Program on Education Policy and Governance Working Papers Series

Gender Differences in Market Competitiveness in a Real Workplace: Evidence from Performance-based Pay Tournaments among Teachers

Victor Lavy
Hebrew University, Jerusalem

PEPG 08-02

Preliminary draft
Please do not cite without permission

Prepared for the CESifo/PEPG joint conference
"Economic Incentives: Do They Work in Education"
Insights and Findings from Behavioral Research

CESifo Conference Center
Munich, Germany
May 16-17, 2008

Gender Differences in Market Competitiveness in a Real Workplace: Evidence from Performance-based Pay Tournaments among Teachers*

Victor Lavy

Hebrew University, Royal Holloway University of London, CEPR and NBER

May 12 2008

Preliminary Draft

Victor Lavy
Department of Economics
Hebrew University
Mt. Scopus, Jerusalem 91905, Israel
msvictor@mscc.huji.ac.il

* Special thanks go to Roy Mill for his outstanding research assistance. I benefited from comments of seminar participants at Boston University.

Gender Differences in Market Competitiveness in a Real Workplace: Evidence from Incentive Tournaments among Teachers

Abstract

In this paper I examine how individual performance in a real work place is affected by the gender mix of the competitive group. The competition is among math, English and Language (Hebrew or Arabic) teachers who participated in a rank-order tournament that rewarded teachers with large cash bonuses based on the test performance of their classes. Teachers were awarded bonuses according to their ranking in the tournament. Ranking was based on a value added measure calculated by the difference between the actual mean performance of the teacher's class and a value predicted on the basis of a regression that controlled for the students' characteristics and a school-level fixed effect. Therefore teachers competed against other teachers of the same subject in their own school and they were told so explicitly. This framework created three distinct types of competitive environment which allow examining whether women succeed differentially when competing only against women or also against men. The evidence suggest that the average ranking, winning rate and awarded prize did not differ by gender nor between teachers in competition groups with only female teachers and in schools with teachers of both genders. I also examine a few mechanisms in the paper that can explain these results. I found no differences by either gender or by the gender mix of the competition group in teachers' awareness and familiarity with the program and its rules, and in effort and teaching methods. However, large gender differences emerge in teachers' expectations about success in the competition and in their perceptions about the effectiveness of the incentive scheme in improving students' achievements. The relative pessimism of female teachers about their success in the tournament and the effect of the program on students' achievement are on average very similar in the two types of competition environment. Examining directly the impact of the bonus program suggests that it did not vary by type of competitive environment and also by male and female teachers. For example, female teachers in only female competition groups and female teachers in mixed gender competition groups were equally effective in improving their students' achievements as a response to the financial incentive.

I. Introduction

The past literature on the gender wage gap provided explanations that rest on gender differences in abilities and preferences and hence in occupational self-selection (Polachek 1981) and on discrimination in the workplace, which leads to differential treatment of men and women with equal preferences and abilities.[1] However, recent studies (Gneezy, Niederele and Rustichini, 2003; Gneezy and Rustichini, 2004; Niederle and Vesterlund, 2007, Passerman, 2007) suggest an additional explanation for gender differences in earnings: women may be less effective than men in competitive environments. These studies test whether men and women differ in their ability to perform in competitive environments, using young adults or children in a controlled lab setting or in sports competitions which allowed for the precise measurement of performance, and which excluded any discrimination or expectation of discrimination. The behavior of men and women in a competitive environment, it is argued, may differ because of differences in skill, talent, and beliefs.[2] In this paper I examine the hypothesis of gender differences in competitiveness in a real work place with adult participants. In particular, I examine how individual performance is affected by the gender mix of the competitive group. The competition is among math, English and Language (Hebrew or Arabic) teachers who participated in a rank-order tournament that rewarded teachers with large cash bonuses based on the test performance of their classes relative to the performance of classes of other teachers in schools. The tournaments, one for each subject, were part of an experiment with individual teachers' incentives implemented in the 2001 academic year in forty nine high schools in Israel. Teachers were awarded bonuses according to their ranking in the tournament. Ranking was based on a value added measure calculated by the difference between the actual mean performance of the teacher's class and a value predicted on the basis of a regression that controlled for the students' socioeconomic characteristics, the level of the course/exam (measured by its credits), and a fixed school-level effect.

1 See Black and Strahan, 2001; Goldin and Rouse, 2000; Wennerås and Wold, 1997; Bertrand and Hallock, 2001). For a review on gender differences in wages, see Blau and Kahn [2000].

2 Gneezy, Niederele and Rustichini (2003) suggest that a competitive environment may produce differences in behavior as subjects adjust their best choices to different strategic environments. In particular, if subjects believe (even if incorrectly) that men are more skillful at solving mazes than women, and effort is costly, one would expect males to put in more effort than females in the winner-takes-all tournaments. A different explanation is based on the argument that preferences over outcomes (that is, over individual effort, payment, and performance) are not independent of the institutional setup in which they are obtained, and in particular of the competitive nature of the institution. The crucial element in this argument is that male and female preferences are affected differently by changes in the institution (its competitiveness, gender composition, etc.).

Therefore teachers were told that they are competing against teachers of the same subject in their own school.

The competitive groups differed in their gender mix. In some all the subject's teachers were female or male and in others there were both male and female teachers. This between groups variation in gender composition allows the comparison of female and male teachers' performance in two distinct competitive environments, competing with teachers of own gender versus competing with teachers of both genders. In particular it allows the examination of whether women succeed differentially when competing only against women or also against men. However, as female teachers were not assigned randomly to one of these two environments, potential endogenous and selective sorting might be a concern. The identification strategy I propose in this paper relies on several natural experiments that suggest that the type of the competition group is practically random. First, none of the forty nine schools had only male or female teachers and having single sex teachers of a given subject is relatively unique. For example, the proportion of female teachers in schools that had a single sex competition groups in one or more of the subjects was 0.59 while in other schools this proportion was almost the same, 0.61. Second, in many of the schools the gender mix was different across the three competition groups. Such significant within school variation suggests that the gender composition of any of the three competition groups in a school was not related to any particular school characteristic and that it is most likely random. Therefore, this within school variation can be exploited for identification by running school fixed effects regressions. Another natural experiment observed in the data is the fact that in many schools some of the English, math and languish teachers did not participate in the competition because their class was not scheduled to take a matriculation exam in the year of the program. Therefore the gender composition of the overall teaching staff of a given subject and that of the respective competition group were not always identical. Since such an event is most likely random, the sub sample that includes all such groups represents a quasi randomized trial. However, the random trial sample includes only few female or male only competition groups, which limit the usefulness of this approach for identifying the treatment effect of interest.

To substantiate my claim that the gender mix of the competition groups is random, I provide evidence in the paper that the three groups of teachers defined by the group gender mix are identical in

their demographics, education, parental schooling and also in the observable characteristics and lagged outcomes of their students. These balancing tests do not necessarily provide a proof of random assignment, as the assumption requires that there be no correlation between treatment status and both observable and *unobservable* background characteristics. However, adding school fixed effect to the analysis and focusing on the within schools variation in the type of competition groups, allow also to control for any potential *unobservable* but constant heterogeneity at the school level. Such school level fixed effect control account also for any *unobservable* heterogeneity at the students at large because the same students took all three exams. Similar perfect balancing is also obtained in the quasi randomized trial sub-sample.

The evidence suggest that the average winning rate and average awarded bonuses did not differ by gender: 42.9 percent of men and 43.6 of women won a bonus. For example, the average conditional (on winning) bonus was $2,790 for men and $2,736 for women. Most important for the hypothesis of interest in this paper, the ranking and bonuses won by women did not differ, on average, between teachers in competition groups with only female teachers and in schools with teachers of both genders. I also examine a few mechanisms in the paper that can explain these results. I found no differences by either gender or by the gender mix of the competition group in teachers' awareness and familiarity with the program and its rules, and in effort in teaching methods. However, large gender differences emerge in the expectations about success in the competition and about the effectiveness of the incentive scheme in improving students' achievements. However, these relatively large male-female differences did not vary by the gender mix of the competition environment.

Another related and interesting question is whether the program effectiveness varied by the gender mix of the competition group. Addressing this question requires evidence about the counterfactual which we base on a comparison group of schools that did not participate in the program. In Lavy (2007) I describe the comparison group derived from a natural experiment based on a measurement error in the assignment variable to the program. Within a sub-sample of the eligible schools this measurement error randomly assigned some schools to treatment and others to a control group. Based on balancing tests, similar to those described above, I show that the three groups of teachers defined by the nature of the gender work environment (only female or male teachers and

teachers of both genders) are identical in observable characteristics of teachers and their students. Lavy (2007) demonstrates that the pay for performance scheme led to significant improvement in the pass rate and test scores of treated students and the evidence presented in this paper show that these gains did not vary by the gender mix of the competition group. When I allowed the treatment effect within competition groups to vary the teacher's gender I find that female teachers in female only groups are as effective as female teachers in mixed gender groups and also as male teachers in these groups. A similar analysis comparing male teachers in male only and in mixed gender groups show also no differences in productivity between the two groups.

The result that the performance of female teachers was not different in the two types of competition environment is different from the evidence obtained from lab experiments and cited above. It is however in line with evidence presented in Kate, Arcidiacono and Walsh (2003) who also examined a non-lab setting and found that in the television game 'The Weakest Link', the performance of female participants was not affected by their opponents' gender. In a follow up study, Kate, Arcidiacono and Walsh (2008) show that the results from the television game differed from those based on replicating the television game in a lab experiment that used much lower stakes, and that increasing the stakes significantly, somewhat bridged the results from the field and lab experiments.[3] However, the results presented in this paper are based on a real work place and not on a television game though it also involved large cash bonuses.

The rest of the paper is structured as follows. The following section presents the experiment, and section III describes the empirical methodology and the results relating to the success of female teachers in the two gender related work environments. Section IV presents evidence on the similarity between math or English female teachers from schools that included only female teachers in these subjects and female teachers from schools that had also male teachers of these subjects. This section also includes the evidence regarding the resemblance between the tournament and comparison teachers. Sections IV-V present results concerning differences by gender in effort, tournament success and productivity. Section 6 concludes.

II. The Pay for Performance Experiment

A. The Experiment

The *Bagrut* (matriculation) examinations are a set of national exams in core and elective subjects that begin in tenth grade, continue in eleventh grade, and conclude in twelfth grade, when most of the tests are taken.[4] The final matriculation score in a given subject is the mean of two intermediate scores. The first is based on the score in the national exams that are administered by an independent agency. The second intermediate score is based on a school-level exam that mimics the national exam in material and format but is administered by the school and scored by the student's own teacher.

In early December 2000, the Ministry of Education unveiled a new teacher bonus experiment in forty-nine Israeli high schools. The main feature of the program was an individual performance bonus paid to teachers on the basis of their own students' achievements. The experiment included all English, Hebrew, Arabic, and mathematics teachers who taught classes in grades ten through twelve prior to matriculation exams in these subjects in June 2001. The program included schools that had a recent history of relatively poor performance in the mathematics or English matriculation exams. Each of the four tournaments (English, Hebrew and Arabic, math, and other subjects) included teachers of classes in grades 10–12 that were about to take a matriculation exam in one of these subjects in June 2001. Each teacher entered the tournament as many times as the number of classes he/she taught and was ranked each time on the basis of the mean performance of each of his/her classes. The ranking was based on the difference between the actual outcome and a value predicted on the basis of a regression that controlled for the students' socioeconomic characteristics, their level of proficiency in each subject, and a fixed school-level effect. Separate regressions were used to compute the predicted passing rate and mean score, and each teacher was ranked twice, once for each outcome. The school

[3] Recent studies addressed the gap between lab and field experiments and highlighted the limitation of the former (see, for example, Harrison and List, 2004, Levitt and List, 2007a and 2007b).

4. Some matriculation subjects are mandatory and many must be taken at the level of three credits at least. Tests that award more credits are more difficult. A minimum of twenty credits is required to qualify for a matriculation certificate. About 55 percent of high-school seniors received matriculation certificates in 2003, i.e., passed enough exams to be awarded twenty credits by the time they graduated from high school or shortly thereafter (Israel Ministry of Education, 2004). In Israel, a high school matriculation certificate is a prerequisite for university admission and one of the most economically important education milestones. Many countries and some American states have similar high school matriculation systems. Examples include 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.

submitted student enrollment lists that were itemized by grades, subjects, and teachers. The reference population was all students enrolled on January 1, 2001, the starting date of the program. All students who appeared on these lists (including dropouts and students who did not take the June 2001 exams, irrespective of the reason) were included in the class mean outcomes at a score of zero.

All teachers who had a positive residual (actual outcome less predicted outcome) in both outcomes were divided into four ranking groups, from first place to fourth. Points were accumulated according to ranking and the total points in the two rankings were used to rank teachers in the tournament and to determine winners and awards, as follows: 30–36 points—$7,500; 21–29 points—$5,750; 10–20 points—$3,500; and 9 points—$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.[5]

The program included 629 teachers, of whom 207 competed in English, 237 in mathematics and 172 in Hebrew or Arabic language. Three hundred and two teachers won awards — of whom 94 were English teachers and 124 were math teachers.

B. The Data

The data I used in this study come from several administrative data files and from a survey the Ministry conducted with the participants (teachers) in the program. An administrative data set provided the following information on all participating teachers: gender, subject of teaching, their school id, i.d of different classes they teach, whether they won and the award size. The teacher survey was conducted during the summer (July-August) following the experiment. The Ministry contracted with a private company (Taldor) to administer a telephone survey of all the teachers in the program. All the subjects were sent a letter in advance, which explained the purpose of the survey and that they would soon be contacted by phone by Taldor. 102 teachers were not interviewed because they could not be contacted by phone or there phone number was not available and 45 teachers refused. 482 teachers completed the survey which represents more than 82% response rate among people whom we were able to contact.

[5] For more details, see Israel Ministry of Education, High School Division, "Individual Teacher Bonuses Based on Student Performance: Pilot Program," December 2000, Jerusalem (Hebrew).

The survey questionnaire included 37 questions, and usually took between 15 to 25 minutes to complete. The survey provided the following information: the grade and credits (3, 4 and 5 credits) of each class the teacher taught, details of his awareness and opinion regarding the program, teacher's subjective assessment about the likelihood that she/he would win a bonus, details about teaching methods (tracking in class, individualized instruction and so on), whether the teacher added after school instruction time during the year and before the matriculation exam period, whether effort and attention was targeted to particular groups of students (weak, average and strong students) and the following background characteristics: gender, age, years of teaching experience, highest degree completed and the academic institution attended, marital status and number of children, country of birth and parental education (mother's and father's highest degree completed). Some of this information was also available in an administrative data file that included all high school teachers in the country and we have used it to complete information for teachers that were not interviewed in the survey and also to compute school level means of the gender composition and also of the competition teams in control schools.

The data for a student's achievement in the 12^{th} grade matriculation exams came from an administrative file that provided the full academic records of each student for the Bagrut exams during high school (grades 10–12) and student characteristics (gender, parental schooling, family size, immigration status-students who recently immigrated). The information for each Bagrut exam included its date, subject, applicable credits, and score. A complementary administrative file provided school level information such as its id, whether it is a Jewish or an Arab school, the religious orientation (secular or religious) of the Jewish schools, and each school's matriculation rate in the years 1999–2001.

III. Identification and Empirical Methodology

The objective of this study is to test whether performance in the incentive tournament varied by gender and whether it was affected by the gender mix of the competition group in school in each subject (the 'treatment'). Since the experiment was not designed specifically to answer this question, teachers were not assigned randomly to 'types' of competition groups in terms of gender composition. However, the sample of schools that participated in the tournament might have produced natural

variation in the proportion of female teachers among the math, English and Hebrew teachers. In particular, this natural variation may have created groups with only female math, English or language teachers in some schools and in other schools groups with mixed genders. Of course this natural process could also have led to within school across groups (subjects) variation in the gender composition. The methodology in this paper exploits this potential natural variation to identify the effect of the treatments defined above.

Table 1 top's panel presents the distribution of school, competition groups and male and female teachers by type of competition groups. Forty nine schools participated in the incentive tournament and therefore there were 49 competition groups in Math, 48 groups in English but only 35 groups in language (Hebrew and Arabic) because 14 schools replaced the language subject with one of the other compulsory matriculation subjects. Therefore there were 132 school-subject level competition groups of which 45 included only female teachers (223 teachers), 70 mixed-gender groups that included in total 149 male and 294 female teachers and 17 groups of only male teachers (58 teachers). The three groups will be denoted as FO (female only), MO (male only) and FM (mixed gender).

The basic statistical strategy can therefore be based on comparing the performance of tournament participants by the three types of competition groups (treatment). A potential problem of course is that the non-random assignment of teachers to FO, MO or FM may result in groups that are different in aspects that may confound the effect of treatment. However, several statistics suggest that the incidence of single sex or a mixed gender competition groups was actually random. First, all 49 schools had both male and female teachers among there staff. The school level proportion of female teachers in the mixed gender competition groups was 0.61 and in the single sex competition groups it was very similar, 0.59. Second in 34 schools there were more than one group type and only in 15 of the schools all the three groups were of the same type, 10 of them are with mixed gender composition. The very low incidence (5 out of 49 schools) where all three competition groups in a school are single sex suggests that being a single sex competition group is not correlated with school observed and unobserved characteristics. Therefore the first identification strategy I use is to exploit the within school variation in the type of competition groups in terms of their gender mix and estimated models

with school fixed effects. The identifying assumption is that within school variation in competition type is random.

Another natural experiment that can be exploited for identification is based on the fact that in many schools some of the English, math and languish teachers did not participate in the competition because their class was not scheduled to take a matriculation exam in the year of the program. For example, a math teacher who taught during the program year an 11^{th} grade class whose matriculation exam was planned for the following year (at end of 12^{th} grade), was not eligible to participate in the bonuses program. Such cases caused in some schools the gender composition of the competition group to be different from that of the overall teaching staff of that subject. The lower panel of Table 1 shows that this was the case for 30 percent of the single sex competition groups. This proportion is 26 percent for the female only competition group (in 59 of the 223 female only competition groups the roster of all teachers in that subject was actually of mixed gender) and 45 percent for the male only competition group. Based on this natural experiment I can define a sub sample of groups where the overall teaching staff of a given subject is of mix gender, but in some cases the competition group is FO or MO. The identifying assumption is that within this sub sample the competition type is random and can be seen as a randomized trial. The limitation of this strategy, however, is that the randomized trial sample includes only 12 FO groups and only 9 Mo groups.

The rich data available allow checking how similar are the three competition groups in various samples and in many dimensions; including teachers' demographics, their schooling attainment and its quality, their parental schooling, and also a range of background variables of their students and school (including lagged achievements that preceded the experiment).

A. Evidence on the Validity of the Identification Strategy: Balancing Tests

The key assumption for the identification strategy outlined above is that the distribution of teachers across the three types of groups is random. To assess the 'observable' part of this assumption I check whether the various characteristics of the teachers, their students and schools are correlated with the three treatment indicators. If teachers are indeed randomly assigned to one of the three teams

types, I would expect to find no significant correlation. This is not necessarily proof of random assignment, as the assumption requires there to be no correlation between the treatment indicators and both observable and unobservable background characteristics of teachers. However, the lack of a significant relationship between the three work environments and observable characteristics suggests that it is unlikely that such a relationship exists with the unobservable characteristics.

Table 3 presents the "balancing tests" for teachers' background, namely a comparison of teachers' characteristics between the three treatment groups. Since the first identification strategy is based on within school comparison, the balancing tests are also based on regressions that include school fixed effects. I present first in column (1) the mean of all male teachers and in column (2) the female-male difference. Male teachers are on average 44 years old, have 18.5 years of teaching experience, are married in 82% of cases, have 1.4 children, 15.2% have a teaching certificate with a degree (not a BA) from a teacher's college, 42.1% have a BA degree, 33.9% an MA degree and 8.8% a PhD. The mean of father's years of schooling is 10.7 and the respective mother's mean is 10.0. As seen in column (2) there are no differences between male and female teachers except in terms of age and years of teaching experience. Female teachers are four years younger and therefore have correspondingly about three years less teaching experience. Fewer women have a PhD degree but this gap is compensated by higher proportion of women with an MA degree. It is important to note that there is no significant difference in salary rank by male and female teachers: women mean rank is lower by 8 percent but it has an estimated large standard error. The salary rank indicator has values from 1 to 17 and most teachers are in ranks 3 and 4.

Column 3 presents the means of teachers from only female groups and column 4 presents the differences between these means and the respective means of female teachers in mixed gender groups. There are no differences at all between these two groups. This perfect within school balancing between these two groups of female teachers is central in this paper because the main hypothesis of interest concerns a comparison between these two groups.

Column 5 presents the differences between the means of female and male teachers in mixed gender groups of teachers. Again, these two groups look identical except for the difference in age and years of teaching experience that were seen in column 2.

Columns 6-7 present the comparison of male teachers in MO teams to male teachers in FM teams. The two groups are identical in terms of their demographic characteristics. However, male teachers in FM groups have a lower proportion of teachers with an MA degree but higher proportion of B.A and PhD degrees. A more significant difference is the significantly higher father's years of schooling among male in FM though this difference is not observed for mother's years of schooling.

Table 3A replicates the balancing tests and analysis for the quasi randomized sub sample. The results are very similar to those presented in Table 3 except that in the comparison of male teachers in MO and FM groups, the first group has older and more experienced teachers.

IV. Results

A. Simple Differences in Performance by Gender and by Competition Group Types

I use the following three measures of teacher's performance: an overall ranking of a teacher in the tournament, a 0/1 indicator of winning a bonus and the amount of the bonus. There are two rankings of teachers, one in the competition based on the pass rate and one based on the average score. Since these two rankings are highly correlated, I only use the percentile ranking of their average. However, the results are identical when I use each of the two base rankings instead of their mean.

Table 3 presents the frequency distribution of the overall and the within competition-type proportion of winners. The win rate ranges from zero to 0.8. There are 12 groups (4 of the FM type, 7 of FO and 1 of MO) where none of the participants won an award. Table 4 presents the mean for these three performance measures by gender and by type of competition groups. On average, men and women had a similar success rate in the tournament. The mean ranking of men was 50.3 and that of women 50.4 and the negligible difference between the two is not significant. There are also no gender differences in ranking based on the pass rate or on the test score. Among men 42.5 percent won an award while for women the respective rate was 42.9. The mean bonus for men is $1,203 and for women it is $1,216. The average award conditional (on winning) among men was $2,790 and among women it was $2,736. Examining the respective gender differences in each of the three tournaments (English, math and language) separately yields very similar results.

Columns 3-4 present a performance comparison of female teachers in FO and FM competition groups. The mean outcomes are marginally higher in the FM group: mean ranking is 51.0 versus 49.8 in FO group, proportion winners is almost 10 percent higher in FM (44.9 versus 40.1) and the mean bonus is 20 percent higher in FM ($ 1,297 versus 1,080). In the next section I test whether these differences in favor of the FM group are statistically significant and whether they remain positive in a controlled comparison.

The comparison between male teachers in MO and FM suggest marginal positive differences in favor of MO teachers but these are very small and most likely not significant different from zero as seen in the next section.

B. Controlled Regression Estimates of Gender Differences by Competition Group Type

Our major interest in this paper is whether female and male teachers were affected by the gender mix of the competition group. For this purpose I estimate the following model:

$$R_{ijs} = \alpha_s + \beta X_{ijs} + \gamma F_{ijs} + \delta FO_{is} + \phi MO_{is} + \mu F_{ijs} + \delta E_{ijs} + \vartheta M_{ijs} + \varepsilon_{ijs} \quad (1)$$

where R is the teacher performance measure, X is a vector of teacher's characteristics, F denotes a female teacher, FO is the female only competition group, MO is the male only competition group and E and M are discrete indicators for English and math teachers, respectively. The main parameters of interest are θ and δ and α_s are the school fixed effects that are included in some of the specifications.

Table 5 presents parameter estimates of regression where the dependent variables are the three principal performance measures, teacher's ranking, whether a bonus was won and its amount. Column 1 presents the mean outcome for men (the constant in the regression) and the simple female difference. The specification presented in column 2 add dummy indicators for the math and English tournaments and in column 3 the two treatments, FO and MO are added and the group left out is the mixed gender groups (FM).

No significant treatment effect for the two treatment indicators is estimated for all three performance measures (column 3). The FO estimates are negative but all three have large estimated standard errors. The highest t-statistic (-1.6) is that of the mean rank performance measure. The

estimates of the MO group are negative for all three measures but they are very small and have large estimated standard errors, therefore they are viewed as practically zero.

In column 4 I present estimates from a specification that includes also teachers' characteristics as controls. These controls include all the background variables for which balancing tests were presented in Table 3. The treatment effect estimates are still not different from zero. The estimates of the control variables are presented in Table 6. The only variable that is significant in this specification is the teacher's salary rank (highest rank is 1 and lowest is 17) and it has a negative sign, implying that the higher the teacher's salary, the better is the teacher's performance in the tournament. Since salary rank is mostly a function of age, years of teaching experience and education, the estimate of the rank variable captures most likely the effect of variation in salary rank determined by unobservables. When these other determinant of salary rank are dropped from the equation the effect of the rank variable is still negative, though smaller by about 30 percent and less precisely measured thought it is still significant or marginally significant. The implication of this result is that financial incentives are more effective among teachers who were found to worth of a promotion beyond the rank they deserve based on their age and their formal schooling. The 'unobservables' that account for the higher salary rank are also positively correlated with its success in the tournament and they perhaps capture unobserved teacher's quality. Allowing for the 'salary rank' variable to vary by gender shows that among women this negative effect is marginally larger ('more negative') but the difference from the estimated effect for men is not statistically significant.

Another interesting result to note is that the teacher's schooling parameter estimates are all negative and not significant except for the indicator of an MA degree. When the salary rank variable is dropped from the regression the negative estimated effects of teacher's schooling are still negative and significant. This pattern is consistent with previous findings of studies that estimated education production functions and found no significant relationship between teachers' schooling levels and students' achievements.

I also estimated a specification where the treatment effects were allowed to have different effects in each of the three subjects. The basic results were unchanged and therefore these results are not reported here.

Column 5 reports estimates from a regression that included also school fixed effects. Any potential effects of school level variables are accounted for in this specification and the estimates are based on within school variation in the type of the competition groups in the three subjects. The estimates of the FO and MO indicators change signs in comparison to estimates without school fixed effects but given their estimated standard errors they remain statistically not different from zero

Columns 6-8 report estimates from regressions of specification similar those of column 3-5, respectively but the FO and MO indicators were replaced in the regression by a continuous measure of the proportion of female teachers in the competition group and its interaction with the female indicator. The estimates of these two variables are never statistically significant in any of the specification for each of the three outcomes.

Table 5A reports results based on the quasi randomized trial sample. The pattern of estimates is very similar to those presented in Table 5. The only exception is the estimated effect of FO on the bonus size which is now positive and significantly different from zero. Since this positive effect is not paralleled by a similar effect on the rank and probability of winning, I tend to discount it and view it as spurious.

B. Gender Related Differences in Program Awareness and Response by Gender composition of Competition Teams

A post program survey with teachers added information about their awareness to the program, their opinion about its efficacy in improving students' achievements and about teaching methods and additional effort. I find no differences by gender in program awareness and knowledge of its details. 90.1% of men and 91.7% of women responded that they knew about the program and 66.3% of men and 67.9% of women said that they received an explanation about it. However, only 51.2% of women and 47.4% of men thought the explanation they received was satisfactory and yet, 62.2% of men and 59.6% of women claimed that they are sufficiently familiar with the ranking criteria for teachers and for bonus winning. These results are presented in the first column in each of panels A-D in Table 7 and their overall pattern does not reveal any gender related differences in these program related variables.

The respondents' answers to the questions about teaching methods do not reveal as well any differences between male and female teachers (first column of panels E-H). For example, 59.3% of male and 54.9% female teachers said that they relied on individualized instruction, 54.7% of male teachers and 52.8% of female teachers grouped their students by ability during lessens. None of these minor gender differences were significant. However, almost all teachers (98.8% of male and 93.8% of female) reported that they adopted their pedagogy to their students, yet the implied small gender difference is still significantly different from zero.

Another dimension of similarity between men and women teachers is in terms of their effort. The questionnaire asked teachers "during the academic year did you have added additional instruction beyond the regular school hours?" and teachers had to choose one of the following: "1. No 2. Yes, during the period before the matriculation exam 3. Yes, throughout the year". Among male teachers 81.9% chose answers 2 or 3 while the respective rate for female teachers was 80.1%. Men added on average per week 2.55 hours while women slightly less, 2.12 hours per week, but there were no differences in how this additional instruction time was targeted to students of different abilities. An almost equal proportion (61% and 62.7%) of male and female teachers reported that the additional, voluntary, instruction time was their own initiative. These results are presented in the first column of panels' I-K of Table 7.

In contrast to the above similarities by gender, I find large gender differences in the teachers' opinion about the program and about of their chances of winning a bonus: 75.8% of the male teachers thought that the program would lead to improvements in their students' achievement while only 60 % of the female teachers shared this view. The implied 15 percentage point difference is significantly different from zero. Furthermore, 3 out of 4 men (76.3%) thought that they would win an bonus, but only 3 out of 5 female teachers (61.0%) had such trust in their winning ability. This 20 percent gap in self-confidence (a difference of 15.3 percentage points relative to 76.3%) is significantly different from zero. Note, however, that despite the fact that women have lower expectations about their ability, we found no difference in performance by gender. Niederle and Vesterlund (2007) report a similar result, that men are substantially more overconfident and that there are no gender differences in performance (in a lab experiment to solve simple two digit addition problems). However, they

concluded that this gender gap in self confidence played an important role in explaining the gender gap in competitive tournament entry, as 73 percent of the men selected the tournament incentive scheme, while only 35 percent of the women made this choice. In the teachers' tournament studied in this paper participants were not offered an alternative to the competitive incentive scheme.

The second to fourth column in each of the panels in Table 7 reports estimates of the effect of the type of the competition groups (FO, MO and FM) based on the first four specifications used in Table 5. Two results should be noted. Firstly, the conditional gender differentials effects are not very different from the unconditional differences reported above. Secondly, the basic patterns of similarity or differences between men and women teachers do not vary with the nature of the competition group with only the following exception. The self confidence of male teachers in winning an bonus is much higher in only male groups than in mixed gender groups.

V. Does Program Effectiveness Vary by the Nature of the Competition Group?

A related question is whether the impact of the program on students' achievements was different by the gender composition of the teachers in each group. In an earlier paper (Lavy 2007) I evaluated the effect of the program on three outcomes in math and English.[6] For the student's outcome I adopted the two criteria that were used to measure teacher's performance, the average score and the pass rate in the matriculation exam in the relevant subject. However, I also evaluated the effect on the test taking rate of students because a change in this outcome could account for some or all of the gain in the average score and the pass rate. The design of the program enabled the implementation of a quasi randomized trial identification strategy based on two features of the program: assignment of schools to the program based on a threshold function of an observable and a measurement error in this variable. Schools were included in the program if their 1999 matriculation rate was equal or lower than a critical value (45 percent). The administrators of the program, unaware that the assignment variable used was measured erroneously, assigned some schools to the program mistakenly. As the measurement error was essentially random and unrelated to the potential outcome,

[6] The effect of Hebrew/Arabic outcomes was not evaluated because schools had the option of replacing this subject with one of the compulsory subjects in the matriculation program and some schools did exercise this option.

the group of schools not too far from the threshold was assigned to treatment at random, condition on the true value of the assignment variable. I also used differences-in-differences estimates in this natural experiment sample based on before and after data in treatment and control schools. The results indicated that incentives increased student achievements by increasing the test taking rate as well as the conditional pass rate and test scores in math and English exams, but mainly for students in the lower half of the ability distribution as measured by their lagged achievements in high school. The improvement in the conditional outcomes of these students accounted for more than half of the increase in the unconditional outcomes in math and somewhat less than half in English. These improvements appeared to result from changes in teaching methods, after-school teaching, and increased responsiveness to students' needs and not from artificial inflation or manipulation of test scores.

In this section I rely on the randomized trial sample of treatment and control schools. In the first part of the section I allow for heterogeneous treatment effect by the indicators FO and MO. I therefore examine if the effectiveness of the program was different in FO or MO groups in comparison to FM groups. Unlike the first part of the paper where estimation was based on samples of teachers, here the unit of observation is the student. To allow for a larger sample and more efficient estimation of the heterogeneity in treatment effect (by FO, MO and FM), I pool the English and math students' samples as I did in studying teachers' performance in the previous section. I therefore first replicate the basic estimation of Lavy (2007) based on the pooled English and math samples and than estimate a model that allows for the main effect of FO and MO as well as interactions between these two variables and the program treatment.

A. Balancing Tests by the Gender Composition of Competition Teams

I first checked whether the various characteristics of the students and schools in the randomized treatment samples are correlated with the three treatment indicators of the competition types groups. Table 8 presents these balancing tests. Column 1 presents the means of student characteristics in groups of only female teachers and columns 2 presents the differences between these means and the respective means of students in groups with both male and female teachers. Overall the two groups of

students are very similar in school and student characteristics and also in terms of lagged achievements in English and math. The only meaningful difference is in the proportion of students in Arab schools.

Column 3 presents the means of student characteristics in groups of only male teachers and columns 4 presents the differences between these means and the respective means of students in groups with both male and female teachers. There are no significant differences between the two groups in school level outcomes such as the mean Bagrut rate in the two years prior to the program. There are also no significant differences in lagged outcomes in math and English. However there are large differences in parental schooling and number of siblings and also in lagged mean overall achievements such as lagged total credits and lagged mean score. These differences resemble the differences found among teachers of these students, again suggestion caution in interpreting the evidence about the comparison of male teachers in MO and FM competition groups.

B. Program Impact Differences by Gender Composition of Competition Groups

Table 9 presents the estimated effects of the program for the three types of competition groups. The estimates are based on panel data that pool the 2000 and 2001 cohorts of students for difference in differences estimation. Columns 1-4 presents estimates using the sample of students in the lower half of the ability distribution and columns 5-8 presents the estimates for the upper half sample. Columns 1 and 5 presents average treatment effect estimates for all types of competition groups. These results suggest that the program had a significant effect on the outcomes of students in the lower half, and zero effect on students in the upper half of the ability distribution. For lower half students the program raised the test taking rate by 7.2%, the test pass rate by 12.4% and the average score by 6.8 points. These are relatively large effects and the largest is that of the pass rate which was raised by almost 25 percent. On the other hand the respective estimates for the upper half are all practically zero. These results are very similar to those reported in Lavy (2007) where the estimation was done separately for math and English while here we pooled the two samples together.

Based on the estimates in columns 2-4 and 6-8 we conclude that the program was equally effective or ineffective in the three types of groups defined by their teachers' gender composition. This similarity is almost perfect in terms of the size of the estimated parameters except for the estimate of the effect on test taking in the MO group which is practically zero while in the FO and FM groups it is

about 8% in both. For example, at the bottom half of the distribution the point estimates of the effect on the pass rate is identical for all three groups, 12.4, 12.9 and 11.6% in FO, MO and FM, respectively. The effect on the test score of students in the lower half is also identical for all three groups, 6.7, 6.9 and 7.0%, respectively. The estimated effect for all three outcomes for the upper half is not significantly different from zero for all three types of competitions. However, I should note that the mean of the outcomes for the control group are marginally lower in the mixed gender group, especially in the 1st and 2ed quartiles which implies that the effect size are larger in the FM group than in the FO and MO groups. For example, the program improved the testing rate by 13 percent in the FM group and by 10 percent in the FO group. For the pass rate the respective rates are 29 and 19 and for the average score they are 18 and 14. This evidence implies that on average the program was marginally more effective in FM groups than in FO or MO groups. However, given the estimated standard error of these estimates, the hypothesis that these estimates are not different from each other cannot be rejected.

The remaining open question is whether the teachers in FO groups were on average less effective than female teachers in FM groups and the respective questions for male teachers.

B. Program Impact Differences by Teacher Gender and by Composition of Competition Groups

The results presented above suggest that the effect of providing incentive to teachers was equally successful in raising students' achievements in groups with only female teachers and in groups with both male and female teachers. In this section I examine whether female teachers in both groups were also equally effective. For answering this question it is sufficient focusing only on the sample of treated schools as there is no need to compare it to evidence from a comparison group. This approach allows doing the analysis not only for the randomized trial sample of 18 treated schools but also for all the 49 treated schools. I estimate the following model:

$$Y_{ijs} = \alpha_s + \beta X_{ijs} + \gamma F_{ijs} + \delta FO_{is} + \phi MO_{is} + \mu F_{ijs} + \delta E_{ijs} + \vartheta M_{ijs} + \varepsilon_{ijs} \qquad (2)$$

Where Y is student's outcome, X is a vector of students and school predetermined characteristics, F is an indicator for female teacher, E and M are indicators for English and math, respectively, and α_s is a school level fixed effect.

In Table 10 I present estimates based on the sample of the lower half of the lagged score distribution. I focus on this sample because the program had no effect on students in the top half in any of the competition groups. Columns 1-2 present estimates based on the randomized sample of 18 schools and columns 3-4 present results based on all 49 schools. Columns 1 and 3 are derived based on a specification without any controls (except a subject main effect for math) and columns 2 and 4 show estimates from a specification that includes as controls school and student level background characteristics. These models are estimated for the three outcomes; the testing rate (panel A), the pass rate (panel B) and the average score (panel C).

The first row in each panel presents the estimates of the differences between female teachers in female only competition groups and female teachers in mixed gender competition groups. The second row in each panel presents the estimates of the differences between female teachers in mixed gender competition groups and male teachers in these groups. The third row in each panel presents the differences between male teachers in male only groups and mixed gender groups.

The overall pattern of the estimates shows no significant differences between female teachers in FO and FM groups. This result is seen in both of the samples, the randomized trial and the full sample, and it is not sensitive at all to whether controls are included in the regressions. The average test rate is 0.918 and there is no difference between female teachers in the FO and FM groups (panel A column 1). Adding controls does not change this equality; the mean test rate is lower by 0.001 in FO than in FM but this infinitesimal difference has a standard error of 0.018. The average pass rate for female teachers is higher by 0.023 (s.e. 0.031) in FO than in FM. The average test score for female teachers is higher by 3.44 (s.e. 2.73) in FO than in FM.

Another interesting result is that there are no differences at all between male and female teachers in FM competition groups. The controlled estimates are all positive, indicating higher effects of female teachers on students' achievements, but they are very small and not significantly different from zero. For example, the estimated effect on average test score of female teachers in FM groups is higher than that of male teachers in these groups by 3.44 (s.e. 2.73) but it is not significantly different from zero.

The results based on the full sample replicates precisely the results from the randomized sample, reinforcing the conclusion that female teachers in FO are as effective as female teachers in FM groups and also as male teachers in FM groups.

A similar analysis comparing male teachers in MO and FM groups show no differences in simple means between the two groups in the randomized and in the full sample. However, adding the controls to the regressions reveal some advantages in the pass rate and mean test score in favor of the former group of teachers. Given the small size of the MO group and the basic imbalance in characteristics between male teachers and their students in the two groups, I think it is hard to draw firm conclusions about productivity differences between them.

VI. Conclusions

In this paper I addressed empirically the question of whether there are gender differentials in performance in a competitive setting in the form of a tournament, when women compete against men. A rank order tournament set the competitive environment where teachers compete against their colleagues in school on the basis of measured improvements in the academic achievements of their students and they are financially rewarded accordingly. I find no such differences and also no overall differences in performance of female and male teachers. As far as I am aware, this study is the first to test of the hypothesis of gender differences in competitiveness based on evidence from a real work place.

These results are different from most recent lab experiments that addressed the same question. This divergence in evidence can be due to the differences between a lab setting and a real work environment and due to differences between the nature of a tournament among teachers who perform real tasks and competition among individuals who compete in solving mazes or running a given distance. Unlike tasks at lab experiments which are instantaneous and last only few minutes or hours, teachers in our experiment had six months to do the best they can to improve their students' performance. Such a lengthy period, which is not unusual in comparison to the time it takes to complete tasks in other real workplaces, allows teachers to plan their strategy, to get feedback from students and supervisors, to observe the actions of their competitors in school and to adjust their

actions. All these is absent in performing tasks in a lab setting and may account for the difference of the results in this study and those based on lab experiments.

The findings reported above are important because of the large proportion of female teachers in schools and the growing interest in incentive programs for teachers in the US, Europe and elsewhere which has led to many new interventions.[7] Many of these programs, whose rationale is the notion that incentive pay may motivate teachers to improve their performance, are based on explicit or implicit ranking within schools. Since the majority of teachers are female and in many schools the teaching staff of some subjects include only female teachers, whether women are less effective under such competitive conditions may have important implications about the effectiveness of financial incentives in schools.

[7] Most recent pay-for-performance programs for teachers in the US include Minnesota's Q-Comp [see http://education.state.mn.us/MDE/Teacher_Support/QComp/index.html], $86 million merit pay initiative, Denver's Pro-Comp, $25 million teachers' pay-for-performance plan (Mitchell, 2005), Florida's E-Comp and STAR programs [http://www.fldoe.org/news/2006/2006_04_05/ValueTable.pdf and http://www.fldoe.org/PerformancePay/, respectively, and Chicago's $27.5 million 2006 pilot of a merit pay initiative for teachers. A recent N.Y. Times article illustrated this trend, commenting that "A consensus is building across the political spectrum that rewarding teachers with bonuses or raises for improving student achievement,…can energize veteran teachers and attract bright rookies to the profession." Other recent interventions of this type include programs in Mobile, Toledo, Columbus, Huston, Charlotte-Mecklenburg and Dallas and scores of other teacher performance-pay experiments are under way nationwide. Many of these programs are financed by federal grants funding experiments based on incentives to educators in 19 states (*Source: Department of Education*). The N.Y Times, **EDUCATION**, June 18, 2007, Long Reviled, Merit Pay Gains Among Teachers By Sam Dillon [http://select.nytimes.com/gst/abstract.html?res=F10715FA395B0C7B8DDDAF0894DF404482]. Pay for teachers' performance has also been implemented in other countries, for example, the U.K's Pay Performance and Management Reform in 2000 (Burgess et al., 2005), the Victorian Government Schools Agreement 2001 in Australia7, Mexico's Carrera Magisterial Program (McEwan and Santibañez, 2005), and Chile's SNED (Vegas, 2005). Smaller scale randomized experiments have been implemented in India (Duflo and Hanna, 2006) and Kenya (Glewwe, Ilias, and Kremer, 2003).

References

Antonovics Kate, Peter Arcidiacono, and Randall Walsh, "Competing Against the Opposite Sex" (June 1, 2003). *Department of Economics, UCSD.* Paper 2003-08.

Antonovics Kate, Peter Arcidiacono, and Randall Walsh, "The Effects of Gender Interactions in the Lab and in the Field," Forthcoming, 2008, *Review of Economics and Statistics*.

Barber, Brad M., and Terrance Odean, "Boys will be Boys: Gender, Overconfidence, and Common Stock Investment," Quarterly Journal of Economics, CXVI (2001), 261–292.

Beyer, Sylvia, "Gender Differences in the Accuracy of Self-Evaluations of Performance," Journal of Personality and Social Psychology, LIX (1990), 960–970.

——, and Edward M. Bowden, "Gender Differences in Self-Perceptions: Convergent Evidence from Three Measures of Accuracy and Bias," Personality and Social Psychology Bulletin, XXIII (1997), 157–172.

Byrnes, James P., David C. Miller, and William D. Schafer, "Gender Differences in Risk Taking: A Meta-Analysis," Psychological Bulletin, LXXV (1999), 367–383.

Conti, Regina, Collins, Mary Ann and Picariello, Martha, "The Impact of Competition on Intrinsic Motivation and Creativity: Considering Gender, Gender Segregation and Gender Role Orientation," *Personality and Individual Differences*, December 2001, 31(8), 1273-89

Corbin, Charles, "Sex of Subject, Sex of Opponent, and Opponent Ability as Factors Affecting Self-confidence in a Competitive Situation", *Journal of Sport Psychology*, 1981, 3(4), 265-270.

Eckel, Catherine C., and Philip J. Grossman, "Differences in the Economic Decisions of Men and Women: Experimental Evidence," mimeo, 2000.

Ehrlinger, Joyce and Dunning, David, "How Chronic Self-views Influence (and Potentially Mislead) Estimates of Performance," *Journal of Personality and Social Psychology*, January 2003, 84(1), 5-17.

Gneezy, Uri., M. Niederle, and A. Rustichini, "Performance in competitive environments: Gender differences," Q*uarterly Journal of Economics*, August 2003, p. 1049-1074.

Gneezy, U., and A. Rustichini, "Gender and competition at a young age," *American Economic Review Papers and Proceedings*, May 2004, 377-381.

Gneezy, Uri, "Executive versus Teachers: Gender, Competition and Self-Selection," Working Paper, in progress.

Goldin, Claudia, and Cecilia Rouse, "Orchestrating Impartiality: The Impact of "Blind" Auditions on Female Musicians," *American Economic Review*, XL (2000), 715–742.

Green, Jerry and Nancy L. Stokey (1983). "A Comparison of Tournaments and Contracts." *Journal of Political Economy* 91: 349-64.

Gupta, Nabanita D., Anders Poulsen, and Marie-Claire Villeval, "Male and Female Competitive Behavior—Experimental Evidence," GATE Working Paper 05-12, Ecully, France, 2005.

Harrison, Glenn and List, John, “Field Experiments," Journal of Economic Literature, December 2004, 42(4), 1009-55.

Holmstrom, B. and P. Milgrom (1991). “Multitask Principal-Agent Analysis: Incentive Contracts, Asset Ownership and Job Design, *Journal of Law, Economics and Organization* 7 (Special Issue), 24–52.

Israel Ministry of Education, The Bagrut 2003 program, an Evaluation”. Jerusalem: Ministry of Education, Evaluation Division, May 2004.

Israel Ministry of Education, Bagrut Test Data 2000, Jerusalem: Ministry of Education, Chief Scientist’s Office, April 2001.

Inzlicht, Michael, and Talia Ben-Zeev, “A Threatening Intellectual Environment: Why Females Are Susceptible to Experiencing Problem-Solving Deficits in the Presence of Males,” *Psychological Science*, XI (2000), 365–371.

Larson, Lia C., “When Girls Stop Competing Against Boys: An Experimental Analysis of the Competitive Behavior of Young Children,” Harvard Honors Thesis, 2005.

Lavy, V. (2002). “Evaluating the Effect of Teachers’ Group Performance Incentives on Students Achievements.” *Journal of Political Economy*, 10 (6), December 2002, 1286–1318.

Lavy, V. (2007) “Performance Pay and Teachers’ Effort, Productivity and Grading Ethics”.

Lazear, E. and S. Rosen. (1981). “Rank-Order Tournaments as Optimum Labor Contracts.” *Journal of Political Economy* 89: 841–64.

Lazear, E. “Performance Pay and Productivity,” *American Economic Review*, December, 2000.

Levitt, Steven and John List, “What Do Laboratory Experiments Measuring Social Preferences Tell Us About the Real World?" *Journal of Economic Perspectives*, Spring 2007, 21(2), 153-74.

Levitt, Steven and John List, “Viewpoint: On the Generalizability of Lab Behaviour to the Field," Canadian Journal of Economics, May 2007, 40(2), 347-70.

Madden, Chris and Kirkby, Robert, “Gender Differences in Competitive Stress”, Perceptual and Motor Skills, June 1995, 80(3, 1), 848-850.

Milgrom, P. and J. Roberts (1992). Economics, Organization and Management, Prentice Hall, New Jersey.

Niederle, Muriel, and Lise Vesterlund, “Do Women Shy Away From Competition? Do Men Compete Too Much?” *Quarterly Journal of Economics*, August 2007. 122(3), 1067-1102..

Paserman Daniele, “Gender Differences in Performance in Competitive Environments: Field Evidence from Professional Tennis Players.” CEPR Working Paper, 2007.

Wennerås, Christine, and Agnes Wold, “Nepotism and Sexism in Peer-Review,” *Nature*, CCCLXXXVII (1997), 341–343.

Table 1 - Distribution of Participants and Competition Groups by Gender Composition of Groups

	Competition Group Type					
	Female Only	Male Only	Mixed Gender			Total
			Female	Male	Total	
	(1)	(2)	(3)	(4)	(5)	(6)
Number of participants*	223	58	294	149	443	724
Number of competition groups	45	17			70	132
Number of schools with all groups under the same category*	4	1			10	15
	Comparing Group Type Classification by Competition Participants and by Roster of Teachers					
Female only group by teachers roster						
Number of participants*	164					
Number of groups	33					
Number of schools with all groups in the same category by competition/roster	3					
Male only group by teachers roster						
Number of teachers		32				
Number of groups		8				
Number of schools with all groups in the same category by competition/roster		0				
Mixed gender group by teachers roster						
Number of teachers	59	26	294	149	443	
Number of groups	12	9			70	
Number of schools with all groups in the same category by competition/roster	0	1			10	

Notes:

* The number of participants (724) is larger than the number of teachers because some teachers participated in the tournament with more than one class.

** Number of schools in tournament = 49

Table 2 - Frequency Distribution of Proportions of Winners by Type of Competition Group

Proportion of winners	Type of Competition Group			Total
	Mixed	Female-only	Male-only	
.000	4	7	1	12
.143	1	0	0	1
.200	0	1	0	1
.250	2	4	1	7
.273	0	1	0	1
.286	2	0	0	2
.308	1	0	0	1
.333	13	5	5	23
.364	1	0	0	1
.368	1	0	0	1
.375	1	3	0	4
.400	6	4	1	11
.417	1	0	0	1
.429	1	1	0	2
.444	3	0	0	3
.455	1	0	0	1
.500	15	15	8	38
.556	1	0	0	1
.571	1	0	0	1
.600	4	1	0	5
.615	2	0	0	2
.667	6	3	1	10
.727	1	0	0	1
.750	1	0	0	1
.800	1	0	0	1
Total	70	45	17	132

Table 3 - Balancing Tests of Teachers' Characteristics by Type of Competition Groups

	Mean of Male (1)	Difference (Female - Male) (2)	Mean of Female in FO (3)	Difference (Female in FO - Female in FM) (4)	Difference (Female in FM - Male in FM) (5)	Mean of Male in MO (6)	Difference (Male in FM - Male in MO) (7)
Age	44.0	-3.963	43.1	-1.10	-4.22	38.9	0.57
	(1.44)	(1.05)		(1.34)	(1.25)		(1.93)
Years of Teaching Experience	18.5	-3.131	17.9	-1.25	-3.166	14.0	-0.21
	(1.49)	(1.23)		(1.13)	(1.45)		(2.25)
Married	.821	-.068	.807	.049	-.088	.741	.092
	(.037)	(.046)		(.062)	(.055)		(.098)
Number of Kids	1.43	.146	1.276	-.001	.198	1.98	-.263
	(.172)	(.164)		(.171)	(.177)		(.561)
Highest Diploma Completed							
Teaching Certificate	.152	-.030	.147	.028	-.052	.080	.091
	(.033)	(.046)		(.041)	(.057)		(.133)
B.A Degree	.421	.057	.424	-.053	.111	.560	.104
	(.042)	(.052)		(.059)	(.061)		(.170)
M.A Degree	.339	.078	.395	.013	.036	.340	-.264
	(.050)	(.063)		(.063)	(.062)		(.122)
Ph.D Degree	.088	-.105	.034	.012	-.095	.020	.069
	(.035)	(.042)		(.021)	(.036)		(.121)
Salary Rank	3.15	-.261	2.972	.072	-.326	2.91	0.727
	(.190)	(.219)		(.171)	(.235)		(.542)
Mother's Years of Schooling	10.7	0.34	11.5	-0.952	.535	9.17	1.53
	0.51	(.522)		(.653)	(.513)		(1.675)
Father's Years of Schooling	9.99	1.14	11.7	-.620	1.28	7.87	3.05
	(.579)	(.599)		(.735)	(.644)		(.638)
Teaches 10th Grade	.150	.006	.193	.072	.014	.155	-.032
	(.031)	(.029)		(.079)	(.026)		(.045)
Teaches 11th Grade	.174	-.005	.193	-.014	.018	.121	.062
	(.037)	(.033)		(.065)	(.035)		(.060)
Teaches 12th Grade	.676	-.001	.614	-.057	-.033	.724	-.030
	(.042)	(.042)		(.088)	(.040)		(.080)

Notes:

1. Standard errors in parentheses are clustered at the school level.
2. FO = Female Only group
MO = Male Only group
FM = Mixed gender group

Table 3A - Balancing Tests of Teachers' Characteristics by Type of Competition Groups in Quasi-Randomized Trial Sample

	Mean of Male (1)	Difference (Female - Male) (2)	Mean of Female in FO (3)	Difference (Female in FO - Female in FM) (4)	Difference (Female in FM - Male in FM) (5)	Mean of Male in MO (6)	Difference (Male in FM - Male in MO) (7)
Age	44.9	-4.144	42.2	-0.68	-4.22	38.8	3.43
	(1.51)	(1.17)		(0.90)	(1.25)		(1.72)
Years of Teaching Experience	19.3	-3.048	18.3	0.47	-3.166	13.3	3.70
	(1.58)	(1.36)		(0.92)	(1.45)		(1.40)
Married	.840	-.079	.746	.012	-.088	.769	.117
	(.037)	(.051)		(.098)	(.055)		(.062)
Number of Kids	1.35	.166	1.15	-.047	.198	2.16	.071
	(.180)	(.169)		(.186)	(.177)		(.635)
Highest Diploma Completed							
Teaching Certificate	.175	-.046	.182	.031	-.052	.136	.153
	(.039)	(.055)		(.052)	(.057)		(.179)
B.A Degree	.392	.128	.568	.133	.111	.545	.023
	(.046)	(.060)		(.068)	(.061)		(.197)
M.A Degree	.336	.002	.227	-.157	.036	.318	-.313
	(.055)	(.063)		(.089)	(.062)		(.183)
Ph.D Degree	.098	-.084	.023	-.006	-.095	.000	.137
	(.040)	(.033)		(.055)	(.036)		(.136)
Salary Rank	3.21	-.340	3.36	.280	-.326	3.04	.942
	(.213)	(.218)		(.411)	(.235)		(.830)
Mother's Years of Schooling	10.8	0.76	11.0	-1.21	.535	7.13	3.37
	.602	(.529)		(.669)	(.513)		(2.27)
Father's Years of Schooling	10.2	1.56	11.4	-.854	1.28	6.76	3.70
	(.648)	(.618)		(.933)	(.644)		(.757)
Teaches 10th Grade	.171	-.001	.237	.077	.014	.308	-.079
	(.036)	(.031)		(.126)	(.026)		(.135)
Teaches 11th Grade	.183	.005	.068	-.116	.018	.115	.037
	(.037)	(.033)		(.044)	(.035)		(.100)
Teaches 12th Grade	.646	-.004	.695	.038	-.033	.577	.043
	(.043)	(.043)		(.103)	(.040)		(.193)

Notes:

1. Standard errors in parentheses are clustered at the school level.

2. FO = Female Only group

MO = Male Only group

FM = Mixed gender group

3. The regressios are based on a sample that exclude groups for which both the roster and the competition classification of the group is female only or male only

Table 4 - Competition Ranks, Proportions of Winners and Bonuses

	Female	Male	Female		Male	
			in FO	in FM	in MO	in FM
	(1)	(2)	(3)	(4)	(5)	(6)
Mean Rank	50.5	50.3	49.8	51.0	50.9	50.0
Rank in the Test Pass-Rate Competition	50.4	50.6	49.9	50.7	50.7	50.5
Rank in the Test Score Competition	50.6	50.0	49.8	51.2	51.2	49.5
Proportion Winners	.429	.425	.404	.449	.431	.423
Bonus: $1,750	.246	.232	.247	.245	.190	.248
Bonus: $3,500	.133	.145	.117	.146	.207	.121
Bonus: $5,750	.041	.034	.036	.044	.034	.034
Bonus: $7,500	.010	.014	.004	.014	.000	.020
Mean Bonus ($)	1,203	1,216	1,080	1,297	1,254	1,201
Number of participants	517	207	223	294	58	149

Note: The bonus figures are based on the exchange rate at the date the program was announced, December 2000 (4 NIS per 1 USD)

Table 5 - Estimates of Effects of the Gender-Mix of Competition Groups on Teachers' Performance

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Mean Rank							
Constant	50.3	48.5	48.4	54.4	56.0	48.5	55.5	57.1
	(1.11)	(1.60)	(1.81)	(9.48)	(12.2)	(1.85)	(9.64)	(13.0)
Female	.205	.477	1.05	-.070	-.047	7.90	2.68	4.12
	(1.49)	(1.65)	(2.18)	(2.76)	(3.10)	(5.04)	(5.64)	(6.69)
FO Group			-.831	-.133	1.80			
			(1.28)	(1.54)	(1.96)			
MO Group			.995	2.00	-.011			
			(2.11)	(2.07)	(2.40)			
Female Prop. in Group (FPG)						1.55	-1.67	-.647
						(4.09)	(4.39)	(5.48)
FPG * Female						-9.94	-3.48	-4.94
						(6.48)	(6.61)	(8.43)
Subject Main Effects		√	√	√	√	√	√	√
Individual Controls				√	√		√	√
School Fixed Effects					√			√
	Winning a Bonus							
Constant	.425	.393	.397	.364	.442	.387	.355	.390
	(.026)	(.028)	(.033)	(.179)	(.234)	(.037)	(.168)	(.231)
Female	.004	.014	.029	.024	.036	.143	.114	.163
	(.031)	(.032)	(.048)	(.054)	(.059)	(.105)	(.111)	(.122)
FO Group			-.036	-.034	.007			
			(.041)	(.043)	(.044)			
MO Group			.008	.012	-.037			
			(.041)	(.054)	(.063)			
Female Prop. in Group (FPG)						.042	-.007	.083
						(.073)	(.109)	(.128)
FPG * Female						-.180	-.132	-.186
						(.124)	(.130)	(.148)
Subject Main Effects		√	√	√	√	√	√	√
Individual Controls				√	√		√	√
School Fixed Effects					√			√
	Bonus Size (NIS)							
Constant	4,865	4,797	4,865	4,444	5,986	4,622	4,770	5,337
	(368)	(643)	(703)	(2,359)	(2,617)	(713)	(2,163)	(2,669)
Female	-52.4	212	470	31.8	18.0	950	-492	-483
	(405)	(420)	(548)	(610)	(666)	(1,126)	(1,302)	(1,546)
FO Group			-635	-474	731			
			(541)	(587)	(596)			
MO Group			103	504	-386			
			(633)	(686)	(1,178)			
Female Prop. in Group (FPG)						566	-568	943
						(1,207)	(1,445)	(1,804)
FPG * Female						-1,207	540	768
						(1,407)	(1,574)	(2,039)
Subject Main Effects		√	√	√	√	√	√	√
Individual Controls				√	√		√	√
School Fixed Effects					√			√
Observations	724	724	724	683	683	724	683	683

Notes:

1. Standard errors in parentheses are clustered at the school level.
2. Regressions in columns (4) and (7) include controls for age, teaching experience, salary rank, overall school's proportion of female teachers, marital status, number of children, education, parents' education, and grade.
3. Regressions in columns (5) and (8) include the same controls as in columns (4) and (7), and also school fixed-effects.

Table 5A - Estimates of Effects of the Gender-Mix of Competition Groups on Teachers' Performance in Quasi-Randomized Trial Sample

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Mean Rank							
Constant	49.9	47.6	47.9	58.5	58.0	46.2	56.7	60.0
	(1.22)	(1.77)	(1.96)	(9.98)	(14.0)	(2.15)	(10.0)	(14.4)
Female	.797	.833	1.00	-.437	.009	11.9	5.75	5.81
	(1.84)	(1.90)	(2.15)	(2.76)	(3.12)	(6.62)	(7.14)	(9.04)
FO Group			-2.04	-.806	.698			
			(1.25)	(1.91)	(2.50)			
MO Group			-.942	-1.17	2.38			
			(2.92)	(2.87)	(3.40)			
Female Prop. in Group						4.51	1.03	-1.30
						(4.82)	(5.39)	(7.04)
FPG * Female						-16.5	-8.68	-8.57
						(9.14)	(9.40)	(12.2)
Subject Main Effects		√	√	√	√	√	√	√
Individual Controls				√	√		√	√
School Fixed Effects					√			√
	Winning a Bonus							
Constant	.417	.386	.395	.399	.438	.352	.351	.340
	(.032)	(.035)	(.036)	(.203)	(.282)	(.052)	(.192)	(.301)
Female	.025	.027	.028	.011	.038	.179	.129	.145
	(.042)	(.041)	(.048)	(.053)	(.058)	(.126)	(.136)	(.161)
FO Group			-.037	-.012	.084			
			(.054)	(.065)	(.054)			
MO Group			-.030	-.043	-.007			
			(.040)	(.064)	(.103)			
Female Prop. in Group						.094	.049	.168
						(.086)	(.136)	(.192)
FPG * Female						-.239	-.170	-.170
						(.163)	(.178)	(.214)
Subject Main Effects		√	√	√	√	√	√	√
Individual Controls				√	√		√	√
School Fixed Effects					√			√
	Bonus Size (NIS)							
Constant	4,743	4,757	4,879	5,363	6,482	4,237	5,619	6,073
	(383)	(788)	(862)	(2,575)	(2,928)	(819)	(2,379)	(3,304)
Female	308	405	452	-30.2	112	942	-988	-1,937
	(484)	(479)	(548)	(625)	(696)	(1,364)	(1,643)	(1,755)
FO Group			-679	-223	1,448			
			(756)	(963)	(642)			
MO Group			-378	162	1,707			
			(834)	(970)	(1,359)			
Female Prop. in Group						1,188	-483	1,214
						(1,298)	(1,638)	(2,328)
FPG * Female						-1,201	1,356	2,895
						(1,907)	(2,219)	(2,360)
Subject Main Effects		√	√	√	√	√	√	√
Individual Controls				√	√		√	√
School Fixed Effects					√			√
Observations	528	528	528	492	492	528	492	492

Notes:

1. Standard errors in parentheses are clustered at the school level.
2. Regressions in columns (4) and (7) include controls for age, teaching experience, salary rank, overall school's proportion of female teachers, marital status, number of children, education, parents' education, and grade.
3. Regressions in columns (5) and (8) include the same controls as in columns (4) and (7), and also school fixed-effects.
4. The regressios are based on a sample that exclude groups for which both the roster and the competition classification of the group is female only or male only

Table 6 - Estimates of Effects of Teachers' Characteristics in the Teachers' Performance Equations

	Mean Rank		Won		Bonus	
	No FE	FE	No FE	FE	No FE	FE
	(1)	(2)	(3)	(4)	(5)	(6)
Math Teacher	.575	1.944	-.016	.030	-321.0	433.4
	(1.460)	(2.357)	(.043)	(.053)	(707.0)	(668.8)
English Teacher	.851	1.188	-.035	-.012	-1212.2	-1006.3
	(2.242)	(3.178)	(.051)	(.063)	(874.2)	(971.3)
Proportion of Female Teachers	-4.09		.003		-166	
in the School	(3.86)		(.086)		(1,345)	
Age	.249	.288	.007	.009	60.8	64.5
	(.237)	(.316)	(.004)	(.006)	(53.6)	(68.0)
Years of Teaching Experience	.110	.135	-.002	-.002	24.5	25.6
	(.300)	(.409)	(.005)	(.007)	(71.4)	(91.7)
Salary Rank	-1.88	-1.87	-.024	-.024	-372	-358
	(.675)	(.754)	(.012)	(.014)	(132)	(149)
Number of Kids	.123	.049	.005	-.002	-62.4	-162
	(.913)	(1.06)	(.016)	(.019)	(200)	(229)
Married	-4.07	-3.55	-.054	-.030	-714	-318
	(3.50)	(3.89)	(.058)	(.066)	(786)	(869)
B.A Degree	-5.13	-5.45	-.072	-.107	-893	-918
	(3.99)	(4.54)	(.067)	(.078)	(960)	(1,095)
M.A Degree	-11.6	-12.4	-.177	-.217	-2,730	-3,134
	(4.48)	(5.08)	.078	.088	(1,102)	(1,231)
Ph.D Degree	-7.64	-6.38	-.019	-.072	-2,362	-1,835
	(8.33)	(9.26)	(.161)	(.161)	(1,694)	(1,923)
Mother's Years of Schooling	.328	.319	-.009	-.008	-1.64	20.0
	(.511)	(.585)	(.009)	(.011)	(146)	(164)
Father's Years of Schooling	-.379	-.424	.003	.002	95.6	66.7
	(.560)	(.638)	(.010)	(.011)	131	148
Teaches 11th Grade	1.12	-.488	.030	.019	-.176	-896
	(4.18)	(4.83)	(.079)	(.089)	(1,214)	(1,428)
Teaches 12th Grade	2.42	.930	.122	.089	617	122
	(4.10)	(5.03)	(.084)	(.097)	(1,246)	(1,400)
Observations	683	683	683	683	683	683

Notes:

1. Standard errors in parentheses are clustered at the school level.
2. These are estimates of the covariates from the regression in columns (4) and (5) of Table 5

Table 7 - Estimates of the Effects on Program Awareness, Effectiveness and on Teachers' effort and Pedagogic Adjustments

	A. Teacher Aware of the Tournament				B. Received Explanation about the Tournament				C. Explanation was Satisfying			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Constant	.901	.925	.989	.972	.663	.683	1.026	0.82	.474	.503	.491	.211
	(.024)	(.035)	(.131)	(.150)	(.045)	(.073)	(.276)	(.301)	(.056)	(.082)	(.273)	(.322)
Female	.016	.004	-.007	-.020	.016	.084	.056	-.027	.038	.157	.136	.127
	(.028)	(.032)	(.030)	(.031)	(.052)	(.052)	(.069)	(.071)	(.063)	(.074)	(.084)	(.095)
FO Group		.016	.025	.037		-.142	-.122	-0.00		-.194	-.196	-.146
		(.040)	(.041)	(.048)		(.075)	(.073)	(.071)		(.085)	(.085)	(.088)
MO Group		-.029	-.032	-.075		.031	-.027	-.338		.176	.270	-.109
		(.055)	(.056)	(.066)		(.099)	(.104)	(.109)		(.121)	(.127)	(.097)
Subject Main Effects		√	√	√		√	√	√		√	√	√
Individual Controls			√	√			√	√			√	√
School Fixed Effects				√				√				√
Observations	608	608	576	576	605	605	573	573	601	601	569	569

	D. Knows the Criteria for Winning				E. Worked with Students in Small Groups				F. Worked with Students Individually			
	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)	(21)	(22)	(23)	(24)
Constant	.622	.728	.904	.537	.686	.605	.844	.624	.593	.564	.819	.721
	(.052)	(.104)	(.223)	(.307)	(.044)	(.079)	(.229)	(.298)	(.039)	(.070)	(.222)	(.276)
Female	-.026	-.007	-.048	-.087	-.058	-.060	-.012	-.010	.042	-.004	-.004	-.007
	(.055)	(.067)	(.083)	(.087)	(.049)	(.063)	(.073)	(.077)	(.050)	(.059)	(.074)	(.076)
FO Group		-.079	-.039	.044		-.012	-.022	-0.04		.031	.022	-.062
		(.085)	(.083)	(.107)		(.054)	(.055)	(.080)		(.082)	(.089)	(.114)
MO Group		-.056	-.093	-.270		.010	-.070	-.112		-.083	-.121	-.328
		(.127)	(.116)	(.129)		(.094)	(.102)	(.136)		(.079)	(.093)	(.119)
Subject Main Effects		√	√	√		√	√	√		√	√	√
Individual Controls			√	√			√	√			√	√
School Fixed Effects				√				√				√
Observations	600	600	568	568	608	608	576	576	608	608	576	576

	G. Divided Students in Class by Level				H. Adjusted Teaching Methods to Student's Level			
	(25)	(26)	(27)	(28)	(29)	(30)	(31)	(32)
Constant	.547	.387	.768	.239	.988	.976	.907	1.074
	(.056)	(.075)	(.199)	(.256)	(.008)	(.017)	(.085)	(.102)
Female	-.019	-.006	.009	-.000	-.050	-.048	-.052	-.062
	(.068)	(.075)	(.071)	(.077)	(.016)	(.015)	(.023)	(.029)
FO Group		.025	.040	.068		-.055	-.071	-0.08
		(.069)	(.074)	(.082)		(.024)	(.022)	(.020)
MO Group		.160	.042	.080		-.034	-.007	-.017
		(.101)	(.089)	(.122)		(.027)	(.031)	(.032)
Subject Main Effects		√	√	√		√	√	√
Individual Controls			√	√			√	√
School Fixed Effects				√				√
Observations	608	608	576	576	608	608	576	576

Notes:

1. Standard errors in parentheses are clustered at the school level.
2. Regressions in the third and fourth columns of each sub-table include controls for age, teaching experience, salary rank, overall school's proportion of female teachers, marital status, number of children, education, parents' education, and grade. Regressions in the fourth column of each sub-table also include school fixed effects.

Table 7 - continued.

	I. Added After School Instruction				J. Number of Hours Added				K. Adding Instruction was Teacher's Initiative			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Constant	.819	.810	.998	1.105	2.551	3.637	5.311	13.84	.605	.611	.583	.893
	(.039)	(.058)	(.188)	(.211)	(.735)	(1.517)	(3.274)	(3.556)	(.052)	(.074)	(.268)	(.247)
Female	-.018	.012	.057	-.002	-.436	-.280	-.617	-1.768	.017	.046	.108	.042
	(.043)	(.063)	(.070)	(.067)	(.832)	(1.100)	(1.330)	(1.595)	(.056)	(.068)	(.070)	(.074)
FO Group		.013	-.006	.004		.122	-.203	0.04		-.032	-.056	-.058
		(.056)	(.057)	(.061)		(.774)	(.962)	(1.670)		(.059)	(.060)	(.049)
MO Group		.081	.061	.035		.821	.791	-1.847		-.063	-.095	-.154
		(.060)	(.065)	(.084)		(1.343)	(1.764)	(2.635)		(.097)	(.087)	(.121)
Subject Main Effects		√	√	√		√	√	√		√	√	√
Individual Controls			√	√			√	√			√	√
School Fixed Effects				√				√				√
Observations	607	607	575	575	179	179	168	168	608	608	576	576

	L. Teacher Thinks Program Will Improve Student's Achievements				M. Teacher Thinks He will Win an Award				N. Teacher Thinks He will Multiple Awards			
	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)	(21)	(22)	(23)	(24)
Constant	.758	.693	1.348	.769	.763	.799	1.260	0.51	.321	.261	.254	.251
	(.046)	(.071)	(.196)	(.284)	(.044)	(.057)	(.154)	(.326)	(.048)	(.067)	(.262)	(.358)
Female	-.158	-.118	-.147	-.166	-.153	-.180	-.201	-.238	-.065	-.071	.008	-.011
	(.044)	(.051)	(.067)	(.074)	(.052)	(.066)	(.073)	(.073)	(.064)	(.086)	(.086)	(.105)
FO Group		.037	.050	.073		.050	.092	0.15		.092	.080	.312
		(.051)	(.051)	(.056)		(.059)	(.060)	(.072)		(.073)	(.080)	(.103)
MO Group		.185	-.018	.049		.039	-.133	-.186		-.001	-.088	-.087
		(.077)	(.079)	(.129)		(.073)	(.074)	(.142)		(.105)	(.111)	(.159)
Subject Main Effects		√	√	√		√	√	√		√	√	√
Individual Controls			√	√			√	√			√	√
School Fixed Effects				√				√				√
Observations	548	548	520	520	485	485	461	461	317	317	304	304

Notes:

1. Standard errors in parentheses are clustered at the school level.
2. Regressions in the third and fourth columns of each sub-table include controls for age, teaching experience, salary rank, overall school's proportion of female teachers, marital status, number of children, education, parents' education, and grade. Regressions in the fourth column of each sub-table also include school fixed effects.

Table 9 - DID Estimates of the Effect of Teachers' Bonuses on ***Math and English*** Outcomes by Competition Group Types

	1st and 2nd Quartiles				3rd and 4th Quartiles			
	All gender comp's	Estimates by teachers' gender comp			All gender comp's	Estimates by teachers' gender comp		
		Mixed	Female-only	Male-only		Mixed	Female-only	Male-only
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Testing rate								
Control group mean	.680	.648	.728	.830	.957	.940	.980	.971
Treatment effect	.072	.083	.076	.015	.005	.000	.015	-.006
	(.034)	(.040)	(.034)	(.038)	(.021)	(.026)	(.017)	(.020)
Pass rate								
Control group mean	.509	.444	.621	.649	.884	.845	.937	.899
Treatment effect	.124	.129	.120	.116	.009	.005	.037	-.103
	(.038)	(.040)	(.046)	(.056)	(.022)	(.028)	(.019)	(.052)
Average score								
Control group mean	41.4	37.4	48.1	53.3	71.3	71.0	71.8	70.8
Treatment effect	6.85	6.74	6.89	6.97	0.92	0.21	2.92	-2.53
	(2.58)	(2.47)	(3.65)	(4.22)	(2.10)	(2.86)	(1.82)	(3.01)
N	9,682	5,687	3,092	903	10,286	6,222	3,452	612

Notes:

1. Standard errors in parenthesis are clustered at the school level.
2. Observations were weighted with frequency weights in order to have similar number of students in control and treatment schools within each group of schools with close true matriculation rate.
3. The by-gender-composition estimates are taken from a single regression with three interaction variables of treatment and gender-composition dummy. The regression includes the gender-composition dummy as main effect.
4. School Fixed-Effects are included.
5. Student level controls include a set of dummy variables for the number of siblings and father and mother education, the school's lagged mean matriculation rate, a dummy for Asia-Africa ethnic background, immigration status, gender dummy, the number of credit units *attempted,* the average score in those attempted units, overall credit units *awarded,* and credit units awarded for the subject in question only.
6. All regressions include a control for math main effect

Table 10 - Differences in Mean Students Outcomes by Teacher's Gender

	18 RT Schools		All Schools	
	No Controls	With Controls	No Controls	With Controls
	(1)	(2)	(3)	(4)
	Testing rate			
Females in FO - Females in FM	.000	-.001	.006	-.010
	(.028)	(.018)	(.022)	(.017)
Females in FM - Males in FM	.009	-.015	.051	.014
	(.021)	(.017)	(.028)	(.020)
Males in MO - Males in FM	-.003	.021	-.018	.038
	(.023)	(.021)	(.023)	(.018)
constant	.918	.906	.841	.560
	(.022)	(.089)	(.026)	(.052)
	Pass rate			
Females in FO - Females in FM	.029	.023	.019	-.000
	(.049)	(.031)	(.040)	(.030)
Females in FM - Males in FM	.012	-.017	.060	.036
	(.033)	(.021)	(.031)	(.028)
Males in MO - Males in FM	.058	.144	.018	.098
	(.050)	(.042)	(.050)	(.037)
constant	.794	.896	.712	.574
	(.049)	(.132)	(.032)	(.062)
	Average score			
Females in FO - Females in FM	2.52	3.44	2.242	1.129
	(4.86)	(2.73)	(3.26)	(2.54)
Females in FM - Males in FM	0.93	-1.55	3.75	2.22
	(4.20)	(2.87)	(2.56)	(2.19)
Males in MO - Males in FM	7.60	13.52	2.25	8.67
	(4.80)	(3.52)	(4.13)	(2.64)
constant	60.0	67.0	55.0	45.3
	(5.45)	(11.36)	(2.58)	(4.82)
Observations	2,911	2,911	7,378	7,378

Notes:

Standard errors in parentheses are clustered at the school level

Regressions in columns (1) and (3) include a control for math main effect

Regressions in columns (2) and (4) include controls for one and two year lagged school mean score, attempted Bagrut credits and score, awarded Bagrut credits (total and in subject), ethnic origin, student gender and the math main effect.

Figure 1: Distribution of Winning Proportion

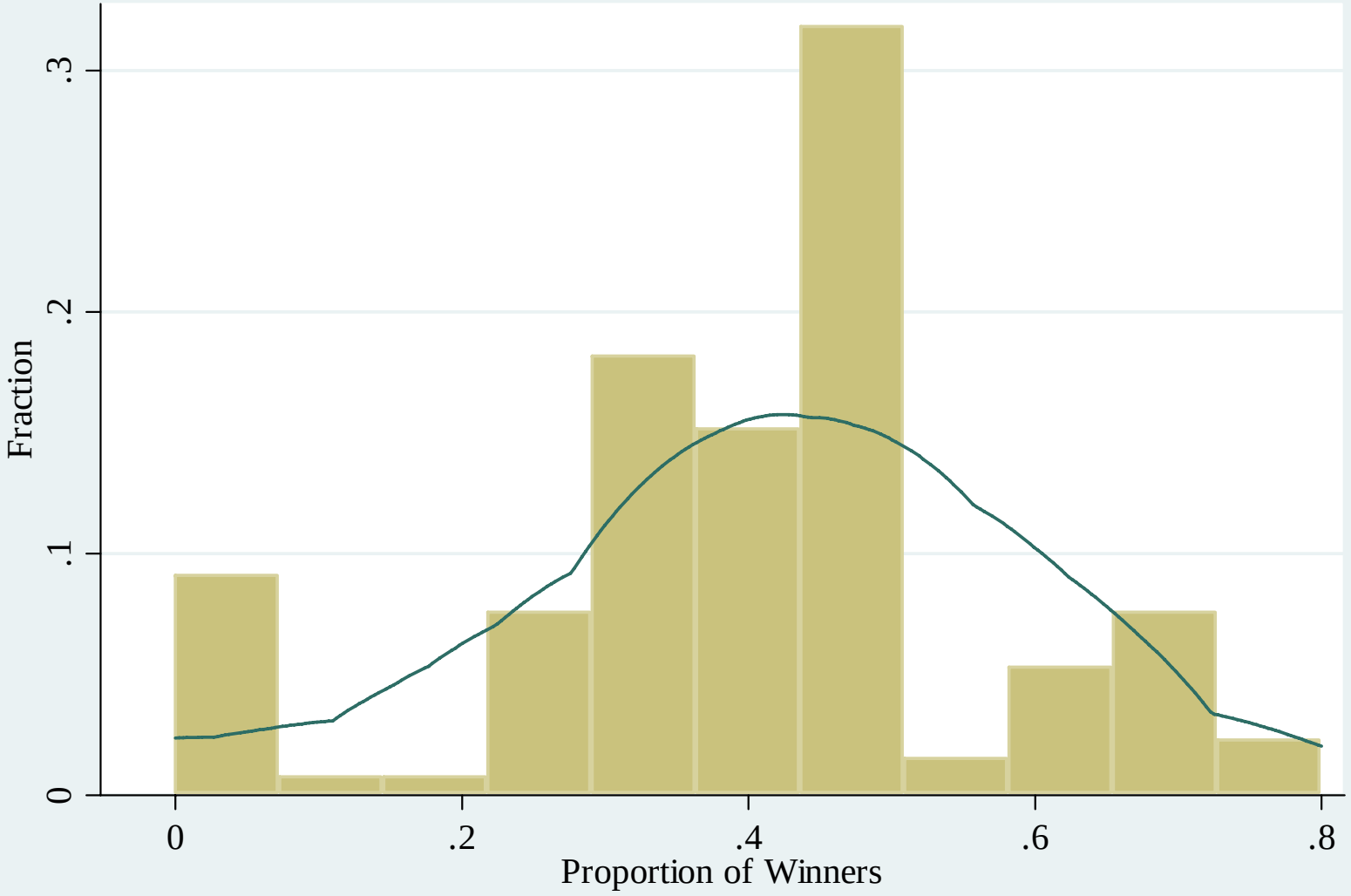

Figure 2: Distribution of Winning Proportion by Competition Group