The extent of the bias varies by subject and test, ranging from 0.05 to 0.25 of the standard deviation of the blind-score distribution. This bias against male students, on average, doubles the gender-score gap because female students outperform male students on the national 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.

The basic results withstand several specification checks. Overall, they do not support the hypotheses that the anti-male bias in the non-blind score reflects statistical discrimination against male students. For example, limiting the sample to schools where boys on average dominates the

[8] The Ministry of Education's formal guidelines about the "internal" score state specifically that the school-level score should measure only the student's level of knowledge in the subject of study, and the school-level exam is intended to measure just that. However, schools are allowed to deviate from the score on the school exam to reflect

average performance of girls, overall or in specific subjects, leave the results basically unchanged. In another check, I restricted the estimation so that it would be based on samples stratified by levels of curriculum (basic, intermediate, and advanced) in each subject. The bias is strongest against low achievers but is measurable even for the most proficient male students. In a second check, I limited the estimation to exams for which the blind and non-blind score distributions overlapped perfectly. This restriction, too, did not change the basic findings.

An interesting and obvious question in this context is whether male and female teachers harbor different stereotypes. Data on teachers' gender was available only for English and math. Ninety-two percent of English teachers in the sample are female but one-third of high-school math teachers in Israel are male. Estimating the model for samples by teacher's gender yields a striking result: male teachers account for all the estimated bias against male students in math. However, since most teachers of other subjects are female, they probably also contribute to the discrimination against boys.

The second part of the study examined how these biases changed when teachers faced the possibility of financial loss due to stereotyped behavior. I exploited two teachers' monetary-incentive experiments that included a specific clause imposing sanctions on teachers if the blind and non-blind test scores of their students were far apart. The results of this analysis suggest that the risk of being disqualified for an award (a bonus payment) did not diminish the discrimination against male students in any way.

The final part of the paper examines two empirical insights gained from experiments in social psychology. The first asks whether discrimination motivated by stereotypes is repressed in the face of revealed information about students' ability. The performance of students on eleventh-grade matriculation exams is used as an elicitor of such information. The second insight is that a "stereotype threat"—the threat of being perceived as a negative stereotype or the fear of poor performance that would confirm the stereotype—may be powerful enough to shape the intellectual performance and academic identities of entire groups of people. In this context, I use the difference between the blind and non-blind scores in exams taken in eleventh grade as an indicator of past discrimination in order to test whether such an event affected performance in twelfth grade.

---

the student's performance on previous exams. As I show below, the distributions of the external and internal scores are very similar and are even identical on many tests.

2. Structure of the Natural Experiment

2.1 The Israeli High-School Matriculation Exam System

Israeli post primary education consists of middle school (grades 7–9) and high school (grades 10–12). High-school students are enrolled either in an academic track leading to a matriculation diploma (bagrut in Hebrew)[9] or in a vocational track leading to a high-school diploma. Bagrut is completed by passing a series of national exams in core and elective subjects beginning in tenth grade, with more tests taken in eleventh grade and most taken in twelfth grade. Students choose to be tested at various levels of proficiency, each test awarding one to five credit units per subject, depending on difficulty. Some subjects are mandatory and, for many, the most basic level of study is three credits. A minimum of twenty credits is required to qualify for a matriculation certificate. About 52 percent of high-school graduates in 2002 and 46 percent of members of the relevant age cohort received matriculation certificates.[10]

The score in each matriculation subject is a weighted average of the score on a national exam and a school-awarded score based solely or mainly on a school-level exam. The national exam is termed an "external exam" and the school exam is called a *matkonet*, derived from the word *matkon*, recipe, meaning that the school exam follows the "recipe" of the national exam. Therefore, school exams use the format of the national exam, cover the same material, and use the same types of questions. However, there is one important difference between the two exams: all national tests are graded externally, each student's exam by two evaluators. The student's identity is concealed; only his or her I.D. number appears on the exam notebook. School-level "matkonet" tests, in contrast, are graded in school by the classroom teacher and are not anonymous.

Students are admitted to post-secondary programs solely on the basis of their matriculation scores. Therefore, higher-education institutions and the National Council for Higher Education,[11] which monitors the level of difficulty of all exams and their comparability from year to year, scrutinize these tests closely. To assure long-term consistency in the national

[9] The matriculation certificate is a prerequisite for university admission and is one of the most economically important education milestones. Many countries and some American states have similar high-school matriculation exams, e.g., the French Baccalaureate, the German Certificate of Maturity (*Reifezeugnis*), the Italian Diploma di Maturità, the New York State Regents examinations, and the recently instituted Massachusetts Comprehensive Assessment System.

[10] See the Israel Ministry of Education web site (www.education.gov.il).

[11] Each higher-education institution undertakes to rank applicants according to the same formula, thus producing an index based on a weighted average of the student's average score on all his/her matriculation exams and the average score on the national exams in three core subjects (math, English, and Hebrew). Students may replace the second component in the formula with an ATS-type score from an examination administered by the National Testing Center.

exams, they are written under contract by an outside independent nonprofit organization that has been doing this for many years.

To assure comparability between school-exam and national-exam scores, the Ministry of Education has since 1996 been using a scheme of sanctions ("Differential Weighting") in cases where the school score deviates substantially from the national-exam scores. The sanctions include disqualification of the school score on the subject for all students and using only the national test scores as the final matriculation outcome for that subject. If such deviations recur in several subjects, the sanction is more severe and may lead to the disqualification of all school scores, in which case only the national exams are used to determine matriculation. Such a ban is usually imposed for a period of two to three years, after which a ministerial committee reviews the case.[12] The purpose of the system is "to insure the validity of the final matriculation scores so that they may be viewed as a unified and credible criterion in the process of admission to higher-education institutions" (Ministry of Education, High School Division web site). The Ministry guidelines to schools indicate that the school score should be submitted to the Ministry before the national matriculation exams and should reflect the student's knowledge of the subject. The exam on which the school score is based is administered several weeks before the national exam.

2.2 *Comparability of School- and National-Exam Scores*

Table 1 presents the mean and the standard deviations of the school and national test scores for a representative sample of tests in seven subjects: six math exams (two at the basic level, two intermediate, and two advanced), four English exams (two at the basic level and one apiece at the intermediate and the advanced level), five in biology (all part of the advanced program), two in Hebrew literature, two in chemistry, three in physics, and one in computer science. The mean gap between the school and national test scores is in most cases positive and significantly different from zero. The standard deviation of the school score is also generally 10–15 percent smaller than in the national exam. In several subjects, however, the internal score is lower than the external score; there are also cases in which the extent of score variation is similar in both cases.

Figures 1–7 present kernal-density estimates of the school and national exams of some of the tests presented in Table 1. The distributions of the external and the school scores are very

[12] A Ministry of Education document describes the rules of the Differential Weighting scheme in detail. For example, if the school score is higher than the national test score by twenty points or more or if it is lower by ten points or more, the case is considered an outlier. If the probability of such an event is 1:10,000, the weights of the two scores will be 30 percent and 70 percent instead of 50 percent each. If the probability of such an event is 1:1,000,000, the two scores are weighted at 10 percent and 90 percent, respectively. If more than 8 percent of scores

different across subjects and tests. Some do not resemble a normal distribution shape. A striking feature in all the figures, however, is the overall similarity between the distribution of the internal and the external scores for each given test in comparison with the large differences between and within subjects. In other words, the distributions of the internal and external scores in each exam change in a similar way from one test to another. In most cases, the school-score distribution is a constant shift, to the right, of the national-exam test score distribution. The direction of the shift is expected after one observes the pattern in differences in means, as shown in Table 2. The figures also reveal that the internal score distribution in many exams is "thinner" at the left tail of the distribution. In many tests, however, the two distributions are almost completely identical, a feature that we exploit in the analysis presented in the sections to come.

2.3 The Blind–Non-Blind Test Score Gap by Gender

Figures 1–7 also present kernal-density estimates of the school and national score distributions by gender. Both score distributions for each test are similar for male and female students, respectively. However, the leftward shift of the school-score distribution relative to the external-score distribution is always larger for female students than for male students. Even in cases where the shift factor is negative, i.e., where the school-score distribution is to the left of the external-score distribution (as occurred in three biology tests, several math exams, and several cases in which the two distributions intersected), the gap is less negative for female students than for male students. The next section estimates these gaps and their statistical confidence intervals more precisely and subjects the main results to several specification tests.

3. The Data

The data used in this study pertain to the school years 2000–2002. The micro student files included the full academic records of each student on matriculation exams taken during high school (grades 10–12) and student characteristics (gender, parents' schooling, family size, and immigration status, i.e., students who recently immigrated) for the three cohorts of high-school seniors in 2000–2002. The information for each matriculation exam included its date, subject, applicable credits, and score. Each matriculation exam is written by members of the Ministry of Education staff and experts from an independent agency. There are two exam "seasons," winter (January) and summer (June), and all students are tested in a given subject on the same date. The exams are graded centrally, each exam by two independent external examiners, and the final

---

were found to be outliers in two of three consecutive years, the ministerial committee may prohibit the school from submitting school scores and advertise its decision in the national print media.

external score is the average of the two. This protocol eliminates the possibility of teachers grading their own students' exams and, thereby, reduces the possibility of cheating.

The school data provide information on the ethnic (Jewish or Arab) nature of each school and its religious orientation (secular or religious) of the Jewish schools. In this study I used a sample that includes only Jewish secular schools, which account for about 60 percent of all schools and students. I excluded the Jewish State-Religious schools and the Arab schools because many of them are either all-male or all-female and in many others classes are segregated by gender. This unique feature may be correlated and confounded with different patterns of gender stereotyping and discrimination in comparison with secular schools.

Table 2 shows the proportions of male and female students by subject and level of study. Advanced literature classes are dominated by women. Biology and chemistry also have majorities of female students. Male students have a higher representation in advanced computer science and in math. Generally speaking, this pattern resembles the pattern observed in the U.S. and the U.K.

4. Estimating Gender-Stereotyped Discrimination

I took advantage of the blind and non-blind nature of the scoring procedures for the external and school exams, across subjects and tests, to identify the effect of the procedure of anonymous evaluation of cognitive ability on the likelihood that male or female abilities would be systematically under- or over-valued. The score of student i on test j is a function of gender (M), whether the evaluation is anonymous (B), and individual and school characteristics (X). The repetitive structure of two scores in each subject (and across various tests in each subject), one blind and the other non-blind, made it possible to use a differences-in-differences estimation strategy. Assuming a linear specification, the score equation may be written as

(1) $$S_{ijb} = \alpha + \lambda M_i + \delta B_{ijb} + \gamma \left(M_i \text{ x } B_{ijb} \right) + X_i \theta + u_{ijb}$$

The coefficients for M and B identify the effects of being male and of a blind scoring procedure, respectively, on the test score. The parameter of interest is that for the interaction between M and B, γ, which measures the difference between the internal scores of male students and those of female students, given the respective difference in the external score. The differences-in-differences nature of the estimation of Equation (1) implies that individual and school fixed effects are implicitly assumed away in this model with regard to the estimated coefficient of interest, γ, as long as they have the same effect on the blind and non-blind scores.

Table 3a presents the estimated parameters for Equation (1) in seven subjects—English, Hebrew literature, math, biology, chemistry, computer science, and physics—obtained from data for 2001. In each subject, the data set is a stacked file including the internal and external scores for each of the exams in the respective subject. All test scores were standardized to a distribution with zero mean and a standard deviation equal to one. This procedure was applied within subjects to each test separately. The sample size varied by subjects, including most students in compulsory subjects but much fewer in elective subjects. Each equation included student's characteristics as controls (mother's and father's years of schooling, number of siblings, immigration status, and a set of dummy variables for ethnic origin—Asia/Africa, America/Europe, former Soviet Union, Ethiopia, and Israel-born)—and two average measures of achievements on external matriculation tests taken in tenth and eleventh grade (lagged outcomes). The two measures are the mean credit-weighted average score on all previous matriculation exams (coding zeros for those who had taken no exams) and the respective overall number of credits. These measures are powerful predictors of students' success on each matriculation exam in twelfth grade in each subject. Table 3 reports only the estimates for the variables of interest in this study (male gender, non-blind test, and the interaction of the two). The standard errors reported in the Table 3 are adjusted for clustering, using formulas set forth by Liang and Zeger (1986).

4.1 Empirical Results

Overall, female high-school students in Israel have higher achievements on the national matriculation exams (blind tests) in all subjects presented in Table 3a except for English. Girls have advantages of 0.10 of a standard deviation of the external-score distribution in math, 0.14 in biology, and 0.475 in Hebrew literature. In physics (0.02), and computer science (0.04), the advantage of female students is positive but not statistically different from zero. The advantage of male students in English is 0.11 of a standard deviation with a 0.014 estimated standard error. The male coefficient may reflect the selective distribution of students among elective subjects and among levels of study (basic, intermediate, and advanced) in compulsory subjects; therefore, these estimates may be biased. However, as will be shown below, the advantage of female students recurs at all levels of study, suggesting that the selection bias may not be very important. For example, girls have on average higher external scores at all three levels of study: basic, intermediate, and advanced.

The mean differences between school scores and national scores, which are conditional on gender and on the interaction between gender and non-blind testing, are very small and seldom significantly different from zero. The largest differences are in English (0.048 estimated standard error) and biology (–0.059 estimated standard error) and in both of these cases they are significantly different from zero.

The main parameter of interest is the estimated coefficients of the interactions between the gender indicator for male students and the non-blind test indicator. These estimates are negative and significantly different from zero in all seven subjects. The highest estimate is in English (–0.180 estimated standard error), the lowest is in literature (–0.053 estimated standard error), and in four of the seven subjects the estimate exceeds one-tenth of a standard deviation in the external-score distribution.

As noted above, adding student, school, or exam fixed effects leads to exactly the same estimates of the interaction between gender and non-blind test because the basic specification of Equation (1) saturates all these fixed effects, since the difference in estimation of the differences occurs at the student level within each subject and exam. These results are presented in Table 3c.

The signs of the estimated coefficients for the interactions of the male and non-blind test indicators are contrary to common perceptions and beliefs. The results obtained from the 2000 and 2001 data are very similar to the findings presented in Tables 3a–3c; therefore, they are not presented in this paper. Insofar as coefficients reflect a bias in the perception of cognitive ability by students' gender, due to stereotypes or other sources of discrimination, then the evidence is that such stereotypes act against male students and not female students. Teachers favor female students by consciously or unconsciously "inflating" their scores on non-blind tests. The direction of the bias enhances the advantage that female students have on the blind test. In English, the bias in favor of female students more than offset these students' "real" disadvantage, as reflected on the blind test (0.180 versus 0.114).

Table 3b presents the results when the percentile-rank score is used as a dependent variable instead of the standardized score. The evidence in this table is very similar to the results in Table 3a.

4.2 Specification Checks

A real "threat" to the interpretation that the estimated anti-male biases, presented above, represent stereotyped perceptions among teachers is the possibility that the blind and non-blind tests measure different abilities or that one of the two reflects attributes not included in the other. A most obvious example would be that the estimated biases reflect non-conformist behavior by male students that the teachers do not like, e.g., absenteeism or discipline problems in the classroom. If teachers adjust scores to reflect such undesired behavior, even though the scores should reflect only the students' knowledge of the subject, a discrepancy between male students' blind scores and non-blind scores would take shape. Although this interpretation may be consistent with the foregoing evidence, other results presented below suggest that the likelihood of this being the source of the bias is very small.

One way to determine whether the aforementioned results reflect a difference in the abilities measured by the two scores is to restrict the empirical analysis to exams for which the distribution of blind and non-blind scores seems absolutely identical. Several of the distributions in Figures 1–7 meet this criterion: intermediate math 3, advance math 2 and 3, basic literature 1, advanced physics 2 and 4, and intermediate and advanced English. Table 4a presents the estimates of the effect of the interaction between the gender and blind-test indicators in each exam within this subset. The results from the sample of these exams are very similar to results based on the pooling of all exams into one sample in each subject: negative and significant estimates of interaction between the male-gender and the non-blind score indicators. This suggests that even in cases where the two score distributions are unquestionably identical, in which it may safely be assumed that they measure the same outcome, the bias is in the same direction and of similar magnitude.

To check the possibility that the non-blind score reflects male students' behavioral problems, one may restrict the sample to students who are least likely to have discipline problems. Students who enrolled in advanced levels of study in each subject, comparable to participants in honors classes in the U.S., may constitute such a sample. I estimated Equation (1) separately for each exam at the basic and advanced levels in each of the seven subjects.[13] The estimated coefficients of the interaction between the gender and blind-test indicators are presented in Table 4b. The sign of the estimated coefficients in the advanced level tests is always negative, significantly different from zero, and—in some cases—even larger than the coefficients estimated from the sample for the basic level of exams.

Another alternative interpretation to the estimates presented thus far may be that they reflect the effect of interactions between the non-blind test variable and other variables that are correlated with gender. To assess this possibility, Equation (1) was augmented by adding the interaction terms of the non-blind indicator with the following dichotomous indicators of students' socio-demographic characteristics: recent immigrant, father with more years of schooling than the school mean, mother with more years of schooling than the school mean, more siblings than the school mean, and ethnic origin. Table 4c presents the results after all these additional interactions were included in Equation (1). First, it should be noted that the inclusion of the nine additional interactions along with the non-blind test indicator does not change the estimated coefficient for the interaction between gender and non-blind test. The coefficient in the math equation, for example, changed from –0.086 to –0.087; in English the change was from –0.180 to –0.176. Second, one of the newly added interactions, that between the non-blind test and immigration status, was positive and significant. The other interactions elicited no systematic pattern in terms of sign or significance level.

Another interpretation of the estimated anti-male bias in the non-blind score, which may be motivated by the average superior performance of girls in the national exams, is statistical discrimination. The blind and the non-blind scores are two sources of information about students' cognitive ability. If teachers are influenced by the expected higher performance of girls on the national exams (as shown in these data), even if male and female students perform at the same level on the school exams, they will receive different scores. In this case, then, the estimated anti-male bias on the non-blind scores may reflect simple statistical discrimination (Cain, 1986).

Statistical discrimination against male students may also occur even if there are no real differences in cognitive performance between boys and girls. This will happen if teachers believe that the non-blind score is a less reliable signal of knowledge for boys than for girls (Aigner and Cain, 1977). Such beliefs may arise in a school context if, for example, male students are known or perceived to cheat more often than girls on school exams. The question of whether such perceptions are based on real evidence or unfounded is irrelevant for the outcome of statistical discrimination.

Some of the evidence presented in Tables 3–4 does not support the interpretation of statistical discrimination. In English, in particular, boys outperformed girls on the national exams by a wide margin but faced bias in their school scores. Similar contrasting comparisons were found in some tests in other subjects as well. In advanced math, for example, boys had a higher blind-score average than girls but sustained a bias in the school score. It is possible, however, that teachers form their expectations about gender differentials in true cognitive ability on the basis of overall performance in all subjects and not only in the subject that they teach. To assess this possibility, I estimated the models separately in two distinct environments: a sample of schools where boys outperform girls on average and a second sample of schools where girls do better on average than boys. First I computed average performance on the basis of matriculation

[13] In several subjects, there is more than one basic or advanced-level exam. The basic curriculum in math (3 credits), for example, involves two exams, one for the first credit and a second for the other two. In cases where

exams taken in eleventh grade. Table 5a presents the results of Estimation Equation (1) for the two samples of schools based on this measure of average performance by gender. I also used school average performance by gender on the basis of all matriculation exams taken by members of the 2000 graduating class. These results are presented in Table 5b.

Focusing first on Table 5a, we see clearly that in the sample of schools where girls outperform boys on average (the upper panel of Table 5a), the male coefficient is negative in all subjects except English and the estimated bias, again, is against boys. The interesting results, however, are seen in the lower panel of Table 5a. In this sample, boys had a higher average external score on twelfth-grade matriculation exams in three subjects: English, computer science, and physics. The male coefficient was negative in the other subjects, but in two of the four it was not significantly different from zero. The most noteworthy result, however, is that in five of the seven subjects the bias against male students is negative and significantly different from zero. In some cases, the bias estimates are even higher in this sample than in the sample of schools when girls dominate boys in average performance.

The results presented in Table 5b generally confirm those in Table 5a. Overall, they do not support the hypotheses that the anti-male bias in the non-blind score reflects statistical discrimination against male students.

4.3 Variance of Discrimination in Accordance with Student Ability

Who, in terms of ability, are the male students who encounter discrimination? Using lagged achievements as a measure of ability, I estimated models that allow the estimated coefficients of discrimination to vary with lagged outcomes. In particular, I allowed the non-blind score indicator to interact with the mean credit-weighted average score on all previous matriculation exams. Using this average score, I coded dummies for each half of the score distribution.

Table 8 reports results of the estimation of Equation (1), which allows the effect of the male and non-blind score interaction to be different for students who are above and below the mean of the lagged average score. An interesting pattern is seen in Table 6: in the two humanities subjects (English and literature), there is no difference between the two estimated coefficients, meaning that below-average and above-average students are equally discriminated against. In all science subjects, however, the coefficient of discrimination is practically zero for above-average students and is high and significant for below-average students. In these subjects, the discrimination factor ranges from one-tenth (math) to one-third (computer science) of a standard deviation in the respective blind-score distribution.

---

there is more than one exam, Table 5 presents results for an arbitrarily chosen exam.

4.4 Can Female Anxiety in 'Money Time' Explain the Male Negative Bias?

Recent studies suggest that women may be less effective than men in competitive environments, even if they are able to perform similarly in non-competitive environments (e.g., Gneezy, Niederle, and Rustichini, 2003). One may argue that national exams create more stress and anxiety than school exams. If teachers expect these factors to have a larger effect on girls than on boys, they may compensate female students for their expected underachievement on the external test by giving them higher school scores. First, it should be noted that the scores on the school exam and the national exam are equally important, both weighted equally in determining the final matriculation score.[14] Therefore, they should cause similar levels of stress and anxiety among students, if they cause any stress at all. Second, in the foregoing results, girls outperformed boys on national exams in almost every subject and at almost every level of curriculum. Third, the evidence that girls "choke" under pressure or in a competitive environment refers only or mainly to situations where stereotyping is a real threat or when tasking is against the opposite gender. Experiments have shown that girls' performance is not affected even in high-stake tests as long as there is no threat of gender stereotyping (e.g., Stangor and Sechrist, 1998). Gneezy, Niederle, and Rustichini (2003) show that the increase in women's performance in a competitive environment is limited only when the rivals are men. The national-exam environment, however, is free of a stereotype threat and any obvious and open competition against the other sex because the scoring is blind. Therefore, there is no reason to expect these factors, even if they are real, to have any gender-differentiated effect. Finally, if girls do suffer more than boys from high-stake exam anxieties, we would expect to find a high correlation between blind versus non-blind scores differences experienced in exams in previous grades and the respective differences in exams in twelve grade. In Section 7, we present data that allow us to examine such correlations in a more general context. The results presented there do not support the "anxiety" hypotheses as an explanation of the anti-male bias.

4.5 Gender of Teacher and Stereotyped Discrimination

Most high-school teachers in Israel are female. This means that the gender bias against male students, reported above, most likely reflects the behavior of female teachers. Still, it is of interest to examine whether the bias against male students varies by teachers' gender in subjects that have a relatively large proportion of male teachers. Data on the gender of English and math teachers were available for a sample of forty secular Jewish high schools. In this sample, about one-third of math teachers but only 8 percent of English teachers were male.

Table 7 presents the results of Estimation Equation (1) in math and English for samples stratified by male and female teachers. The evidence in Column (1) shows that the average bias against male students in math in this sample (–0.089) is almost identical to the bias estimated in the full sample (–0.086, as reported in Column 3 of Table 3a). In the sample where teachers' gender is identified, however, all of the bias is associated with the behavior of male teachers, at –0.209 as against –0.030 for female teachers, a result that is also not significantly different from zero.

In English, the bias in the sample used in Table 6 is again similar to the bias estimated in the full sample, as reported in Tables 3a–3c. In this case, however, the male- and gender-estimated biases are very similar, although, notably, only a small fraction of English teachers are male, which may explain the high estimated standard errors in these regressions.

I also asked whether the male-teacher bias in math varies in accordance with teacher characteristics such as age, education (B.A. versus M.A.), and marital status. I found no systematic patterns in the estimated discriminatory behavior of male teachers against male students.

5. Do Financial Incentives Affect Stereotyping and Discrimination?

The research literature in social psychology suggests that stereotyping and discrimination, although partly automatic, are individually controllable and responsive to social structure. In this section of the paper, I ask whether the estimates of stereotype discrimination by gender in the evaluation of cognitive ability, presented above, change in a situation where such behavior becomes costly to the evaluator. For this purpose, I use data from two Israeli programs that rewarded high school teachers with cash bonuses for improvements in their students' performance on high-school matriculation exams. The first program rewarded all teachers in a given school, as a group, on the basis the students' average performance. Performance was measured as the value-added of two main outcomes, the average dropout rate and the school-average matriculation rate. The program was implemented in the form of a rank-order tournament among schools. The program included sixty-eight schools in its first year (1995) and

[14] Interestingly, in France, school exams are also an integral part of the matriculation system but the final scores are based solely on the national exams. The school exams are called by a name that has the same meaning as the Hebrew name of the corresponding exams in Israel.

was expanded to eighty schools in its last year (2001). The monetary bonuses were modest—up to $1,000 per teacher. For further details of this program, see Lavy (2002). The second program offered individual monetary incentives to English, Hebrew, and mathematics teachers in forty-nine high schools in the 2001 school year. This program was also structured as a rank-order tournament among teachers, in each subject separately. Teachers were rewarded on the basis of their performance relative to other teachers of the same subjects. The relative measurements were based on comparison of the achievements of each teacher's students with predicted values using regressions. Two measurements of students' achievements were used as indicators of teachers' performance: the passing rate and the average of the blind and non-blind scores on each matriculation exam. The total sum to be awarded in each tournament was predetermined and individual awards were determined on the basis of rank and a predetermined award scale: first place—$7,500; second place—$5,750; third place—$3,500; and fourth place—$1,750. These awards are significant relative to the mean gross annual income of high-school teachers ($30,000) and the fact that a teacher could win several awards in one tournament if he or she prepared more than one class for a matriculation exam. The program included 629 teachers, of whom 207 competed in English, 237 in mathematics, 148 in Hebrew or Arabic, and 37 in other subjects that schools preferred over Hebrew. Three hundred and two teachers won awards—94 English teachers, 124 math teachers, 67 Hebrew and Arabic teachers, and 17 in other subjects. Three English teachers won two awards each, twelve math teachers won two awards each, and one Hebrew teacher won two first-place awards totaling $15,000.[15]

Both bonus programs included an explicit stipulation about sanctions that would apply to teachers who violated the official Ministry Differential Weighting rules in regard to discrepancies between school-level and national-level scores, as described in Section 3 above. The stipulation determined that such teachers would be disqualified from both bonus programs. Consequently, teachers for whom 7 percent or more of students had a large discrepancy between scores would be ineligible for bonus payments. Therefore, we used the sample of teachers in these two programs to test whether the discrimination against male students was lower when a cost was attached to stereotyped behavior.

[15] For further details, see Ministry of Education, High School Division, "Individual Teacher Bonuses Based on Student Performance: Pilot Program," December 2000, Jerusalem (Hebrew). Lavy (2003) presents evidence on the effect of the experiment on teachers' effort and productivity and on students' achievements.

Table 8a presents results for a sample of students whose English and math teachers participated in the individual bonus experiment. As a comparison group, the table also shows results for all other students. The evidence pertains to the school year in which the bonus program was implemented, 2001, and also for the pre-program year, 2000, which will allow for differences-in-differences comparison. Note, however, that the bonus program allows us to study the behavior of English and math teachers only. The estimate of the coefficient for interaction between the gender and non-blind score indicators is largely identical between bonus-program teachers and all other teachers: –0.076 versus –0.086 in math and –0.154 versus –0.182 in English. The estimates for the program teachers (the first estimate in each pair) are somewhat lower, but a similar pattern is observed from the results of the pre-program year, 2000. The comparison of the evidence before and after the program gains importance given that the results in the table also suggest that students in the bonus program have lower achievements in math and English than other students in Israel's secular high schools.

The evidence in Table 8a suggests that imposing a cost for stereotyped discrimination did not decrease the incidence of this social phenomenon among English and math teachers. However, this non-sensitivity may also be results of low teacher awareness of the differential-weighting stipulation that was built into the bonus program, even though a post-program survey among the teachers indicated that 80 percent were familiar with the program and its rules.[16]

Tables 8b and 8c present results for a sample of students whose teachers participated in the group bonus experiment. Since the cognitive performance measures used to rank schools included the average matriculation rate and the average number of credits in all subjects, we may examine the effect of the incentives on teachers of each of the seven subjects studied in this paper. We present data for the years 2001 (Table 8b) and 2000 (Table 8c). The sample is very small in several subjects (chemistry and computer science in particular). The estimated coefficients for the interaction between male gender and non-blind test score are negative and significantly different from zero in all subjects except biology, in which the estimate is positive but imprecisely estimated. The results for the English, math, and literature teachers resemble those obtained from the overall sample. The estimates for the science teachers are generally higher than those obtained from the full sample (except in biology, as noted). These results again support the conclusion that stereotyped discrimination did not succumb to the threat of financial

[16] We conducted a follow-up survey of teachers in the program during the summer vacation following the school year. Seventy-four percent of teachers were interviewed. Very few of the intended interviewees were not interviewed; most of these cases were due to wrong phone numbers or failure to reach teachers by telephone despite several attempts. The survey results show that 92 percent of the teachers knew about the program, 80 percent had been briefed about its details—almost all by their principals and the program coordinator—and 75 percent thought

cost, but these results may also reflect the fact that the bonuses were modest and had the nature of group incentives that are susceptible to free-rider behavior.

6. Do Stereotypes Succumb to Revealed Information about Student Ability?

Since some students take matriculation exams in eleventh grade, some twelfth grade teacher have a precise signal about their ability, i.e., the score on the external exam taken in the same subject in eleventh grade. This situation is most relevant in math, since many schools (270 out of 363) administered a math matriculation exam in eleventh grade as part of their math curriculum. The rules concerning school scores and external scores apply to matriculation exam taken in eleventh grade as well as they do to those taken in twelfth grade. Thus, the twelfth-grade math teachers have ample opportunity to become highly familiar with each student in their class. This has important implications for stereotyping. According to recent studies (Kunda and Thagard, 1996; Oakes et al., 1994), perceivers rarely use stereotypes to judge individuals if they have abundant information about them.

The issue, then, is whether the bias against male students is lower when this information is available to the twelfth-grade teacher than when it is not. Some 23,170 students from 270 schools took a math matriculation exam in both eleventh and twelfth grades. Column 1 in Table 9 presents the estimates of Equation (1) on the basis of this sample. On the external exam, the score of male students was 0.093 of a standard deviation lower than that of female students. The gender gap in this case was identical to that in the full sample (Column 3 in Table 3). The gender bias against male students in this sample was negative and significant and was even higher than in the overall sample, –0.118 versus –0.086. This implies that the possession of information from a previous exam in the same subject that is clearly indicative of the student's true ability, and that is known to the student, his parents, and his classmates, does not lead to the elimination of discrimination.

Columns 2 and 3 in Table 9 present results from two samples. The first sample corresponds to the sample of students whose eleventh-grade math scores surpassed the school-level mean; the second is composed of students below the mean. Columns 4–6 present the results for three sections of the external-score distribution. Table 9 shows clearly that although the bias is lower for students who performed better than others on previous math exams, these students still experienced a bias in twelfth grade of about –0.080 of a standard deviation. The bias against those at the bottom of the eleventh-grade distribution was one-fifth of a standard deviation.

---

that the information was complete and satisfactory. Almost 70 percent of the teachers were familiar with the award criteria.

The results presented in Table 9 include evidence that strengthens the case against the statistical-discrimination interpretation discussed in Section 4.2. The estimate of the non-blind intercept presented in the table is higher for students who were below the average of their class or in lowest third of this distribution, implying that students who did poorly on the national math exam in eleventh grade received (or were "compensated") by a higher school score. This result is opposite the result that one would expect were the statistical-discrimination interpretation valid.

7. Do Students React to Stereotyped Discrimination?

The "stereotype threat" theory (Steele, 1997) focuses on the consequences for individuals who contend with negative stereotypes related to their intellectual abilities. Stereotype threat is the psychological burden imposed by stereotype-based suspicions of inferiority in achievement. Stereotype threat has been shown to undermine academic achievement in two ways: by interfering with performance in mental tasks and, over time, by encouraging students to protect their self-esteem by disengaging from the threatened domain. Wanting not to perform badly, another possible result of stereotype threat, has been shown in experiments to impair performance in difficult cognitive tasks, either by simply distracting the performer or by eliciting a self-protective withholding of effort. Spencer, Steele, and Quinn (1979) showed that stereotype threat impairs intellectual functioning even in a group that is unlikely to have any sense of collective inferiority.

The social-psychology literature on stereotypes and discrimination suggests that students may react to a stereotyped situation in a variety of ways but that three ways are the most common. The first is avoidance of the stereotype situation. Female math majors, for example, may "dress down"—wear less markedly feminine clothing—in math classes than in humanities classes (Hewitt, 1996).[17] The second is confirmation of the stereotype, e.g., to perform below actual ability in a test that involves a stereotype threat. Such underachievement may result, for example, from anxiety and evaluation apprehension (Aronson, Quinn, and Spencer, 1998). The third prevalent response a stereotyped situation is an attempt to refute the stereotype by making more effort.

Students' own perceptions of a stereotyped situation may be based on experience. For example, students who on a previous exam had a school score that was lower than the external blind score may blame some of the discrepancy on stereotype discrimination. According to the theories presented above, such students may respond in such a way as to refute the stereotype and make an effort to do better on the next exam. Alternately, they may succumb to the stereotype and do worse on the second external exam than on the first, thus confirming the stereotype, as it were. Some insight into these two possible reactions to a previous stereotyped signal may be gained by estimating Equation (1) using two samples based on the stereotyped signal that the students experienced in the past—the sample of students who had a positive difference between their blind and non-blind scores on the eleventh-grade math matriculation exam, and a second sample including those who had a negative difference between the two scores.

Table 10 reports the results of the estimation of Equation (1) on the basis of these two samples. Column 1 presents, for comparison, the results based on the sample of students who had math matriculation scores in both eleventh and twelfth grades (identical to Column 1 in Table 11). Column 2 presents results for students who had a positive difference between the two scores and Column 3 shows results for students who had a negative difference. The two respective estimated coefficients, –0.103 and –0.119, both precisely estimated, are not very different. Therefore, they do not provide support, in this limited sample and specific context, for the hypotheses that students' cognitive performance is affected by what may be interpreted as a previous stereotyped experience.[18]

8. Conclusions

Recent convincing evidence suggests that women face discrimination in the labor market in terms of both employment opportunities and wages (e.g., Goldin and Rouse, 2001). However, the question of whether discrimination against women is also partly responsible for the observed gender differences in human-capital investment and occupational choice has not been directly addressed. This paper addressed itself to one important alleged form of discrimination against women that may cause some of the skewed gender pattern in productivity-enhancing investments: discrimination against girls in schools by teachers. This discrimination, it is argued, emanates from a stereotyping of cognitive ability that causes female students to underperform in,

---

[17] This suggests that in situations where women feel at risk of confirming a negative gender stereotype, they take steps to avoid projecting stereotypically feminine traits, thereby reducing the risk of being viewed through the lens of stereotype and being treated accordingly (Aronson, Quinn, and Spencer, 1998)

[18] We should note here that convergence toward the mean in test scores may confound in the data the reaction of students to past stereotyped experience. Such a convergence, however, would act in the direction of reducing the twelfth-grade difference between the blind and the non-blind scores: students whose eleventh-grade non-blind score was randomly lower than their actual potential might have a higher non-blind score in twelfth grade, one that approximates more closely or exceeds their actual potential. This would lower the difference relative to the blind score. However, convergence toward the mean may be relevant on the external exam as well, complicating the implications of this statistical process in this case.

and to shy away from, math and science subjects in secondary and post-secondary schooling—a state of affairs that also affects occupational choice, of course.

The evidence presented in this study does not confirm the commonly held belief that schoolteachers discriminate against female students. On the contrary: on the basis of a natural experiment that compared two evaluations of student performance—a blind score and a non-blind score—the bias estimated was clearly against boys. This direction of the bias was replicated in all seven subjects of study, in humanities and science subjects alike, at various level of curriculum of study, among underperforming and best-performing students, in schools where girls have a higher average performance than boys and in schools where boys are better on average. The anti-male bias among teachers widened the gap between male and female students because the latter, on average, outperform the former in almost all subjects.

The magnitude of the bias against male students is relatively large and may harm students first by causing their final scores in some matriculation exams to fall below the passing mark and, thereby, by failing to qualify them for a matriculation diploma. This failure may block students, at least temporarily, from post-secondary schooling, implying lower schooling attainment. Second, since admission to various university department is based solely on the average score on the matriculation exams, the bias against boys lowers their average matriculation score and, by so doing, may reduce boys' chances to admission their preferred fields of study. Both effects have negative future labor-market consequences.

9. References

Aigner, Dennis J., and Glen C. Cain. 1977. “Statistical Theories of Discrimination in the Labor Market.” Industrial and Labor Relations Review, Vol. 30 (2), January, 175-87.

Angrist, Joshua D., and Victor Lavy, “Achievement Awards for High School Matriculation: Evidence from Randomized Trials”. NBER Working paper 9389, December 2002.

Aronson J., D. Quinn, and S. J. Spencer, “Stereotyping Threat and the Academic Underperformance of Minorities and Women,” in Swim, J. K., and C. Stanger (eds.), Prejudice, Academic Press, 1998.

Blank, Rebecca, “The Effects of Double-Blind versus Single-Blind Refereeing: Experimental Evidence from the American Economic Review,” American Economic Review, December 1991, 81 (5), pp. 1041–1067.

Cain, Glen C. 1986. “The Economic analysis of Labor Market Discrimination: A Survey.” In Orley Ashenfelter and Richard Layard, eds., Handbook of Labor Economics (Amsterdam: North Holland), 693-786.

Carr, M., and D. L. Jessup (1997), “Gender Differences in First-Grade Mathematics Strategy Use: Social Metacongitive Influences,” Journal of Educational Psychology, 89(2), pp. 318–328.

Deaux, Kay, and Marianne LaFrance, “Gender,” in The Handbook of Social Psychology, vol. II, chap. 17, McGraw-Hill, 1998.

Equal Opportunity Commission, Facts About Women and Men in Great Britain 2003, January 2003, <www.eoc.org.uk>.

England, Paula, “The Failure of Human Capital Theory to Explain Occupational Sex Segregation,” Journal of Human Resources, Summer 1982, 17 (3), pp. 358–370.

Fennema, E., and L. E. Hart (1994), “Gender and the JRME,” Journal for Research in Mathematics Education 25(6): pp. 648–659

Fiske, T. Susan, “Stereotyping, Prejudice, and Discrimination,” in The Handbook of Social Psychology, vol. II, chap. 25, McGraw-Hill, 1998.

Gneezy Y., M., Niederle and A. Rustichini,. “Performance in Competitive Envronments: Gender Differences,” Quarterly Journal of Economics, August 2003, 1049-1074.

Goldin, Claudia, and Cecilia Rouse, “Orchestrating Impartiality: The Impact of “Blind” Auditions on Female Musicians,” American Economic Review, 90 (4), September 2000, pp. 715–742.

Hallinan, M. T., and A. B. Sorensen (1987), “Ability Grouping and Sex Differences in Mathematics Achievement,” Sociology of Education 60 (April), pp. 63–72.

Hyde, J., and S. Jaffee, "Perspectives from Social and Feminist Psychology," Educational Researcher, 27 (5), pp. 14–16.

Jones, Melinda, Social Psychology of Prejudice, Prentice Hall, 2001

Madigan, T., 1997, Science Proficiency and Course Taking in High School: The Relationship of Science Course-Taking Patterns to Increases in Science Proficiency Between 8th and 12th Grades, National Center for Education Statistics, NCES 97-838. Washington, DC: U.S. Department of Education.

Madon, S., L. Jussin, S. Keiper, J. Ecles, A. Smith, and P. Palumbo, "The Accuracy and Power of Sex, Social Class, and Ethnic Stereotypes: A Naturalistic Study in Person Perception," Personality and Social Psychology Bulletin, 1998, pp. 1304–1318.

Middleton, J., and P., Spanias. "Motivation for Achievements in Mathematics: Findings, Generalizations, and Criticisms of the Research" Journal of Research in Mathematics Education, 30 (1), 65-88.

Lavy, Victor, "Evaluating the Effect of Teachers' Group Performance Incentives on Pupils' Achievements," Journal of Political Economy, 110(6), December 2002, pp. 1286–1317.

Lavy, Victor, "Paying for Performance: The Effect of Teachers' Financial Incentives on Students' Scholastic Achievements," CEPR Working Paper, number ?, 2003.

Liang, Kung-Yee, and Scott L. Zeger, "Longitudinal Data Analysis Using Generalized Linear Models," Biometrika 73 (1986), pp. 13–22.

National Science Foundation, "Women, Minorities, and Persons with Disabilities in Science and Engineering: 1994," Arlington, VA, 1994 (available on-line at <www.nsf.gov/sbe/srs/wmpdse94>).

National Science Foundation, "Women, Minorities, and Persons with Disabilities in Science and Engineering: 1994. Arlington, VA, 1998 (available on-line at <www.nsf.gov/sbe/srs/nsf99338>).

Spencer, S. J., Steele, C. M., and D. M., Quinn. "Stereotype Threat and Women's Math Performance." Journal of Experimental Social Psychology, 1999, 35, 4-28.

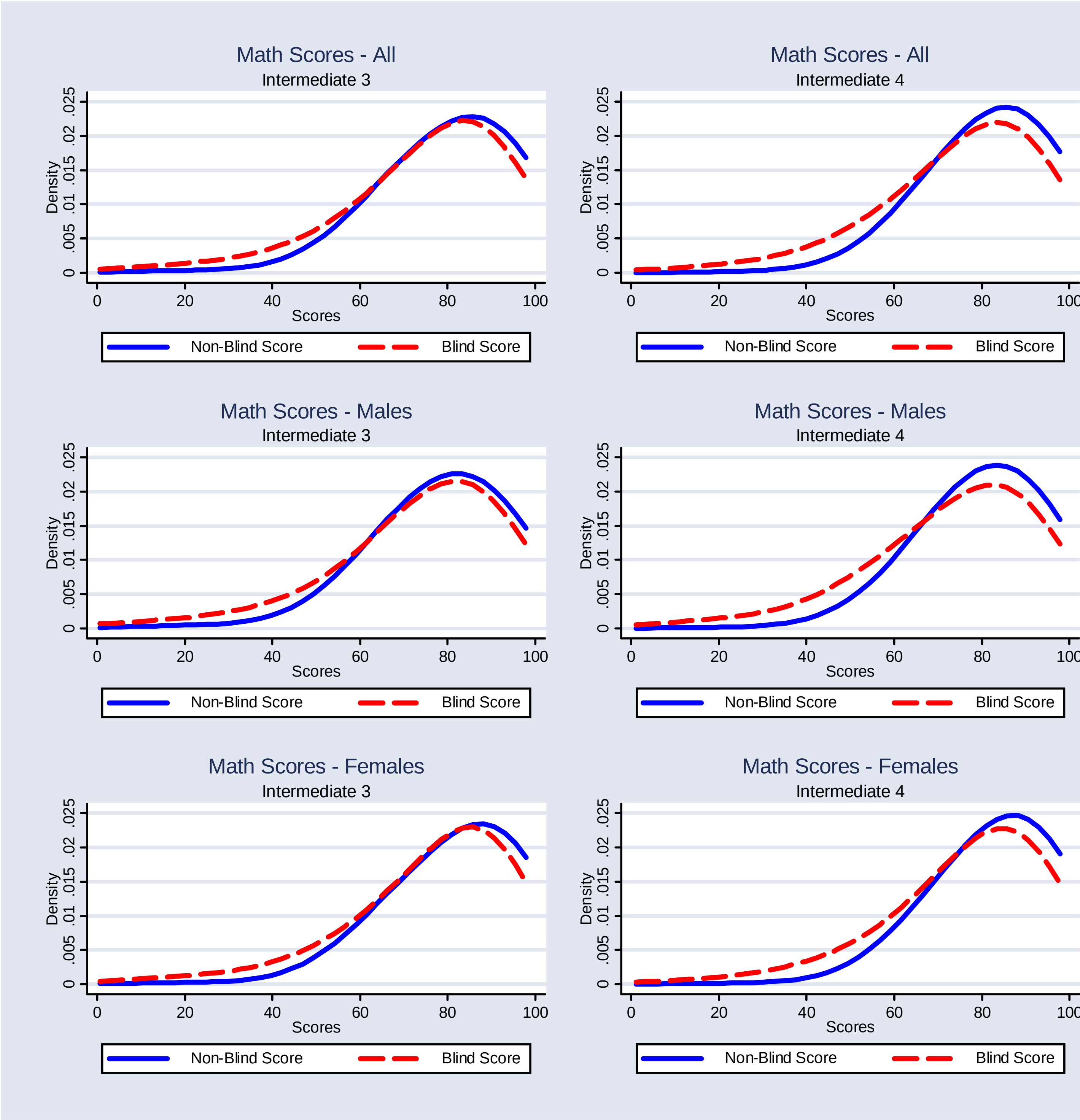
Stangor, C., and G. B., Sechrist., "Conceptualizing the Determinants of Academic Choice and Task Performance Across Social Groups," in Swim, J. K. and C. Stanger (eds.), Prejudice, Academic Press, 1998.

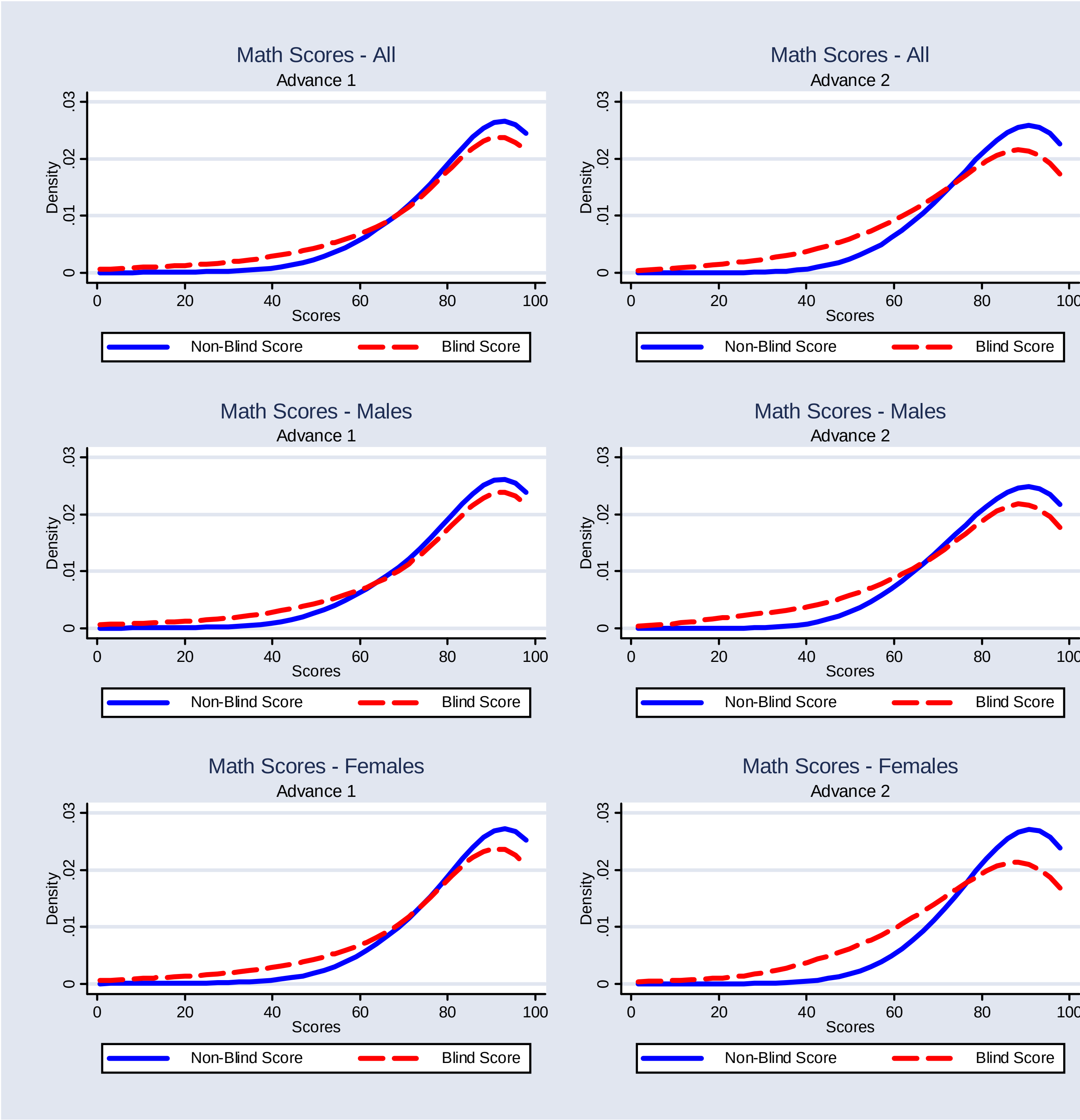
Steele, C. M., "A Threat in the Air: How Stereotypes Shape Intellectual Identity and Performance". American Psychologist, 1997, 52, 613-629.

Rebhorn, Leslie S., and D. D. Miles (1999), "High-Stakes Testing: Barrier to Gifted Girls in Mathematics and Science?" School Science and Mathematics, 99 (6), pp. 313–318.

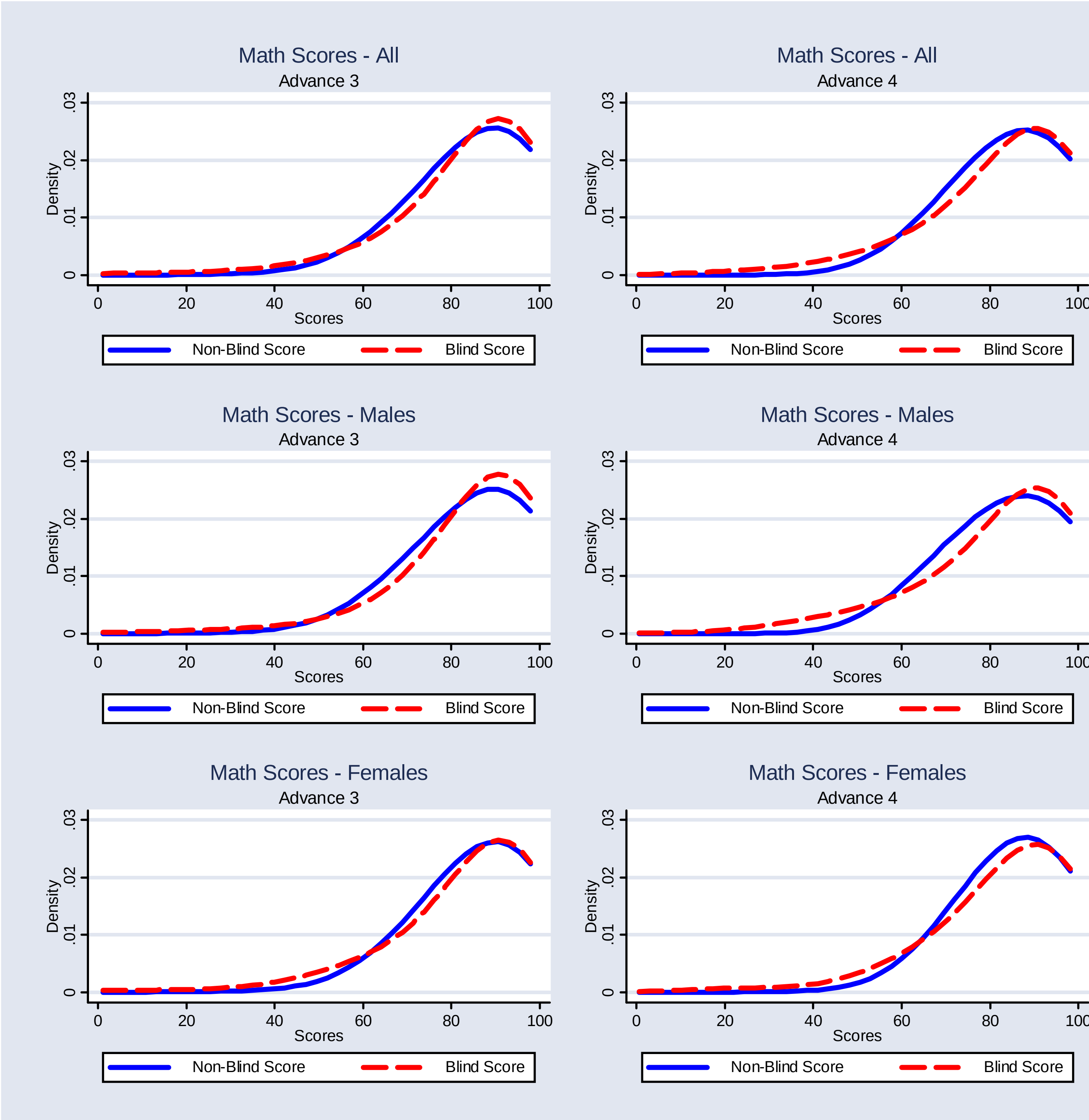
Tiedemann, J. (2000), “Parents’ Gender Stereotypes and Teachers’ Beliefs as Predictors of Children’ Concept of Their Mathematical Ability in Elementary School,” Journal of Educational Psychology, 92 (1), pp. 144–151.

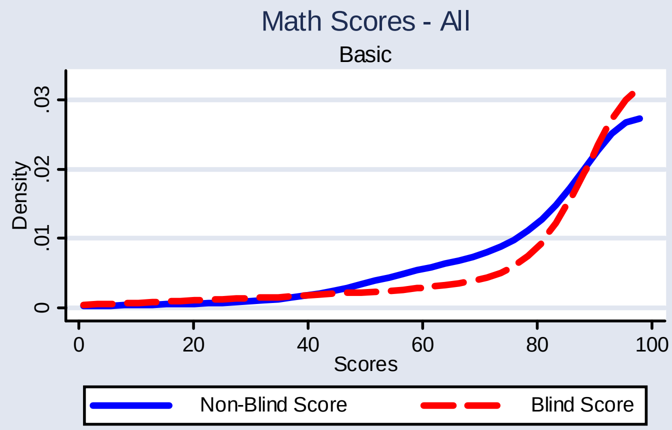
Figure 1: Math Tests Scores Distributions

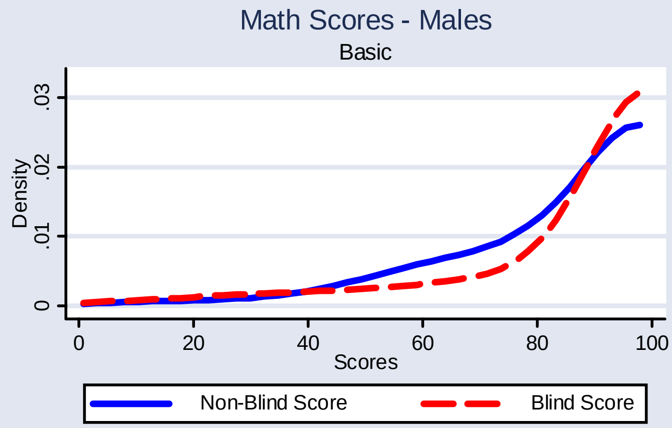




Math Scores - All
Advance 1
Math Scores - All
Advance 2
Math Scores - Males
Advance 1
Math Scores - Males
Advance 2
Math Scores - Females
Advance 1
Math Scores - Females
Advance 2
Density
Scores
Non-Blind Score
Blind Score


Math Scores - All
Advance 3
Math Scores - All
Advance 4
Math Scores - Males
Advance 3
Math Scores - Males
Advance 4
Math Scores - Females
Advance 3
Math Scores - Females
Advance 4
Density
0
.01
.02
.03
Scores
0
20
40
60
80
100
Non-Blind Score
Blind Score


Math Scores - All
Basic
Density
0
.01
.02
.03
0
20
40
60
80
100
Scores
Non-Blind Score
Blind Score


Math Scores - Males
Basic
Density
0
.01
.02
.03
0
20
40
60
80
100
Scores
Non-Blind Score
Blind Score

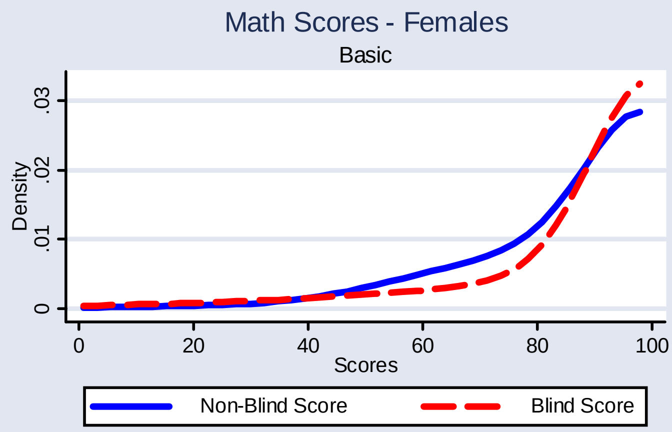
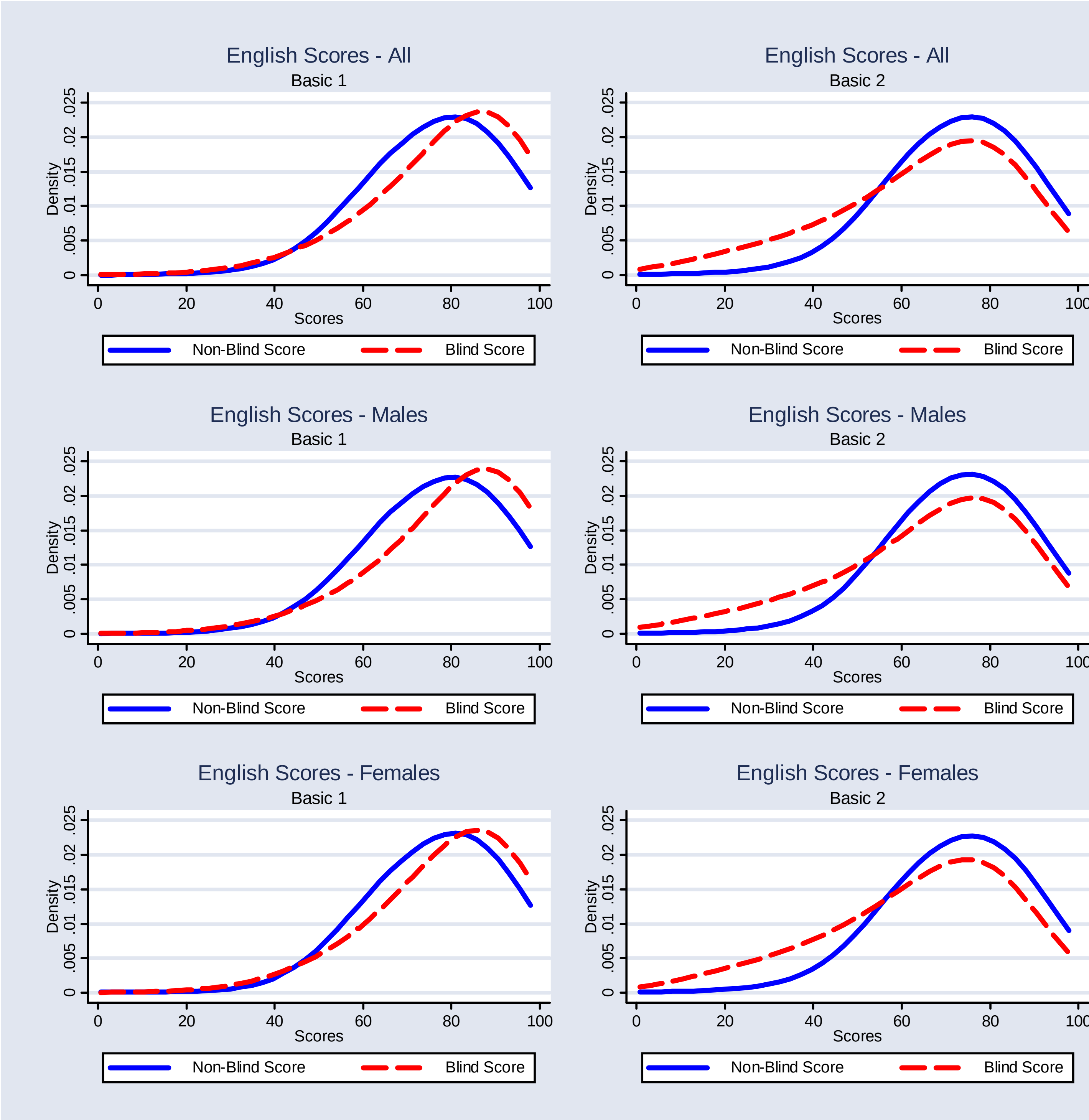

Math Scores - Females
Basic
Density
0
.01
.02
.03
0
20
40
60
80
100
Scores
Non-Blind Score
Blind Score

Figure 2: English Tests Scores Distributions



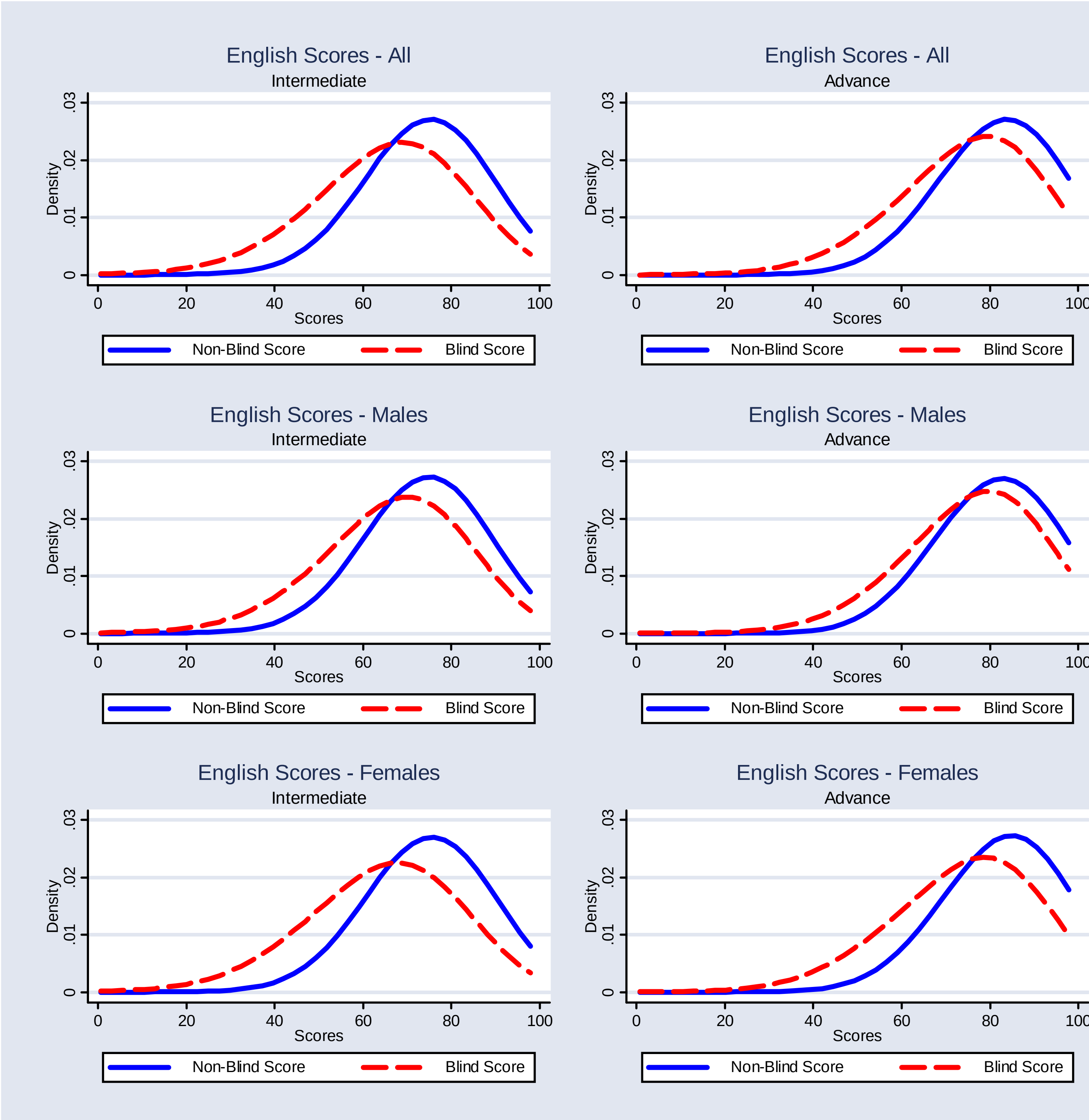
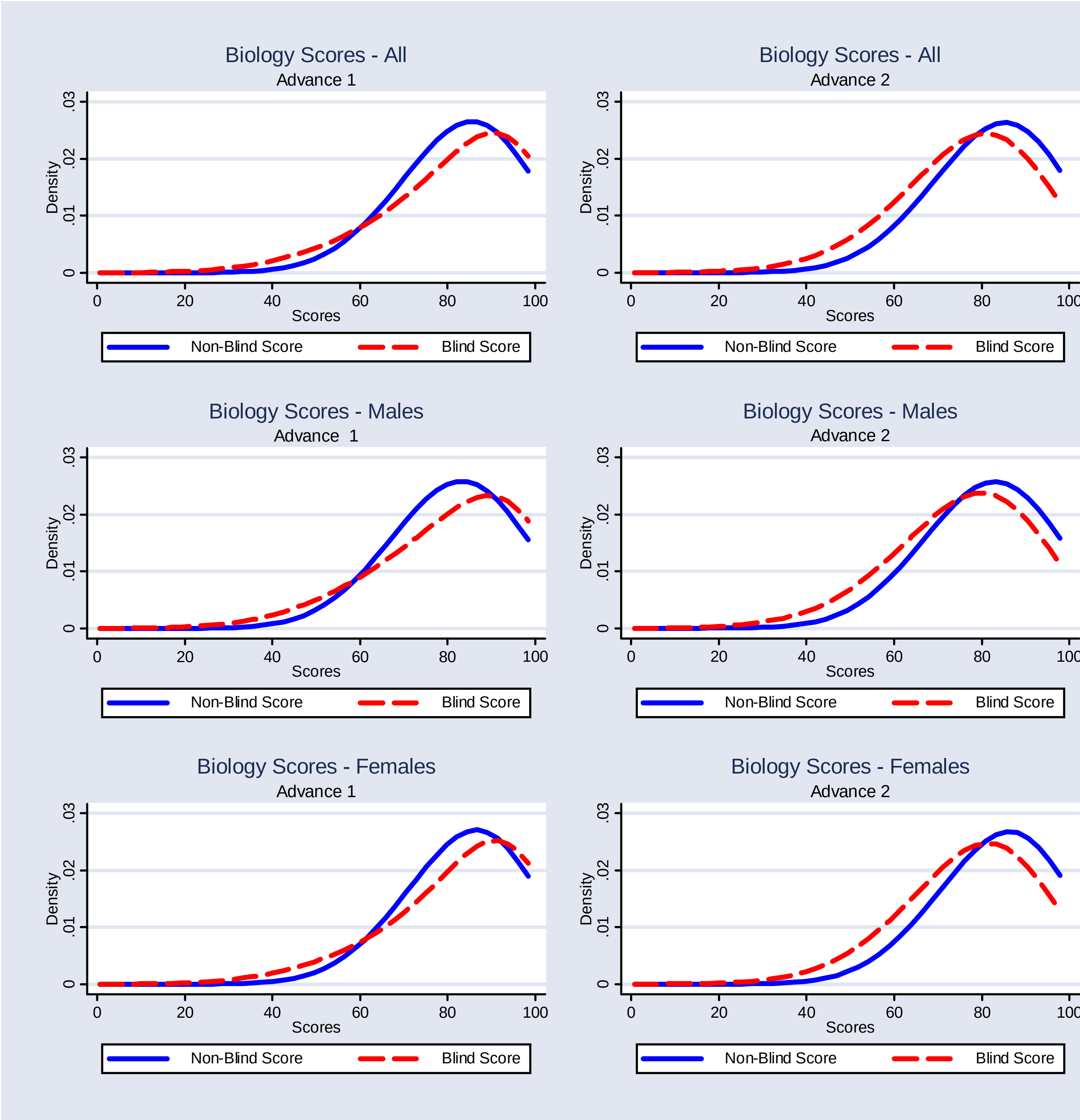
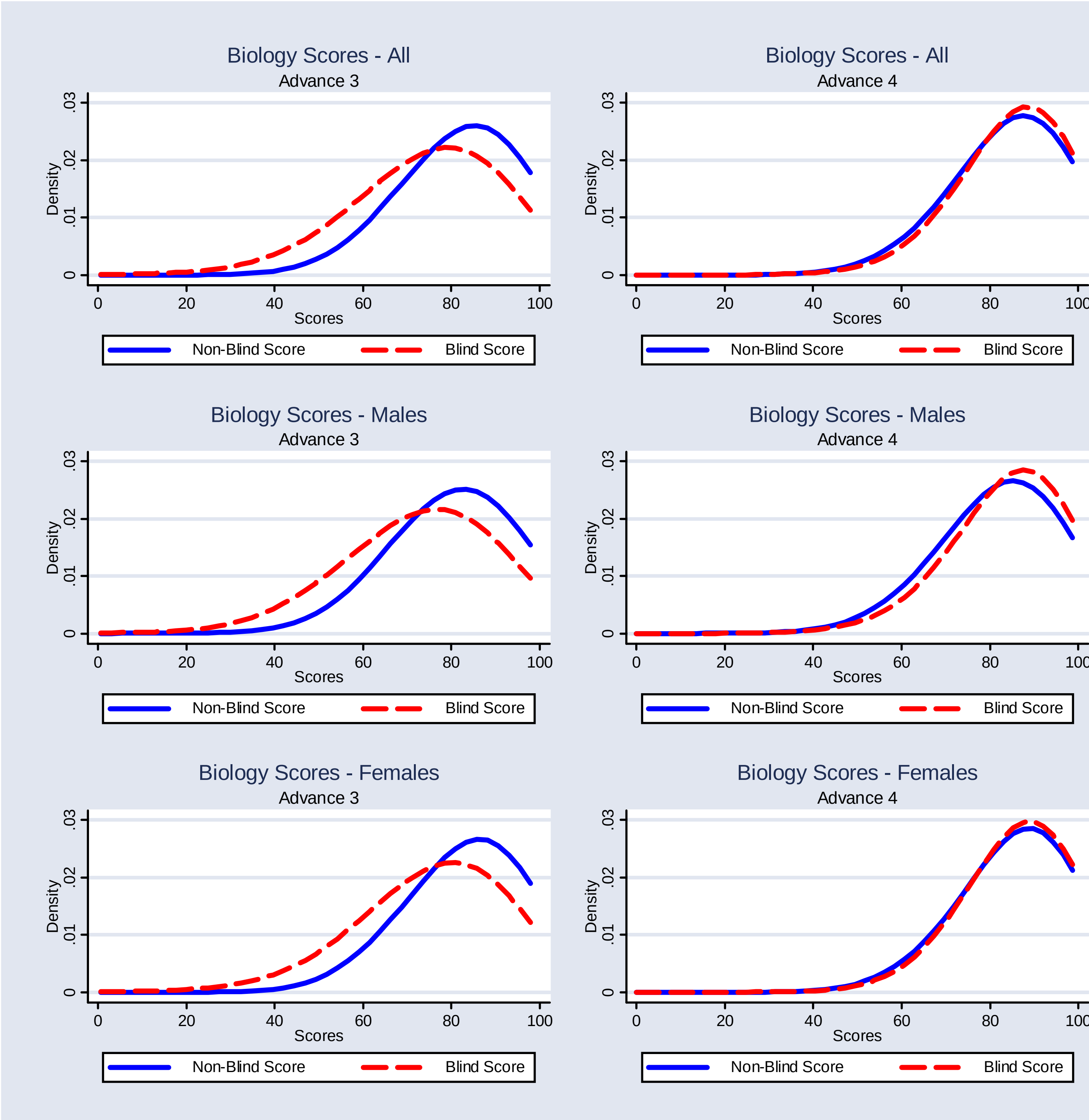

English Scores - All
Intermediate
Density
Scores
Non-Blind Score
Blind Score
English Scores - All
Advance
English Scores - Males
Intermediate
English Scores - Males
Advance
English Scores - Females
Intermediate
English Scores - Females
Advance

Figure 3: Biology Tests Scores Distributions




Biology Scores - All
Advance 3
Density
Scores
Non-Blind Score
Blind Score
Biology Scores - All
Advance 4
Density
Scores
Non-Blind Score
Blind Score
Biology Scores - Males
Advance 3
Density
Scores
Non-Blind Score
Blind Score
Biology Scores - Males
Advance 4
Density
Scores
Non-Blind Score
Blind Score
Biology Scores - Females
Advance 3
Density
Scores
Non-Blind Score
Blind Score
Biology Scores - Females
Advance 4
Density
Scores
Non-Blind Score
Blind Score

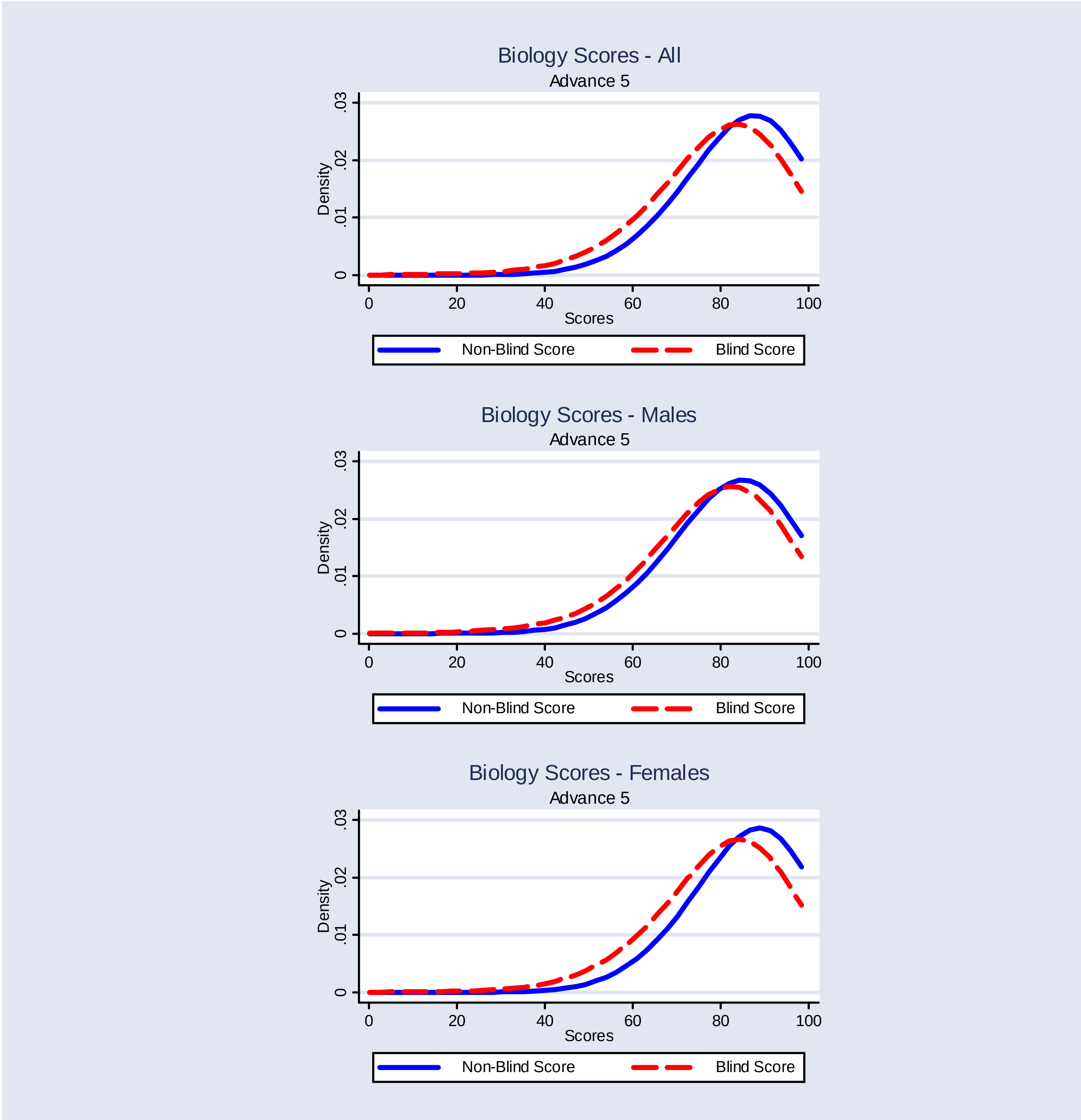

Biology Scores - All
Advance 5
Density
0
.01
.02
.03
0
20
40
60
80
100
Scores
Non-Blind Score
Blind Score
Biology Scores - Males
Advance 5
Density
0
.01
.02
.03
0
20
40
60
80
100
Scores
Non-Blind Score
Blind Score
Biology Scores - Females
Advance 5
Density
0
.01
.02
.03
0
20
40
60
80
100
Scores
Non-Blind Score
Blind Score

Figure 4: Literature Tests Scores Distributions

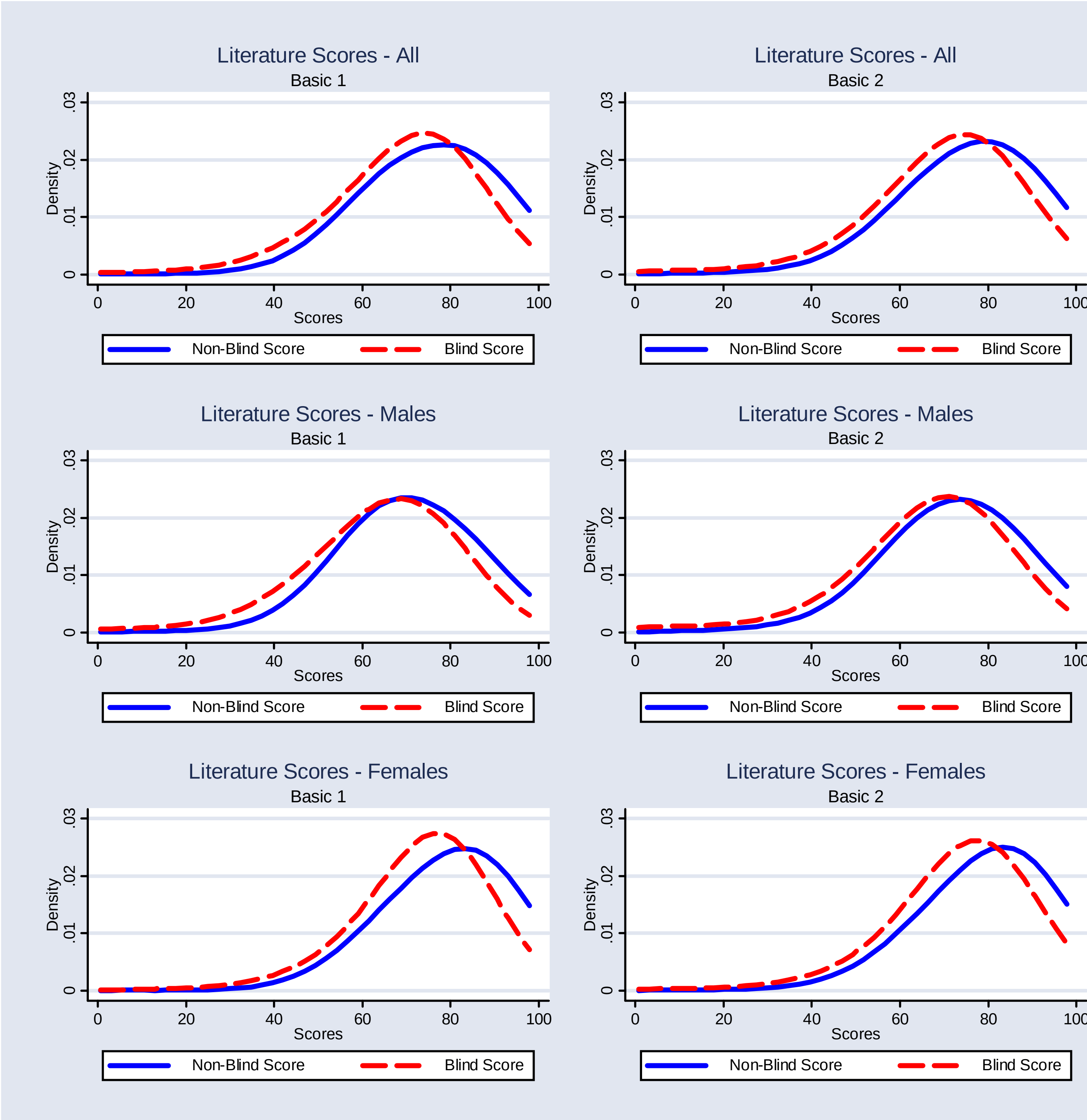
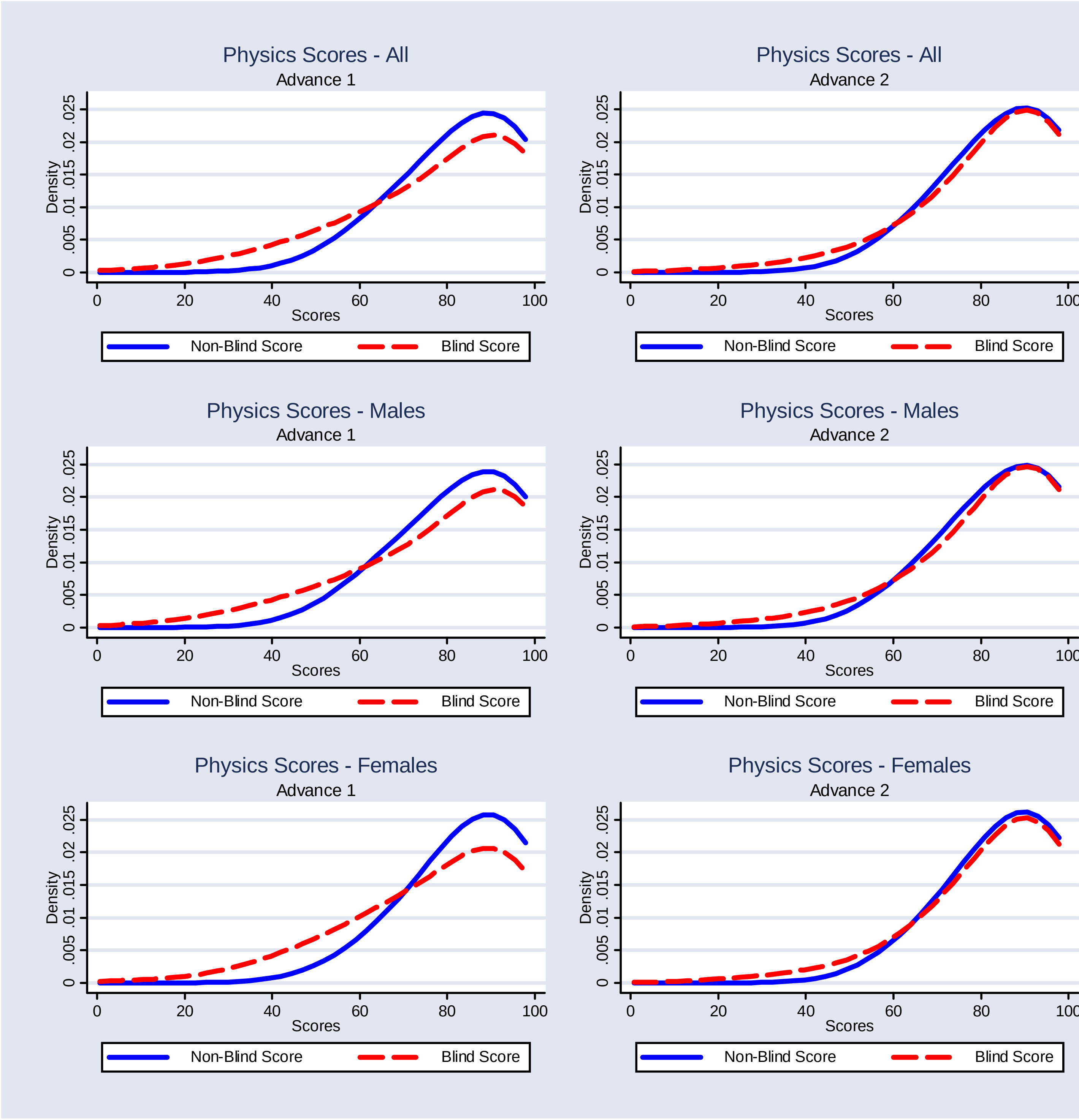


Figure 5: Physics Tests Scores Distributions



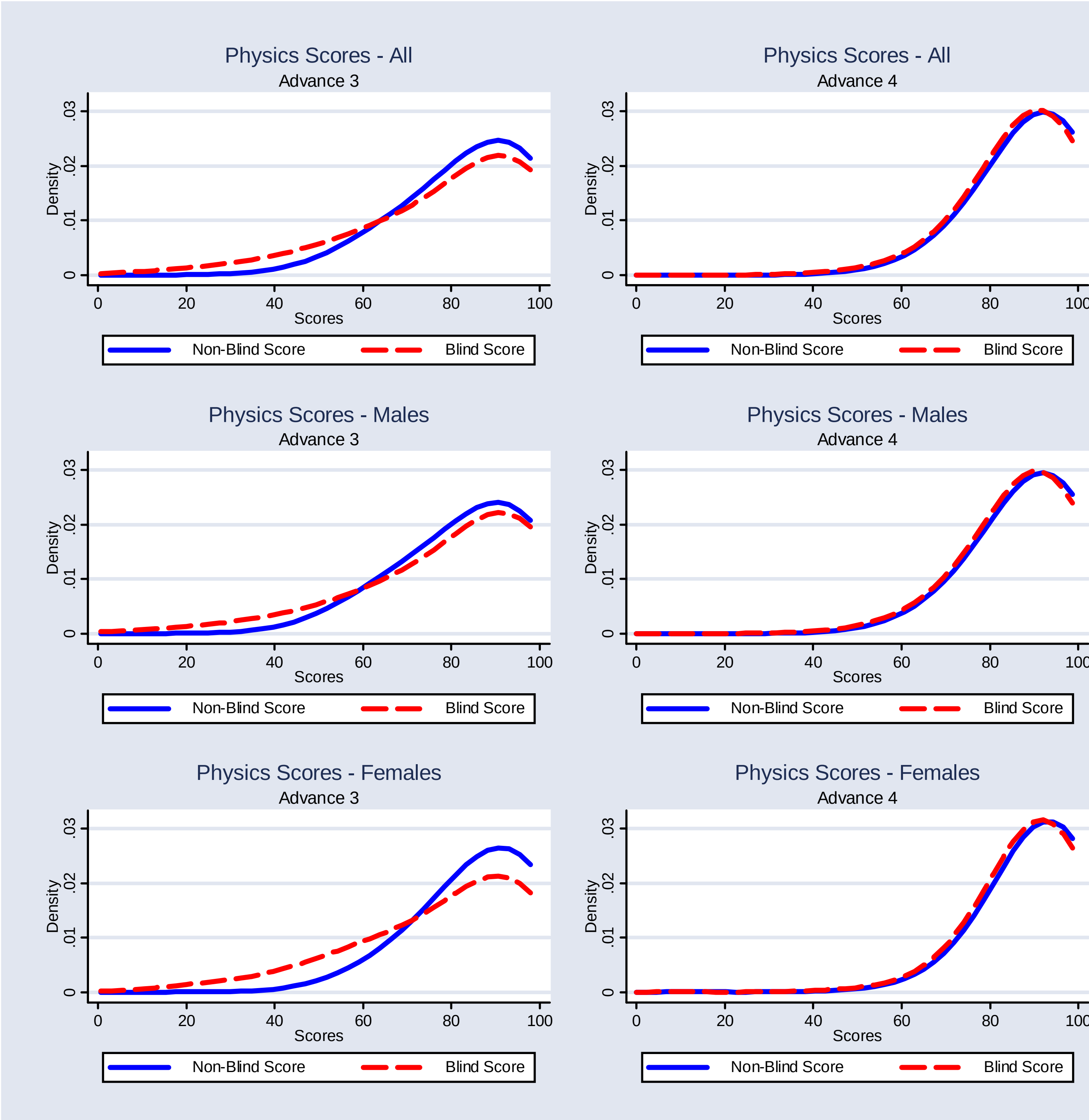

Physics Scores - All
Advance 3
Physics Scores - All
Advance 4
Physics Scores - Males
Advance 3
Physics Scores - Males
Advance 4
Physics Scores - Females
Advance 3
Physics Scores - Females
Advance 4
Density
Scores
Non-Blind Score
Blind Score

Figure 6: Chemistry Tests Scores Distributions

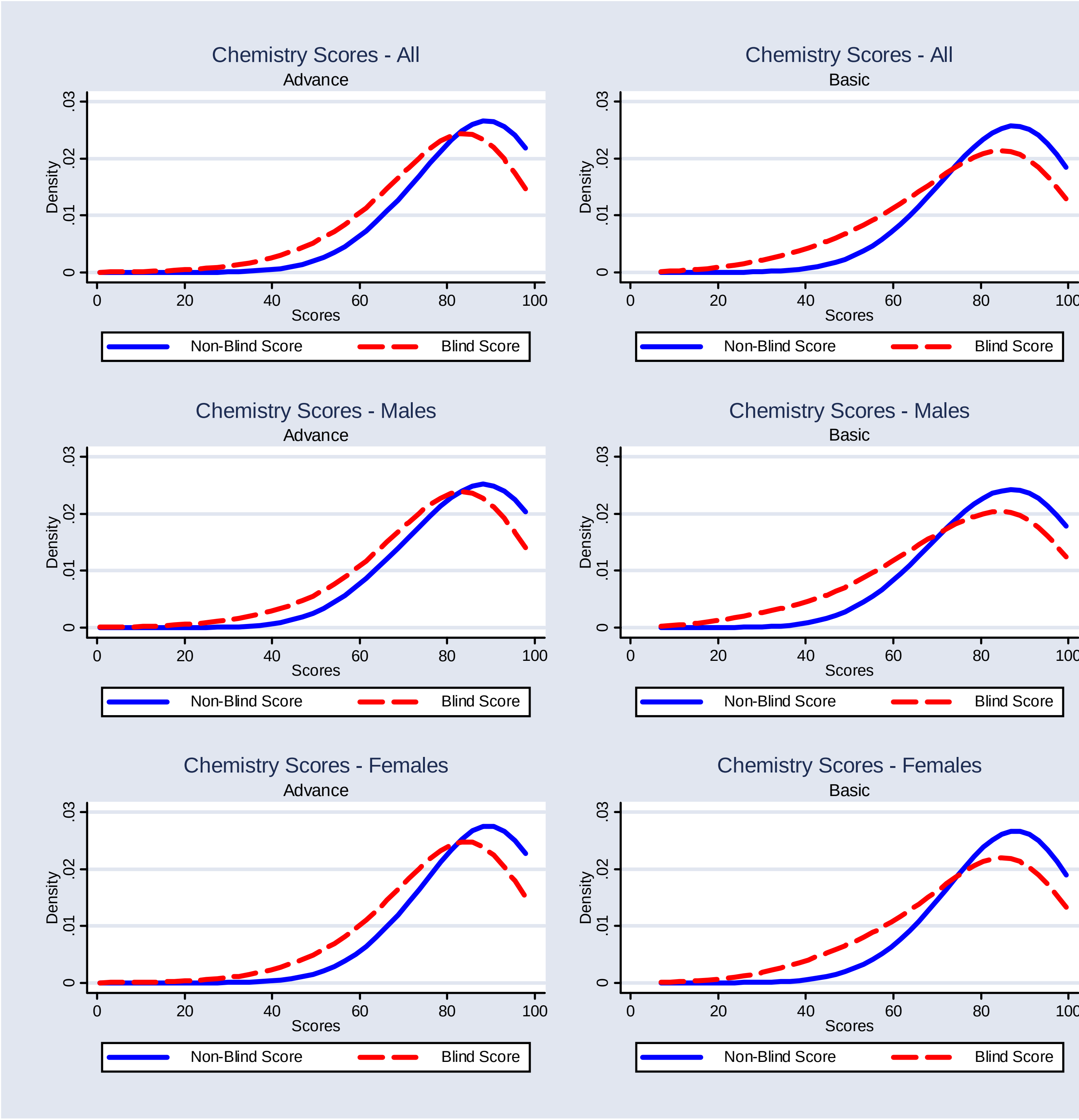


Figure 7: Computer Science Tests Scores Distributions

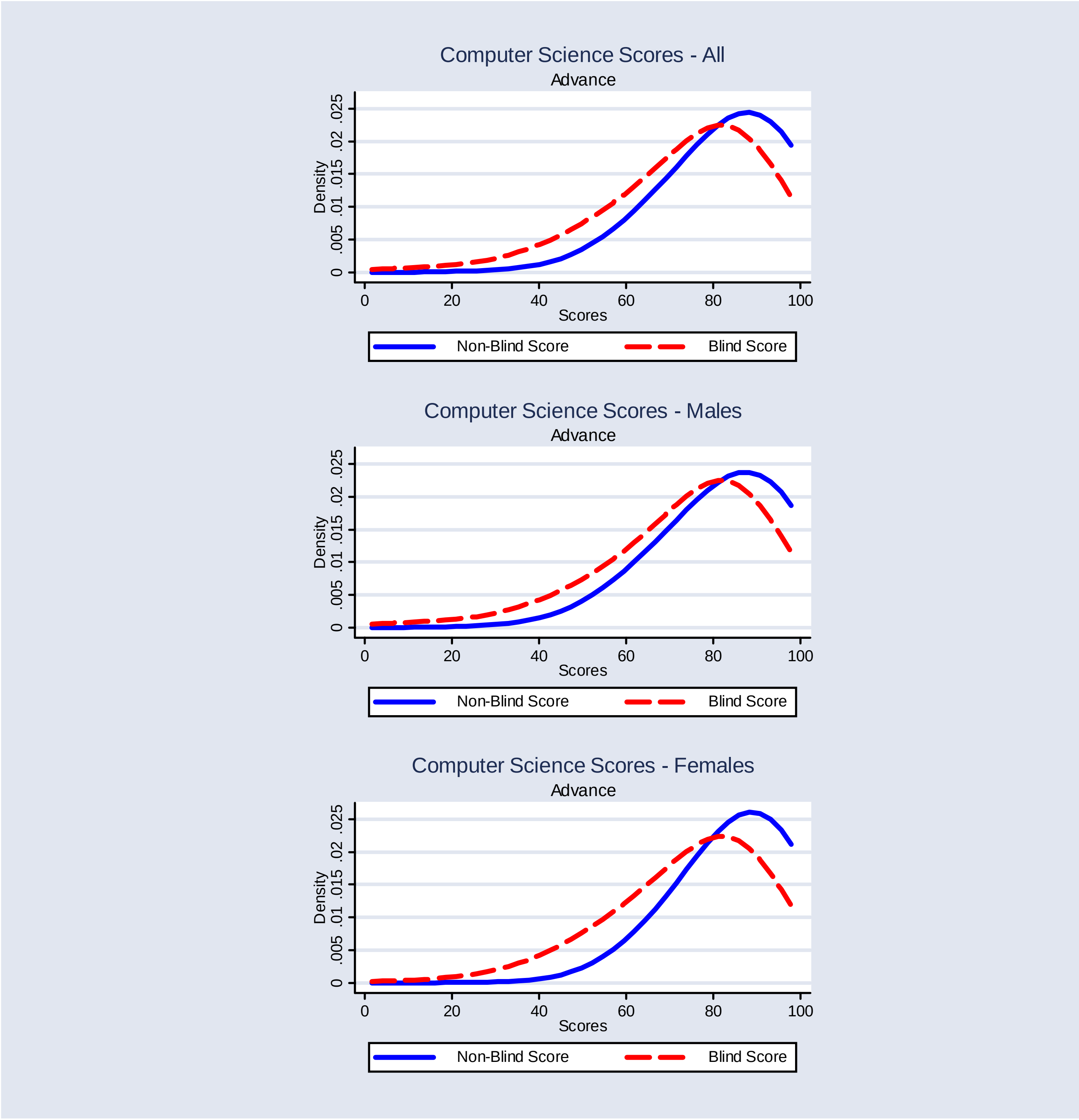


Table 1: Means and Standard Deviations of National and School Scores

Subject and Test (1)	Number of Observations (2)	Mean School Score (3)	Mean External Score (4)	T test for the difference in means (5)
Math, basic 1	16,647	70.0 (17.2)	67.5 (16.3)	6.2
Math, basic 2	11,862	88.6 (18.1)	91.2 (19.1)	7.2
Math,Intermediate 1	7,346	80.8 (15.5)	80.6 (21.0)	0.3
Math, Intermerdiate 2	4,490	80.9 (15.1)	75.2 (18.7)	8.6
Math, Advanced 1	3,983	87.3 (13.1)	81.3 (20.2)	12.1
Math, Advanced 2	2, 197	85.9 (12.8)	84.4 (15.7)	2.3
English, Basics 1	8,130	77.2 (14.1)	80.1 (15.5)	8.0
English, Basics 2	8,986	73.1 (14.1)	64.9 (20.4)	18.1
English, Intermediate	11,466	74.1 (10.9)	65.4 (14.7)	27.3
English, Advanced	13,843	82.3 (10.8)	75.1 (13.8)	27.5
Biology, Advanced 1	5,274	83.4 (11.2)	83.1 (14.9)	0.6
Biology, Advanced 2	5, 277	82.6 (11.4)	77.0 (13.5)	18.9
Biology, Advanced 3	5,276	82.7 (11.6)	74.5 (15.3)	15.4
Biology, Advanced 4	5,101	85.0 (10.8)	86.1 (9.9)	3.2
Biology, Advanced 5	5,079	85.2 (10.8)	79.89 (12.8)	11.8
Literature, Basic	32,908	76.1 (14.3)	69.9 (15.8)	25.6
Literature, Advanced	3,097	82.1 (11.3)	75.7 (10.7)	20.0
Chemistry, Basic	964	84.9 (12.2)	76.5 (17.2)	2.1
Chemistry, Advanced	3,616	85.7 (11.3)	78.4 (14.5)	14.1
Physics, Advanced 1	4,854	84.0 (13.2)	78.2 (20.0)	11.6
Physics, Advanced 2	1916	86.2 (13.2)	82.8 (16.4)	4.1
Physics, Advanced 3	4927	84.7 (13.5)	79.4 (19.9)	10.3
Physics, Advanced, 4	3,023	89.6 (9.8)	88.1 (10.2)	3.6
Computer, Advanced	3,546	83.5 (13.7)	73.8 (17.7)	14.2

Notes: The t statistics in column (4) reflect estimated standard errors that are corrected for clustering by school.

Table 2: Distribution of Female and Male Students By Level of Study in Elective Subjects

Subject	Basic Level		Advance Level	
	Female	Male	Female	Male
English[a]	51	49	49	51
Literature	53	47	84	16
Biology	-	-	67	33
Chemistry	61	49	61	49
Computer Science	-	-	40	60
Mathematics	54	46	46	54

Notes: English, Math and Literature are compulsory subjects at basic level. Biology, chemistry, computer science and physics are electives.

Table 3a: Estimated Gender Bias By Subject, Dependent Variables Are Standardized Scores

	English	Literature	Math	Biology	Chemistry	Computer Science	Physics
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Constant	0.041	0.214	0.044	0.268	0.057	0.096	0.012
	(0.019)	(0.022)	(0.145)	(0.031)	(0.043)	(0.520)	(0.035)
Male	0.114	-0.475	-0.105	-0.139	-0.122	-0.037	-0.015
	(0.014)	(0.018)	(0.145)	(0.022)	(0.040)	(0.042)	(0.035)
Non-blind score	0.048	0.024	0.019	-0.059	-0.025	0.004	0.041
	(0.018)	(0.015)	(0.015)	(0.222)	(0.036)	(0.047)	(0.033)
Male x non-blind score	-0.180	-0.053	-0.086	-0.125	-0.058	-0.122	-0.130
	(0.013)	(0.014)	(0.012)	(0.023)	(0.033)	(0.420)	(0.024)
Number of students	84,850	75,568	109,928	52,888	9,562	8,006	29,992
Number of schools	359	328	363	190	196	237	242

Notes: Standard errors are corrected for clustering and are presented in parentheses.

Table 3b: Estimated Gender Bias By Subject, Dependent Variables Are Standardized Scores, controls for student's characteristics and lagged outcomes are included

	English	Literature	Math	Biology	Chemistry	Computer Science	Physics
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Male	0.117	-0.483	-0.072	-0.057	-0.061	-0.059	0.092
	(0.014)	(0.015)	(0.014)	(0.020)	(0.032)	(0.040)	(0.028)
Non-blind score	0.048	0.024	0.020	-0.059	-0.025	0.004	0.041
	(0.018)	(0.015)	(0.015)	(0.222)	(0.036)	(0.047)	(0.033)
Male x non-blind score	-0.180	-0.053	-0.086	-0.125	-0.058	-0.122	-0.130
	(0.013)	(0.014)	(0.102)	(0.023)	(0.033)	(0.420)	(0.024)
Number of students	84,850	75,568	109,928	52,888	9,562	8,006	29,992
Number of schools	359	328	363	190	196	237	242

Notes: Standard errors are corrected for clustering and are presented in parentheses. The regressions include as controls students background characteristics: father and mother schooling, number of siblings and 6 dummies as indicators of ethnic origin (Africa/Asia, America/Europe, Israel, Soviet Union Republics, Ethiopia and a category for students with missing data on ethnic origin.

Table 3c: Estimated Gender Bias By Subject, Dependent Variables Are Percentile Rank Scores

	English	Literature	Math	Biology	Chemistry	Computer Science	Physics
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Constant	50.783	56.139	50.973	57.351	52.206	52.342	48,673
	(0.576)	(0.731)	(0.671)	(0.063)	1.275	(1.663)	(1.001)
Male	3.449	-14.779	-3.225	-4.332	-3.123	-0.496	0.342
	(0.412)	(0.486)	(0.396)	(0.665)	(1.223)	(1.328)	(1.007)
Non-blind score	1.425	0.571	0.066	-1.823	-0.957	-0.332	0.879
	(0.533)	(0.485)	(0.486)	(0.591)	(1.001)	(1.516)	(1.058)
Male x non-blind score	-5.405	-1.050	-2.247	-3.397	-1.336	-3.501	-3.822
	(0.373)	(0.393)	(0.334)	(0.683)	(0.921)	(1.282)	(0.715)
Number of students	84,850	75,568	109,928	52,888	7,232	8,006	29,992
Number of schools	359	328	363	190	180	237	242

Notes: Standard errors are corrected for clustering and are presented in parentheses.

Table 4a: Estimates of Gender Bias in Tests Where Blind and Non-Blind Score Distributions are Identical

	Biology Basic (1)	Biology Advanced (2)	Math Basic (3)	Math Advanced (4)	English Basic (5)	English Intermediate (6)	English Advanced (7)	Literature (8)
Male	-0.089	-0.029	-0.099	0.181	0.074	0.179	0.160	-0.489
	(0.026)	(0.027)	(0.038)	(0.047)	(0.025)	(0.023)	(0.019)	(0.015)
Blind-test	-0.105	-0.101	-0.016	-0.029	0.004	0.043	0.084	0.027
	(0.031)	(0.035)	(0.044)	(0.050)	(0.024)	(0.026)	(0.024)	(0.016)
Male X Non-Blind Test	-0.162	-0.216	-0.062	-0.147	-0.078	-0.224	-0.272	-0.057
	(0.037)	(0.037)	(0.034)	(0.048)	(0.021)	(0.022)	(0.018)	(0.015)
Number of students	10,202	10,158	8,980	4,394	16,260	22,932	27,686	65,816
Number of schools	178	177	167	127	338	334	306	317

Notes: Standard errors are corrected for clustering and are presented in parentheses. Dependent Variables Are Standardized Scores. The regressions include as controls students background characteristics: father and mother schooling, number of siblings and 6 dummies as indicators of ethnic origin (Africa/Asia, America/Europe, Israel, Soviet Union Republics, Ethiopia and a category for students with missing data on ethnic origin.

Table 4b: Estimated Gender Bias in Exams By Level of Study

Level of Study	English	Literature	Biology	Math	Chemistry	Physics
	(1)	(2)	(3)	(4)	(5)	(6)
Basic Level:	-0.074	-0.057	-0.211	-0.099	0.032	-0.178
	(0.025)	(0.045)	(0.096)	(0.017)	(0.070)	(0.113)
Number of Students	8,111	32,908	225	16,647	964	213
Number of Schools	338	317	45	354	89	36
Advanced level:	-0.272	-0.070	-0.102	-0.147	-0.072	-0.140
	(0.018)	(0.045)	(0.38)	(0.029)	(0.033)	(0.029)
Number of Students	13,843	3,097	5,274	2,197	3,616	4,854
Number of Schools	306	161	185	127	180	227

Notes: Standard errors are corrected for clustering and are presented in parentheses. Dependent Variables Are Standardized Scores. The regressions include as controls students background characteristics: father and mother schooling, number of siblings and 6 dummies as indicators of ethnic origin (Africa/Asia, America/Europe, Israel, Soviet Union Republics, Ethiopia and a category for students with missing data on ethnic origin.

Table 4c: Estimated Gender Bias Versus Other Biases

	English (1)	Literature (2)	Math (3)	Biology (4)	Chemistry (5)	Comp. Science (6)	Physics (7)
Interaction of Non-Blind Score with:							
Male	-0.176 (0.013)	-0.053 (0.014)	-0.087 (0.012)	-0.130 (0.023)	-0.069 (0.032)	-0.123 (0.042)	-0.127 (0.024)
Recent immigrant	0.377 (0.044)	-0.032 (0.059)	0.146 (0.047)	0.202 (0.052)	0.440 (0.165)	-0.080 (0.031)	0.163 (0.099)
Father's schooling	-0.040 (0.015)	-0.012 (0.014)	-0.024 (0.014)	0.026 (0.022)	0.004 (0.031)	-0.018) (0.034)	0.015 (0.026)
Mother's schooling	-0.027 (0.016)	-0.016 (0.014)	-0.003 (0.014)	0.002 (0.022)	0.026 (0.033)	0.011 (0.040)	-0.054 (0.025)
Number of siblings	0.012 (0.014)	-0.001 (0.012)	-0.017 (0.013)	(0.013 (0.020)	0.061 (0.048)	0.011 (0.040)	-0.050 (0.032)
Ethnic origin: America-Europe	0.007 (0.327)	0.494 (0.244)	-0.189 (0.281)	-0.235 (0.046)	0.437 (0.345)	0.482 (0.064)	0.980 (0.105)
Ethnic origin: Asia-Africa	0.028 (0.330)	-0.463 (0.243)	-0.193 (0.285)	-0.281 (0.042)	0.380 (0.353)	0.482 (0.070)	0.980 (0.106)

Notes: Standard errors are corrected for clustering and are presented in parenthesis. The regressions include as controls all the level variables used in all the interactions (father and mother schooling, number of siblings, 6 dummies as indicators of ethnic origin), the number of matriculation credit units achieved in 11th grade, average score on 11th grade matriculation exams.

Table 5a: Estimated Gender Bias When Schools Were Divided to Two Samples According to Average Performance of Female and Male Students in Eleventh Grade Matriculation Exam

	English	Literature	Math	Biology	Chemistry	Computer Science	Physics
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
			Schools where girls are better than boys				
Male	0.123	-0.459	-0.063	- 0.047	-0.076	-0.106	0.074
	(0.013)	(0.022)	(0.015)	(0.024)	(0.042)	(0.053)	(0.029)
Male x non-blind score	-0.179	-0.052	-0.084	-0.128	-0.084	-0.131	-0.141
	(0.016)	(0.018)	(0.013)	(0.028)	(0.044)	(0.047)	(0.024)
Students	58,236	46,276	77,952	39,864	6,112	5,450	22,570
Schools	212	175	214	126	107	121	150
			Schools where boys are better than girls				
Male	0.131	-0.515	-0.064	-0.055	-0.044	0.110	0.187
	(0.019)	(0.022)	(0.025)	(0.035)	(0.052)	(0.067)	(0.061)
Male x non-blind score	- 0.197	-0.060	-0.117	-0.123	-0.014	-0.128	-0.172
	(0.024)	(0.024)	(0.022)	(0.040)	(0.049)	(0.060)	(0.054)
Students	25,422	29,018	30,826	12,340	3,289	1,958	6,452
Schools	118	134	119	53	70	54	47

Notes: Standard errors are corrected for clustering and are presented in parentheses. Dependent Variables Are Standardized Scores. The regressions include as controls students background characteristics: father and mother schooling, number of siblings and 6 dummies as indicators of ethnic origin (Africa/Asia, America/Europe, Israel, Soviet Union Republics, Ethiopia and a category for students with missing data on ethnic origin.

Table 5b: Estimated Gender Bias When Schools Were Divided to Two Samples According to Average Score In Matriculation Exams of Twelve Grade in 2000

	English (1)	Literature (2)	Math (3)	Biology (4)	Chemistry (5)	Computer Science (6)	Physics (7)
				Schools where girls are better than boys			
Male	0.124	-0.503	-0.065	- 0.080	-0.047	0.018	0.085
	(0.015)	(0.019)	(0.015)	(0.020)	(0.034)	(0.043)	(0.029)
Male x non-blind score	-0.191	-0.049	-0.097	-0.112	-0.092	-0.206	-0.137
	(0.014)	(0.016)	(0.012)	(0.024)	(0.037)	(0.047)	(0.028)
Students	67,836	53,518	88,676	45,130	6,860	4,788	22,152
Schools	248	199	244	149	117	107	143
				Schools where boys are better than girls			
Male	0.120	-0.423	-0.067	0.058	-0.087	0.013	0.086
	(0.034)	(0.026)	(0.027)	(0.059)	(0.077)	(0.086)	(0.064)
Male x non-blind score	- 0.148	-0.085	-0.079	-0.186	0.028	-0.066	-0.109
	(0.032)	(0.029)	(0.029)	(0.075)	(0.080)	(0.084)	(0.051)
Students	13,374	19,754	17,948	6,004	2,162	1,788	6,074
Schools	72	96	78	27	49	41	48

Notes: Standard errors are corrected for clustering and are presented in parentheses. Dependent Variables Are Standardized Scores. The regressions include as controls students background characteristics: father and mother schooling, number of siblings and 6 dummies as indicators of ethnic origin (Africa/Asia, America/Europe, Israel, Soviet Union Republics, Ethiopia and a category for students with missing data on ethnic origin.

Table 5c: Estimated Gender Bias When Schools Were Divided to Two Samples According to the Score in the External in Each Subject of Twelve Graders in 2000

	English	Literature	Math	Biology	Chemistry	Computer Science	Physics
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
			Schools where girls are better than boys				
Male	0.131 (0.032)	-0.481 (0.016)	-0.083 (0.021)	- 0.067 (0.021)	-0.027 (0.037)	0.003 (0.053)	0.047 (0.031)
Male x non-blind score	-0.148 (0.028)	-0.059 (0.015)	-0.080 (0.017)	-0.122 (0.026)	-0.082 (0.044)	-0.169 (0.055)	-0.130 (0.032)
Students	15,156	71,848	47,732	38,496	5,766	4,024	15,040
Schools	79	272	162	121	92	83	99
			Schools where boys are better than girls				
Male	0.124 (0.015)	-0.507 (0.094)	-0.056 (0.018)	-0.064 (0.047)	-0.113 (0.062)	0.067 (0.062)	0.147 (0.047)
Male x non-blind score	- 0.196 (0.021)	-0.081 (0.093)	-0.105 (0.015)	-0.116 (0.049)	-0.032 (0.054)	-0.167 (0.066)	-0.135 (0.038)
Students	66,054	1,424	58,892	12,638	3,256	2,566	13,186
Schools	241	23	160	55	74	66	92

Notes: Standard errors are corrected for clustering and are presented in parentheses. Dependent Variables Are Standardized Scores. The regressions include as controls students background characteristics: father and mother schooling, number of siblings and 6 dummies as indicators of ethnic origin (Africa/Asia, America/Europe, Israel, Soviet Union Republics, Ethiopia and a category for students with missing data on ethnic origin.

Table 6: Estimated Gender Biases By Students' Ability

	English	Literature	Math	Biology	Chemistry	Physics	Computer Science
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Interaction of non-blind score x male x 11th grade average score:							
Above mean	-0.087	-0.182	-0.016	-0.013	0.017	0.026	0.097
	(0.014)	(0.015)	(0.013)	(0.032)	(0.035)	(0.080)	(0.049)
Below mean	-0.086	-0.169	-0.098	-0.236	-0.157	-0.299	-0.359
	(0.017)	(0.017)	(0.014)	(0.032)	(0.041)	(0.033)	(0.050)

Notes: Standard errors are corrected for clustering and are presented in parenthesis. The regressions include as controls all the level variables used in all the interactions (father and mother schooling, number of siblings, 6 dummies as indicators of ethnic origin), the number of matriculation credit units achieved in 11th grade, average score on 11th grade matriculation exams.

Table 7: Estimated Gender Biases by Male and Female Teachers

	Math			English		
	All Teachers (1)	Male Teachers (2)	Female Teachers (3)	All Teachers (4)	Male Teachers (5)	Female Teachers (6)
Constant	-0.050	-0.173	0.009	-0.086	-0.109	-0.084
	(0.060)	(0.074)	(0.068)	(0.061)	(0.221)	(0.023)
Male	-0.039	0.113	-0.113	0.124	0.045	0.130
	(0.041)	(0.079)	(0.036)	(0.043)	(0.192)	(0.034)
Non-blind test	-0.020	0.044	-0.051	-0.011	0.186	-0.029
	(0.046)	(0.047)	(0.062)	(0.045)	(0.164)	(0.033)
Male X non-blind test	-0.089	-0.209	-0.030	-0.165	-0.136	-0.163
	(0.030)	(0.051)	(0.033)	(0.032)	(0.166)	(0.047)
Number of students	10,058	3,286	6,772	7,478	558	6,920
Number of schools	40	29	38	39	11	39

Notes: Standard errors are corrected for clustering and are presented in parenthesis. The regressions include as controls all the level variables used in all the interactions (father and mother schooling, number of siblings, 6 dummies as indicators of ethnic origin), the number of matriculation credit units achieved in 11^{th} grade, average score on 11^{th} grade matriculation exams.

Table 8a: Estimated Gender Biases From a Sample That Includes Teachers in the Incentives Program and a Sample of All Other Teachers

	Math 2001		Math 2000		English 2001		English 2000	
Variables	Incentives (1)	No Incentives (2)	Incentives (3)	No Incentives (4)	Incentives (5)	No Incentives (6)	Incentives (7)	No Incentives (8)
Constant	-0.095	0.054	-0.175	0.055	-0.057	0.054	-0.107	0.066
	(0.076)	(0.024)	(0.049)	(0.021)	(0.075)	(0.020)	(0.081)	(0.019)
Male	-0.056	-0.112	-0.005	-0.068	0.102	0.113	0.156	0.135
	(0.051)	(0.015)	(0.035)	(0.015)	(0.055)	(0.015)	(0.047)	(0.015)
Non-blind test	-0.054	0.024	0.073	0.051	-0.046	0.053	-0.951	0.027
	(0.059)	(0.016)	(0.064)	(0.017)	(0.056)	(0.019)	(0.073)	(0.018)
Male X non-blind test	-0.076	-0.086	-0.109	-0.137	-0.154	-0.182	-0.111	-0.181
	(0.039)	(0.013)	(0.041)	(0.012)	(0.043)	(0.014)	(0.050)	(0.014)
Number of students	6,668	99,780	7,918	110,156	5,312	77,062	6,676	94,561
Number of schools	22	366	21	323	22	354	21	324

Notes: Standard errors are corrected for clustering and are presented in parenthesis.

Table 8b: Estimated Gender Bias in Schools Included In Program with Teachers Group Monetary Incentives, By Subject, year 2001

	English	Literature	Math	Biology	Chemistry	Computer Science	Physics
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Constant	0.041	0.166	0.009	0. 290	-0.249	-0.142	-0.184
	(0.058)	(0.056)	(0.062)	(0.116)	(0.179)	(0.201)	(0.127)
Male	0.138	-0.487	-0.119	-0.176	-0.033	0.151	0.016
	(0.043)	(0.052)	(0.051)	(0.104)	(0.187)	(0.171)	(0.140)
Non-blind score	0.021	-0.090	0.045	-0.189	0.268	0.151	0.071
	(0.049)	(0.043)	(0.052)	(0.074)	(0.161)	(0.137)	(0.093)
Male x non-blind score	-0.208	-0.078	-0.093	0.068	-0.293	-0.258	-0.088
	(0.050)	(0.053)	(0.038)	(0.083)	(0.156)	(0.150)	(0.066)
Number of students	6,796	4,862	8,054	2,190	608	638	2,160
Number of schools	26	25	26	15	22	23	21

Notes: Standard errors are corrected for clustering and are presented in parentheses.

Table 8c: Estimated Gender Bias in Schools Included In Program with Teachers Group Monetary Incentives, By Subject, year 2000

	English	Literature	Math	Biology	Chemistry	Computer Science	Physics
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Constant	0.043	0.033	0.054	0. 290	0.106	0.113	-0.120
	(0.053)	(0.074)	(0.065)	(0.106)	(0.140)	(0.157)	(0.083)
Male	0.097	-0.435	-0.195	-0.261	-0.014	-0.060	-0.077
	(0.042)	(0.062)	(0.035)	(0.065)	(0.162)	(0.163)	(0.089)
Non-blind score	- 0.039	0.076	0.046	-0.194	-0.027	0.008	0.122
	(0.045)	(0.044)	(0.057)	(0.090)	(0.097)	(0.143)	(0.059)
Male x non-blind score	-0.155	-0.111	-0.070	-0.119	-0.082	-0.079	-0.111
	(0.049)	(0.071)	(0.035)	(0.111)	(0.074)	(0.125)	(0.074)
Number of students	7,868	5,324	9,334	1,218	824	652	2,590
Number of schools	26	26	26	18	22	21	23

Notes: Standard errors are corrected for clustering and are presented in parentheses.

Table 9: Estimated 12h Grade Math Gender Bias by 11th Grade Math Blind Test Score

	All Students	11th grade math score				
		Below school average	Above school average	First Third	Second Third	Third Third
	(1)	(2)	(3)	(4)	(5)	(6)
Male	-0.093	-0.102	-0.103	-0.106	-0.089	-0.105
	(0.022)	(0.035)	(0.024)	(0.041)	(0.024)	(0.035)
Non-blind test	0.345	1.165	-0.509	0.944	0.333	-0.021
	(-0.557)	(0.046)	(0.050)	(0.160)	(0.560)	(0.118)
Male X non-blind test	-0.118	-0.169	-0.087	-0.213	-0.085	-0.078
	(0.021)	(0.038)	(0.020)	(0.045)	(0.024)	(0.029)
Number of students	23,170	8,516	14,654	6,904	9,220	7,046
Number of schools	270	270	270	220	235	221

Notes: Standard errors are corrected for clustering and are presented in parenthesis. The regressions include as controls all the level variables used in all the interactions (father and mother schooling, number of siblings, 6 dummies as indicators of ethnic origin), the number of matriculation credit units achieved in 11th grade, average score on 11th grade matriculation exams.

Table 10: Estimated 12th grade math gender bias by 11th grade bias math

	All Students (1)	Sample of students whom their 11th grade blind score is higher than non-blind score (2)	Sample of students whom their 11th grade blind score is lower than non-blind score (3)
Male	-0.093 (0.022)	-0.128 (0.030)	-0.058 (0.025)
Non-blind test	0.345 (0.557)	-0.475 (0.050)	1.131 (0.047)
Male x non-blind test	-0.118 (0.021)	-0.103 (0.025)	-0.119 (0.031)
Number of Students	28,170	13,008	10,162
Number of Schools	270	256	222

Notes: Standard errors are corrected for clustering and are presented in parenthesis. The regressions include as controls all the level variables used in all the interactions (father and mother schooling, number of siblings, 6 dummies as indicators of ethnic origin), the number of matriculation credit units achieved in 11th grade, average score on 11th grade matriculation exams.