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

Victor Lavy

Hebrew University and Royal Holloway University of London

Abstract

Recent lab and field experiments suggest that women are less effective than men in a competitive environment. In this paper I examine how individual performance in a real work place is affected by a competitive environment and by its gender mix. The competition is among math, English and Language teachers who participated in a rank order tournament that rewarded teachers with large cash bonuses based on the test performance of their classes. The evidence suggests that based on various performance criteria there were no gender differences in performance under competition: neither between teachers in competition groups with only female teachers nor groups with both genders. Regarding mechanisms that explain these results, I found no differences by either gender or by the gender mix of the competition group in teachers' awareness and knowledge of the program and in effort and teaching methods. Women however were more pessimistic about the effectiveness of teachers' performance pay, and more realistic than men about their likelihood of winning bonuses.

I. Introduction

Recent evidence from laboratory experiments suggests that men and women have different attitudes, and respond differently to competition in the workplace.[1] These studies suggest that a sizeable part of the gender earnings gap could be explained by women being less effective than men in competitive environments even if they are able to perform similarly to men in noncompetitive environments. For example, Gneezy et al. (2003) and Gneezy and Rustichini (2004) report results from a lab and field experiment, respectively, that an increase in the competitiveness of the environment significantly increased men's performance, but not women's, even though there was no performance gender gap when participants were paid according to a piece rate. This effect was stronger when men competed against women. In the lab experiment women also improved their performance but only when competing against their own gender. Doman and Falk, 2006, and Niederle and Vesterlund 2007, added related experimental results that suggest that women select the competitive environment (such as winner take all) less often than men. These findings are very important in the context of schools and teachers because of the recent expansion of performance-pay compensation schemes for teachers in the US and elsewhere[2] and because women comprise a larger fraction of the teacher labor market compared to other occupations with similar skills. For example, in 2006 21 percent of employed women in Israel were teachers versus only 5 percent among employed men. In the US the respective rates were 13 percent for women and 5 percent for men. If women are indeed less productive in a competitive environment as suggested by these studies, this may cast doubt over whether pay for performance and merit pay can improve school quality and students' academic achievements, as many of these schemes involve some form of competition between teachers.[3]

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 a competitive environment whereby monetary bonuses are paid based on a performance in a tournament. I test for gender differences in this effect and whether it varies by the gender mix of the tournament participants ('the competitive group'). However, I cannot examine the hypothesis that women shy

[1] 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 (Altonji and Rebecca, 1999, M. Blank, 1999, Blau and Kahn, 2000, Goldin and Rouse 2000, Bertrand and Kevin, 2001, Blau and Kahn, 2006). Save-Soderbergh, (2007) provide an explanation based on differences in wage bargaining strategies.

[2] The most recent pay-for-performance programs for teachers in the US include Minnesota's Q-Comp, Denver's Pro-Comp, Florida's E-Comp and STAR and Chicago's 2006 pilot program. See Lavy (2008) for more details about these programs.

[3] For example, the E-Comp Plan in Florida (2006) awards the top 10 percent of teachers in each school a five percent bonus based on student learning gains in terms of test scores in the State standardized tests.

away from competition when it is optional as competition is exogenously imposed on teachers in this case.

The competition is among math, English and Language 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 other teachers' classes in their school. 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 based on their tournament ranking. 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' and class' characteristics, and a *fixed school-level effect*. Therefore teachers were explicitly informed that they were competing against teachers of the same subject in their own school.

This experiment allows the examination of the impact of competition on performance, by gender and by gender mix of completion group, because 30 percent of participants were male and because the competitive groups in the experiment differed in their gender mix. In some subjects all the subject teachers were female or male and in others they were both genders. This between groups variation in gender composition allows comparison between female and male teacher performance in two distinct competitive environments, firstly competing with own gender and secondly with both genders. However, as female teachers were not assigned randomly to one of these two environments, I use two different identification strategies to deal with the potential endogenous and selective sorting of teachers to different types of competition groups. Firstly, in many of the schools the gender mix was different across the three competition groups, which indicated that the gender composition of the competition groups was not a school characteristic. I use this within school variation for a within school regression analysis where the identification assumption is that conditional on school fixed effects the gender composition of the different competition groups is not correlated with the potential performance of teachers in each group. The second identification strategy is based on the fact that in a sub-sample of schools, the gender composition of some of the competition groups was actually different from the gender composition of the entire teaching staff in that subject. This occurs because only teachers whose class was scheduled for a matriculation exam at the end of the year were included in the program. As a result a third of the single sex competition groups were actually a sub group of a mixed gender peer group. I will demonstrate in the paper that the way teachers are assigned to classes is unsystematic, a feature for which I find consistent support in the data. This feature naturally also

makes the non-participation of some teachers in the program an unsystematic and practically a chance event. Therefore, within the sub sample that includes all schools where there is a difference between the gender composition of the competition group and of the overall staff of that subject, and all schools with mixed gender competition groups, the incidence of having only female or only male teachers is actually random and so I view this sample as a randomized trial.

Along with the two identification approaches outlined above, 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. By using the within schools variation in the type of competition groups (by adding a school fixed effect to the analysis) I also control for any potential *unobservable* but constant heterogeneity at the school level.

I first examine the possibility of gender differences in competitiveness and the role of the gender mix of the competitive group based on teacher level outcomes related to their tournament performance such as their rank, probability of winning a bonus and bonus amount won. Secondly, I examine directly how the effectiveness of pay incentives in improving student academic achievements varied by the gender of the teacher and by the gender mix of the competition group. This analysis has the additional benefit of allowing me to compare the performance of teachers in the tournament to a randomized comparison group that did not participate in the program and to address the question whether women in competition perform better or worse than under no competition.[4]

The evidence suggests that female and male performance under competition did not differ, and female performance did not differ by the gender mix of the competition group. This analysis also shows that female teachers improved their performance in the competitive environment relative to the non-competitive one. I also examine a few mechanisms that can explain these results. For example, I found no differences by either gender or by the gender mix of the competition group in teachers' awareness of the program and in effort and in teaching methods. However, large gender differences emerge in the expectations about success in the competition and in perceptions about the effectiveness of the incentive scheme. However, these relatively large male-female differences did not vary by the gender mix of the competition environment.

These results are different from the evidence obtained from the lab experiments but they are consistent with the findings of Antonovics et. al., (2003) who used the television game 'The Weakest

[4] The presentation order of these two sets of evidence (teachers' performance based on tournaments' outcomes versus teachers' performance based on improvement in students' achievements) is immaterial and could therefore be reversed.

Link' and found that the performance of female participants was not affected by their opponents' gender. In a later study, Antonovics et al (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.

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 and V presents results concerning differences by gender in effort, tournament success and productivity. Section 6 concludes.

II. A Summary of the Related Literature

The first studies in this literature focused on the effect of competition on the relative performance of males and females. Gneezy, Niederle and Rustichini (2003) show that women are less effective than men in competitive environments (tournaments), even if they are able to perform similarly in noncompetitive environments (piece rate schemes). This effect is stronger when women have to compete against men than in single-sex competitive environments and the authors interpret this finding to suggest that women may be able to perform in competitive environments per se. They also argue that the behavior of men and women in a competitive environment may differ because of differences in skill, talent, and beliefs. Gneezy and Rustichini (2004), using boys and girls as subjects in a field experiment, concluded that competition improves performance relative to a non-competitive environment for boys, but not for girls whom, unlike the results from the lab experiment, their performance did not improve relative to the non-competitive environment even when two girls were competing against each other. Paserman (2007) provides evidence based on naturally occurring data of tennis players in Grand Slam tournaments. His main dependent variable is the percentage of unforced errors and the measure of the competitive environment is the importance of the point in determining the result of a game, set and match. This study's results suggest that men's performance in terms of unforced errors does not depend on importance of the point while women's performance is affected negatively. No mixed gender pairings are discussed in this study.

Several subsequent experiments examined potential gender differences in preference for competition. Niederle and Vesterlund (2007) investigate whether men and women, who perform equally well in mixed competitive environments, select into competitive environments. They observe that men typically select into competitive environments, while even high ability women select into noncompetitive environments. Furthermore, the choices of women are not driven by actual ability. While a significant part of this difference is driven by differences in confidence about one's ability, some additional gender effect remains. Dohmen and Falk (2006) find that even after controlling for productivity, women are about 15% less likely to enter a variable pay scheme than men. Similar results are reported in Gupta et al (2005). More recently, Gneezy, Leonard, and List (2008) explore whether there are gender differences in selecting into competitive environments across two distinct societies: the Maasai in Tanzania and the Khasi in India. Subjects in their experiment are not exogenously allocated into a payment scheme and do not know their competitors.

Overall, this body of research suggests that women may be less able when competing against men, and that furthermore, even able women opt out of competitive incentive schemes and hard tasks in general. The gender gaps found in these experiments are often very large. For example, in Gneezy et al (2003) the gender gap in performance in mixed gender tournaments was 33%. However, a continuation of this exploration yielded somewhat different results. Antonovics et al, (2008) who used data from the television game the Weakest Link and examined outcomes in competitive environments with a special focus on whether gender composition influences performance, argue that male performance is sensitive to the gender composition of the other contestants: men perform better against women than other men. In contrast, the authors report that women's performance is not influenced by gender composition. These findings are at odds with the extant literature but the authors cannot make any unequivocal arguments why that is the case.

Manning and Saidi (2008) have attempted to provide an estimate of the portion of the gender pay gap in the UK that can be attributed to these differing attitudes to competition. They find very modest evidence for differential sorting into performance pay schemes by gender, and small effects of performance pay on hourly wages and on the gender pay gap. Unlike the laboratory studies, this study finds no significant effect of the gender mix in the job on the responsiveness to performance pay.[5] These findings are not based on an experimental design but the authors suggest that they are in line with evidence of other studies with an experimental design from the real world (e.g. Bandiera, Barankay and Rasul, 2005; Paarsch and Shearer, 2007).

II. The Pay for Performance Experiment

A. The Experiment

In December 2000, the Ministry of Education announced 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.[6] The program included schools that had a recent history of relatively poor performance in the mathematics or English matriculation exams. The competition was structured as a separate rank order tournament in each subject and teachers entered the tournament as many times as the number of classes they taught and was ranked each time on the basis of the mean performance of each of their 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.[7] Separate regressions were used to compute the predicted passing rate and mean score, and each teacher was ranked twice, once for each outcome (see Lavy, 2008, for more details). Because ranking was based on within school comparison, as a school fixed effect was included in the regression used for prediction and ranking, teachers were informed explicitly that the tournament was a competition rather than a bonus scheme whereby everyone simply worked toward a prize.

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.

The program included 629 teachers, of whom 207 competed in English, 237 in mathematics and 172 in Hebrew or Arabic. As the experiment was announced during the school year, teachers were

---

[5] Lemieux, Macleod and Parent (2006) also report relatively modest estimates of the effects of performance pay schemes on average earnings.

[6] The matriculation ('*Bagrut*') exams are a set of national high school exams. See Angrist and Lavy (2008) for more details and the similarity of the *Bagrut* program to high school exit exams in other countries.

[7] The teachers were explicitly told that they were competing against their colleagues in their school. They were not told, however, that they are competing against their male peers but this is element in the experiment is not different from the lab experiment. In both cases it is assumed that participants are implicitly aware of the gender of their competitors.

not strategically assigned to groups in the participating schools. Three hundred and two teachers won awards — of whom 94 were English teachers and 124 were math teachers.

B. The Data

The data I use in this study comes from several administrative data files and from a survey the Ministry conducted with the program participants (teachers). An administrative data set provided the following information on all participating teachers: gender, subject of teaching, their school ID, ID 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 their phone number was not available and 45 teachers refused. 482 teachers completed the survey which represents more than 82% response rate among teachers who could be contacted.

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 or 5) of each class the teacher taught, details of their awareness and opinion regarding the program, teachers' subjective assessment about the likelihood that they 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 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'). Even though teachers were not assigned randomly to 'types' of competition groups in terms of gender composition, a natural variation in the proportion of female teachers among the math, English and Hebrew competition groups led in some schools to groups with only female math, English or language teachers and in other schools to groups with mixed genders.

Table 1's top panel presents the distribution of school, competition groups and male and female teachers by type of competition group. Forty nine schools participated in the incentive tournament and therefore there are 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 are 132 school-subject level competition groups of which 45 include only female teachers (223 teachers), 70 mixed-gender groups that include 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 average group size is 5 in FO, 6.3 in FM and 3.4 in MO (see Figure 1).

The empirical strategy is therefore based on comparing the performance of tournament participants by the three types of competition groups (treatment). Of course a potential problem is that the non-random assignment of teachers to FO, MO or FM may have resulted in groups that are different in aspects that may confound the effect of treatment. I use two different identification strategies to deal with potential selection in the incidence of the types of competition groups in a school. The first approach is based on a large within school variation in the gender composition of competition groups. In 34 schools there are more than one group type and only in 15 of the schools all the three groups are 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 level observed and unobserved fixed characteristics. I therefore use this within school variation in the type of competition groups to

estimate models with school fixed effects. The identifying assumption is that within school, variation in competition type is uncorrelated with potential teacher or student performance even though such selection could take place at the school level but in such a case the school fixed effect eliminates it. Evidence that supports the view that even at the school level it is very unlikely that the gender composition of the completion groups is selectively determined is the fact that none of the forty nine schools have only male or female teachers among their overall teaching staff. For example, the proportion of female teachers in schools that have a single sex competition group in at least one subject is 0.59 while in other schools this proportion is almost the same, 0.61. Therefore the incidence of having only male or only female teachers in a given subject in a school is not a particular feature of the school or its teaching staff.

The second identification strategy is based on a type of natural experiment that results from the fact that in many schools some of the English, math and language teachers did not participate in the competition because they did not teach a class scheduled for a matriculation exam in June 2001, during the 2001 academic year. This 'chance' variation in timing in some cases caused 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 is the case for 30 percent of the single sex competition groups. This proportion is 26 percent for the female only competition group (in 12 of the 45 female only competition groups the roster of all teachers in that subject is actually of mixed gender) and 55 percent for the male only competition group. The hiring and assignment procedures or practice of teachers to subjects, grades and classes, and in particular to classes that will take a matriculation exam that year, support my claim that the event of being the teacher of a class scheduled to take a matriculation exam at end of the academic year 2001 is indeed a random event. Matriculation exams are taken mostly at end of 12^{th} grade but in some study tracks (basic, intermediate and advanced level) in some subjects the study program includes two exams, the first taken at the end of 11^{th} grade. High school teachers in Israel specialize in one subject (this is particularly the case in math and English teaching), and there are no different formal hiring standards for male and female teachers. A teacher is typically assigned to a class for its complete matriculation study program, most commonly for 11^{th} and 12^{th} grade, and sometime even in 10^{th} grade. Thus it so happened that in 2001 some of the math, English or language teachers taught a 10^{th} or 11^{th} grade class scheduled for an exam a year later and therefore these teachers could not be included in the program. In such cases the teachers that were included in the program can be viewed as a random sample from the whole staff teaching that subject in the school. Therefore the participants and non participants teachers are identical in all characteristics

as indeed we see based on observables such as age, years of teaching experience, gender and schooling.

Based on this natural experiment I can define a sub sample of groups where the overall teaching staff of a given subject is of mixed 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 therefore treatment assignment represents a randomized trial. The limitation of this strategy is, however, that the randomized trial sample includes only 12 FO groups and only 9 MO groups.

The rich data available allow checking the similarity of the three competition groups in various samples and in many dimensions; including teacher 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).

More evidence supporting the view that the gender composition of the completion group is not a result of some selection process at the school level is the fact that none of the forty nine schools had only male or female teachers among their overall teaching staff. For example, the proportion of female teachers in schools that have a single sex competition group in at least one subject is 0.59 while in other schools this proportion is almost the same, 0.61. Therefore the incidence of having only male or only female teachers in a given subject in a school is not a particular feature of the school or teaching staff.

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. Table 2 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 first present 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 fathers' years of schooling is 10.7 and the

respective maternal 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. These differences simply reflect the higher probability of men to serve in the military and to be drafted for a longer time (three years) than women (two years) which delay in comparison to women their entry to post secondary schooling and into the labor force.

Fewer women have a PhD degree but this gap is compensated by the higher proportion of women with an MA degree. It is important to note however 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 a large estimated 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 between these two groups. This perfect within school balancing between these two groups of female teachers is central in this paper because one of the main hypotheses 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 a higher proportion of B.A and PhD degrees. A more significant difference is the significantly higher fathers' years of schooling among men in FM although this difference is not observed for mothers' years of schooling.

Table 2A replicates the balancing tests and analysis for the quasi randomized trial sample. The results are very similar to those presented in Table 2 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: the overall ranking of a teacher in the tournament, a 0/1 indicator of winning a bonus and the bonus amount. 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.[8]

Table 3 and Figures 2-3 present 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 have a similar success rate in the tournament. The mean ranking of men is 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. This can also be seen in Figures 4 and 5 that present the Kernel density of the pass rate and the mean score residual on which the respective ranking is based. Among men 42.5 percent win an award while for women the respective rate is 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 is $2,790 and among women it is $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 the FO group, the proportion of 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 significantly 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 react differently to competition and to the gender mix of the competition group. For this purpose I estimate the following model:

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

[8] The results are also identical when the raw residuals are used instead of the percentile ranking.

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

Table 5 presents parameter estimates of regressions where the dependent variables are the three principal performance measures, teacher ranking, whether a bonus is 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 also includes dummy indicators for math and English tournaments and in column 3 the specification includes the two treatment indicators, FO and MO, and the excluded group is of mixed gender (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 positive for all three measures but they are very small and have large estimated standard errors, therefore they can be viewed as practically zero.

In column 4 I present estimates from a specification that also includes teacher characteristics (including all the background variables presented in Table 2) as controls. 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 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 heterogeneity unobserved by the researcher but observed by the school headmaster. When these other determinants 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 although it is still significant or marginally significant. The implication of this result is that financial incentives are more effective among teachers who were promoted 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 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 teachers' 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 teachers' schooling are still negative and not 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.[9]

I also estimate a specification where the treatment effects are allowed to have different effects in each of the three subjects. The basic results are unchanged. Similarly, adding the number of competitors in each group as a control in the regressions left the results unchanged.

Column 5 reports estimates from a regression that also includes 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. The estimates are also relatively small, in particular those of the mean rank and of the winning a bonus outcome, when compared to their respected mean or standard deviation, suggesting that we cannot reject the null of zero effects. The estimate of FO on the mean rank is 1.8, only 3.6 percent of the mean of the dependent variable and it is even a lower proportion of the mean minimum rank needed for receiving a bonus; the estimated effect on the probability of winning a bonus is even smaller, 0.007 relative to mean of .425. Therefore I view these size effects as small enough to allow ruling out economically important effects on teacher performance. The respective coefficient estimates for the bonus size are larger relative to their mean and even though they are also not statistically different from zero we can less convincingly claim that they are actually zero.

Columns 6-8 report estimates from regressions of a specification similar to those of column 3-5, respectively, but the FO and MO indicators are replaced 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 specifications for each of the three outcomes.

[9] See for example Rivkin, Hanushek and Kain (2005).

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, nor in the estimates in column 8 of this table, and not in the respective estimates presented in the next section about the effect of the program on students' achievements, I tend to discount it and view it as spurious.

I also use the indictor of whether a teacher qualifies for a bonus, namely if they have a positive residual in both ranking criterions. The results for this outcome are very similar to the evidence presented in tables 5 and 5A. The bottom line from these results, which are not shown here, is that the gender mix of the competition group does not influence the probability of having a positive residual in any one or in both of the two ranking criteria.

B. Gender Differences in Program Response by Gender and by mix of Competition Teams

A post program survey of teachers added information about their awareness to the program, their opinion regarding its efficacy in improving students' achievements, teaching methods and additional effort.

First, 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. This result is presented in the first column in panel A in Table 7. Secondly, respondent answers regarding the teaching methods questions also do not reveal any differences between male and female teachers (the first column of panels B-E). 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 lessons. None of these minor gender differences were significant. However, almost all teachers (98.8% of male and 93.8% of female) reported that they adapted 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 add additional instruction beyond the regular school hours?" and teachers chose 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' F-H of Table 7.

In contrast to the above similarities by gender, I find large gender differences in the teachers' opinion regarding the program and 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 first column of panel I). The implied 15 percentage point difference is significantly different from zero.[10] Furthermore, 3 out of 4 men (76.3%) thought that they would win a bonus, but only 3 out of 5 female teachers (61.0%) had such trust in their winning ability (panel J). This 20 percent gap in 'self-confidence' (a difference of 15.3 percentage points relative to 76.3%) is significantly different from zero. Two important points should be noted with respect to this evidence. Firstly, that women's assessment of their prospects for success in the tournament is more accurate and closer to the actual mean win rate (43 percent for women and men) and so the gap between men and women in self assessment of winning reflects the over confidence of men. Secondly, despite the fact that women have more accurate expectations of their ability, I find 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 where the task was 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 report estimates of the effect of the type of 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 differential 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 a bonus is much higher in only male groups than in mixed gender groups.

V. Does Program Effectiveness Vary by Teacher's Gender and by Type of Competition Group?

In this section I test whether the impact of the program on students' achievements was different by teacher's gender and by the gender composition of the competition group. In an earlier paper (Lavy 2008) I evaluated the effect of the program on the two outcome criteria that were used to measure teacher's performance, the average score and the exam pass rate of students. The program design enabled the implementation of a 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.[11] The results indicated that incentives increased student achievements 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. 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 that I used in Lavy (2008). It should be emphasized that this randomized trial sample is a sub-sample of the sample I used in the analysis in section IV above and that it also includes a control group of school which will allow me to identify the treatment effect on students by MF, FO and MO competition teams and also to identify the effect of competition on the performance of female and of male teachers.[12]

I first allow for a heterogeneous treatment effect by the indicators FO and MO to examine if the program effectiveness differs in FO or MO groups in comparison to FM groups. The number of competition groups of the three types is smaller in this sample than in the full sample of schools which I use for the analysis in section IV. Furthermore, unlike the first part of the paper where estimation is based on samples of teachers, here the unit of observation is the student.

A. Balancing Tests by the Gender Composition of Competition Teams

I first check 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

[10] This doubt about the effectiveness of pay for performance may be viewed as consistent with the findings of Niederle and Vesterlund (2007) that women prefer fixed to piece rate compensation.
[11] See Lavy (2008) for more details.
[12] The basic results presented in section IV are unchanged when the language groups are not included but their inclusion improves the power of the estimates because it increases the total number of groups. Another difference between the two samples is that the randomized trial sample does not include the language teachers and their students, as explained in Lavy (2008).

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 this suggests the use of 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. As in Lavy (2008), the estimates are based on panel data that pool the 2000 and 2001 cohorts of students for difference in differences estimation. Columns 1-4 present 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.[13]

Based on the estimates in columns 2-4 and 6-8 I 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%. 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 test scores 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 competition. However, I should note that the mean of the control group's outcomes are marginally lower in the mixed gender group, especially

in the 1st and 2nd quartiles which implies that the effect size is 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 questions are whether female teachers in FO groups are on average less effective than female teachers in FM groups, and correspondingly male teachers in MO groups.

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

The results presented above suggest that the effect of providing incentives to teachers is 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 are equally effective. To answer this question it is sufficient to focus only on the sample of treated schools as there is no need to compare it to evidence from a control group. This approach allows performing the analysis not only for the randomized trial sample of 18 treated schools but also for all 49 treated schools. I estimate the following model:

$$Y_{ijs} = \alpha_s + \beta X_{ijs} + \gamma F_i + \delta FO_{is} + \theta MO_{is} + \phi E_{ijs} + \lambda M_{ijs} + \varepsilon_{ijs} \quad (2)$$

Where Y is student's outcome and all other notations are identical to that of equation 1.

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, as shown above, 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 treated schools and columns 3-4 present results based on all 49 treated 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 pass rate (upper panel) and the average score (lower panel).

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

[13] These results are very similar to those reported in Lavy (2008) where the estimation was done separately for math and English while here I pool the two samples.

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 pass rate for female teachers in FO is higher by 0.023 (s.e. 0.031) than in FM; the average test score for female teachers in FO is higher by 3.44 (s.e. 2.73) than in FM. However, even though the latter estimate is not significantly different from zero we should be cautious in its interpretation because it is not precisely measured and its 95% confidence interval includes high positive estimated effects (up to almost 9 as well as negative estimates as low as -2.0).

The result that female teachers perform equally well in the two types of competition groups under competition, in conjunction with the result presented in Table 9, that FO and MF groups perform better under competition than under no competition, lead necessarily to the conclusion that female teachers perform better under competition than under no competition.

Another interesting result is that there are no differences between male and female teachers in FM competition groups. The controlled estimates are all positive, indicating the higher effect 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 lower than that of male teachers in these groups by 1.55 (s.e. 2.87) but it is not significantly different from zero.

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

A similar analysis comparing male teachers in MO and FM groups shows 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 address empirically the question of whether men and women teachers improve their performance in competitive setting in the form of a tournament, and whether there are gender differentials in performance when women compete against men. A rank order tournament sets 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. As far as I am aware, this study is the first to test the hypothesis of gender differences in competitiveness based on evidence from a regular work place. I find no overall differences in performance of female and male teachers and no such differences by the competitive environment in terms of gender composition.[14]

These results are different in some respects from the evidence of the lab experiments that address the same question. Several possible explanations can account for this difference. In the real world most tasks are not completed instantaneously and workers have time to plan, receive feedback, observe rivals and adjust strategy and actions. Also, the competition studied in this paper is based on the regular activity of the participants for which they are educated and trained for and therefore female teachers may have more self confidence and are less intimidated in competing against male rivals. Of course, there is always the possibility that the results differ because they are based on heterogeneous populations. Another possible source of difference is that the lab experiments are clearly set up as a zero-sum game while the teachers' tournament experiment allowed the proportion of winners to vary across groups. Even though teachers were fully aware that they are competing against their colleagues, if they helped each other anyway, for example by helping other teachers' students or helping other teachers become more effective at their instructions etc., then the tournament is a different experiment to what the lab imposes. Finally, it can be argued that since women so heavily dominate the teaching profession, this is one of the few workplace environments in which we may expect to not find differences in competition by gender. This argument, however, cannot explain the vigorous response of female teachers to the incentive scheme without any difference from the response of male teachers.[15]

[14] Recent interesting and related evidence suggests that young women respond more effectively than young men to achievement awards. Angrist and Lavy (2008), show such results for senior high school students and Angrist and Oreopoulos (2007), for first year college students.

[15] Differences between lab and field experiments are discussed in Harrison and List (2004), Levitt and List (2007a) and (2007b). Lazear et al, (2004) provide an argument based on sorting in an experimental setting that might lead to differences with real world results.

The evidence presented in this paper provide additional support for the potential benefit from introducing incentives and competition in schools and for further experimenting with these ideas in educational systems. However, it is important to note that the findings reported here reflect short term adjustments by teachers but it is possible that a performance pay scheme may in the longer term have an effect through sorting on teaching staff composition. Women may be more likely to leave the teaching profession under such a pay compensation system.

A question can be raised about the external validity of the findings presented above to other occupations as teachers, especially male, might be different from others in the labor force. For example, it can be claimed that as men employed in the teaching profession are highly selective, they may be less likely to engage in competition and be less threatening to women than men in other contexts. It can also be argued in this regard that it is easier for teachers to collude and behave strategically as the group of participants is relatively small and therefore caution is called for in extrapolating the lessons from this study for other occupations. However, as noted earlier, the teaching profession is an important and large segment of the labor force, especially for women, and therefore the evidence is important even if men employed in the teaching profession are not perfectly representative of men in other types of employment.

References

Altonji, Joseph, G. and Rebecca, M. Blank, 1999, "Race and Gender in the Labor Market", in (Ashenfelter, Orley and David Card, eds) Handbook of Labor Economics. Volume 3C. Handbooks In Economics, vol. 5. Amsterdam: North-Holland, pages 3143-3259.

Angrist J. and V. Lavy, (2009) "The Effect of High-Stakes High School Achievement Awards: Evidence from a Group-Randomized Trial" *American Economic Review,* Vol. 99, No. 4, September: 1384-1414.

Angrist J. D. Lang and P. Oreopoulos (2009) "Incentives and Services for College Achievement: Evidence from a Randomized Trial" *American Economic Journal: Applied Economics*.

Antonovics, Kate, Peter Arcidiacono and Randy Walsh (2005) "Games and Discrimination: Lessons from the Weakest Link", *Journal of Human Resources*, 40, 918-947.

Antonovics, Kate, Peter Arcidiacono and Randy Walsh (2009) "The Effects of Gender Interactions in the Lab and the Field", *Review of Economics and Statistics*.

Bertrand, Marianne and Hallock, Kevin F. "The Gender Gap in Top Corporate Jobs." *Industrial and Labor Relations Review,* 55(1), October 2001, pp. 3-21.

Bandiera, Oriana, Iwan Barankay and Imran Rasul (2005), "Social Preferences and the Response to Incentives: Evidence from Personnel Data", *Quarterly Journal of Economics*, 120, 917-962.

Blau, Francine, and Lawrence M. Kahn, "Gender Differences in Pay," *Journal of Economic Perspectives*, XIV (2000), 75–99.

Blau, Francine D. and Kahn, Lawrence M., (2006) "The U.S. Gender Pay Gap in the 1990s: Slowing Convergence", *Industrial and Labor Relations Review*, 60, 45-66.

Datta Gupta, Nabanita, Anders Poulsen and Marie-Claire Villeval (2005) “Male and Female Competitive Behavior: Experimental Evidence”, IZA DP 1833.

Dohmen, Thomas and Armin Falk, 2006, “Performance Pay and Multi-Dimensional Sorting: Productivity, Preferences and Gender”, IZA DP No. 2001.

Gneezy, U, 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, U, Leonard, K.L. and List, J.A. (2009) “Gender Differences in Competition: Evidence from a Matrilineal and a Patriarchal Society.” *Econometrica.*

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

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

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

Lavy, V. “Using Performance-Based Pay to Improve the Quality of Teachers”, *The Future of Children*, Spring 2007, 87-110.

Lavy, V. (2009) “Performance Pay and Teachers’ Effort, Productivity and Grading Ethics”, *American Economic Review*, Vol. 99, No. 5, December, 1979-2011.

Lazear, Edward P., Ulrike Malmendier and Roberto Weber (2004) “Sorting in Experiments,” unpublished, Stanford University.

Lemiuex, Thomas, W. Bentley Macleod and Daniel Parent (2006) “Performance Pay and Wage Inequality”, University of British Columbia.

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.

Manning Alan and Farzad Saidi, (2008), “Understanding the Gender Pay Gap: What’s Competition Got to Do with It?” Centre for Economic Performance, London School of Economics.

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.

Paarsch, Harry J. and Bruce Shearer, (2007) ``Do Women React Differently to Incentives? Evidence from Experimental Data and Payroll Records," *European Economic Review*, 51, 1682-1707.

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

Polachek, SolomonW., “Occupational Self-Selection: A Human Capital Approach to Sex Differences in Occupational Structure,” *Review of Economics and Statistics*, LXIII (1981), 60– 69.

Rivkin, Steven G., Eric A. Hanushek, and John F. Kain, "Teachers, Schools, and Academic Achievement", *Econometrica*, Vol. 73, No. 2 (March, 2005), 417–458

Save-Soderbergh, Jenny (2007) “Are Women Asking for Low Wages? Gender Differences in Wage Bargaining Strategies and Ensuing Bargaining Success”, Swedish Institute for Social Research.

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 - 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 2A - Balancing Tests of Teachers' Characteristics by Type of Competition Groups in Quasi-Randomized Trial Sample

	Mean of Male	Difference (Female - Male)	Mean of Female in FO	Difference (Female in FO - Female in FM)	Difference (Female in FM - Male in FM)	Mean of Male in MO	Difference (Male in FM - Male in MO)
	(1)	(2)	(3)	(4)	(5)	(6)	(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 3 - 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 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. Worked with Students in Small Groups				C. Worked with Students Individually			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Constant	.901	.925	.989	.972	.686	.605	.844	.624	.593	.564	.819	.721
	(.024)	(.035)	(.131)	(.150)	(.044)	(.079)	(.229)	(.298)	(.039)	(.070)	(.222)	(.276)
Female	.016	.004	-.007	-.020	-.058	-.060	-.012	-.010	.042	-.004	-.004	-.007
	(.028)	(.032)	(.030)	(.031)	(.049)	(.063)	(.073)	(.077)	(.050)	(.059)	(.074)	(.076)
FO Group		.016	.025	.037		-.012	-.022	-0.04		.031	.022	-.062
		(.040)	(.041)	(.048)		(.054)	(.055)	(.080)		(.082)	(.089)	(.114)
MO Group		-.029	-.032	-.075		.010	-.070	-.112		-.083	-.121	-.328
		(.055)	(.056)	(.066)		(.094)	(.102)	(.136)		(.079)	(.093)	(.119)
Subject Main Effects		√	√	√		√	√	√		√	√	√
Individual Controls			√	√			√	√			√	√
School Fixed Effects				√				√				√
Observations	608	608	576	576	608	608	576	576	608	608	576	576

	D. Divided Students in Class by Level				E. Adjusted Teaching Methods to Student's Level			
	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)
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.

	F. Added After School Instruction				G. Number of Hours Added				H. 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

	I. Teacher Thinks Program Will Improve Student's Achievements				J. Teacher Thinks He will Win an Award				K. 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 8 - Balancing Tests of Students' and School Characteristics by Competition Group Types in Treated Schools

	Mean of FO	Difference (FO - FM)	Mean of MO	Difference (MO - FM)
	(1)	(2)	(3)	(4)
A. School Characteristics				
Religious school	.283	.044	.486	.258
		(.147)		(.292)
Arab school	.036	-.159	.514	.411
		(.124)		(.257)
Lagged "Bagrut" rate	.460	.001	.436	-.026
		(.034)		(.044)
Two-years Lagged "Bagrut" rate	.514	.023	.497	-.001
		(.023)		(.031)
B. Student Background				
Father education	11.0	.914	8.01	-2.65
		(.630)		(.855)
Mother education	11.2	1.13	7.56	-3.14
		(.713)		(.678)
Number of siblings	2.70	-.283	4.15	1.39
		(.405)		(.626)
Gender (male=1)	.490	-.047	.624	.113
		(.058)		(.078)
Immigrant	.012	-.004	.024	.011
		(.008)		(.022)
Asia-Africa ethnicity	.259	.048	.159	-.075
		(.039)		(.085)
C. Student Lagged Achievements				
Math credits gained	.381	.158	.195	-.088
		(.173)		(.140)
English credits gained	.066	-.044	.073	-.025
		(.021)		(.071)
Total credits attempted	5.86	1.23	3.93	-1.22
		(.652)		(.624)
Total credits gained	4.66	.935	2.91	-1.24
		(.537)		(.336)
Average score	63.3	2.19	51.1	-11.9
		(2.32)		(1.92)
Observations (FO, FM, MO, Total)	1,702	3,031	508	5,241
Groups (FO, FM, MO, Total)	13	17	6	36

Notes:

1. Standard errors in parenthesis are adjusted for school level clustering.
2. The schools status of nationality and religiosity does not change. Any change in the means across years reflects relative changes in the number of students in a cohort.
3. MO = Male teachers only FO = Female teachers only FM = Both female and male teachers
4. The sample consists of 2001 students in the 18 Randomized Treatment schools

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)
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
3. The by-gender-composition estimates are taken from a single regression with three interaction variables of treatment and gender-composition dummy.
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)
	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 Group Size by Competition Groups

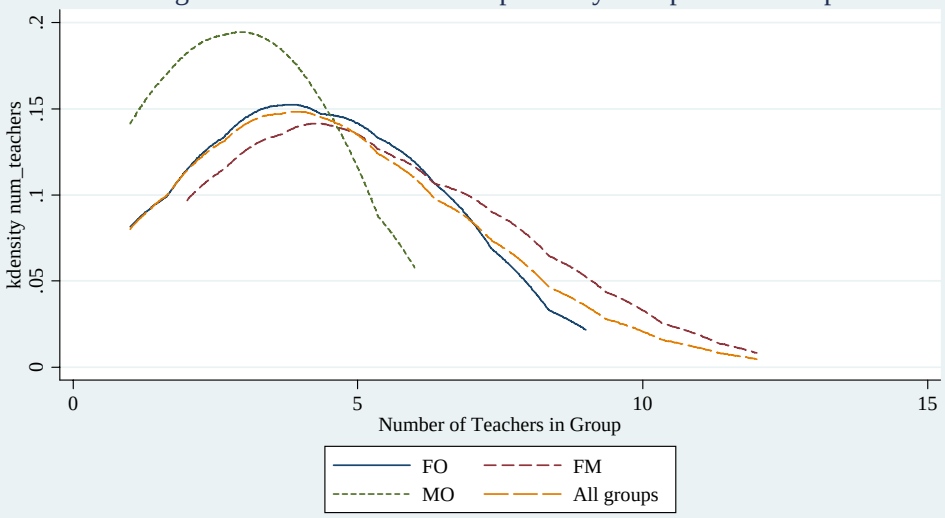

Figure 2: Distribution of Winning Proportion

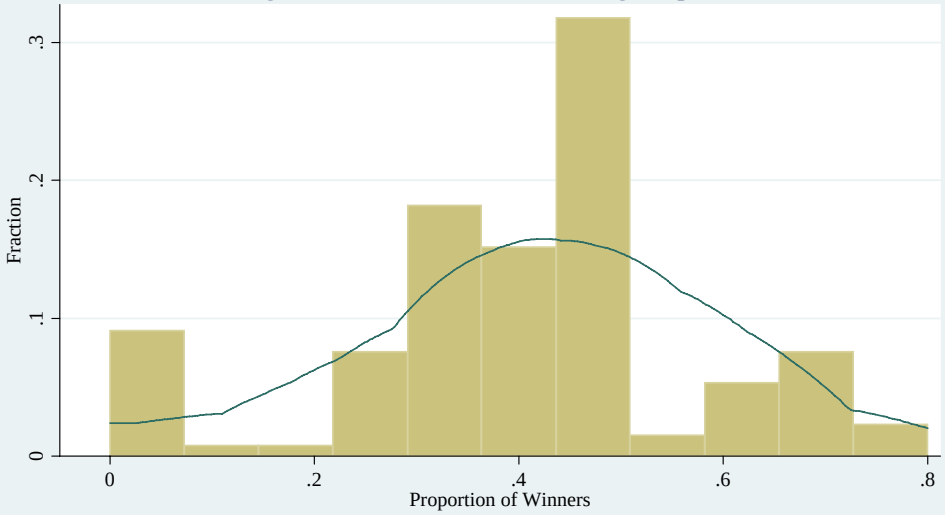

Figure 3: Distribution of Winning Proportion by Competition Groups

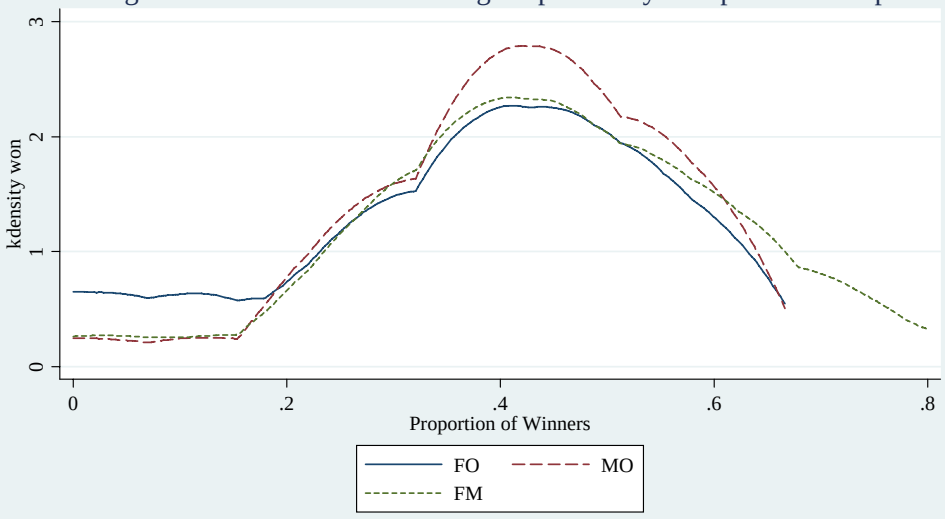

Figure 4.A: Kernel Density of Mean Score Residua

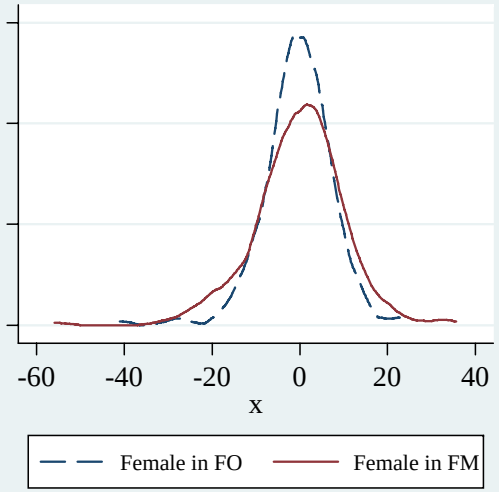

Figure 4.B: Kernel Density of Mean Score Residua

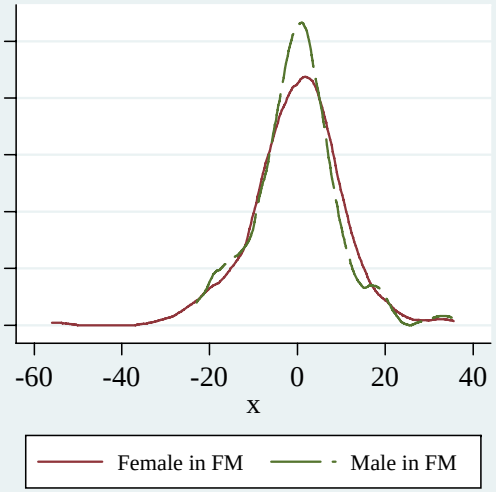

Figure 4.C: Kernel Density of Mean Score Residual

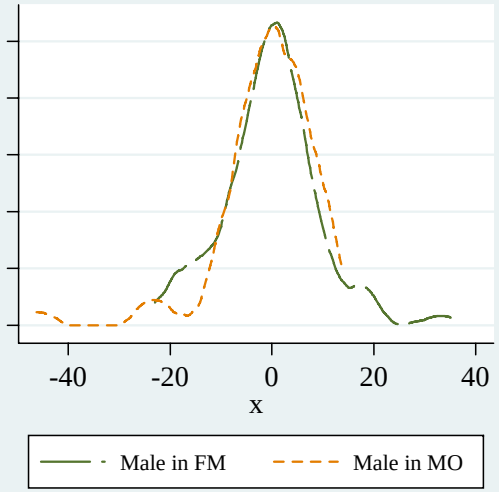

Figure 5.A: Kernel Density of Mean Pass Residual

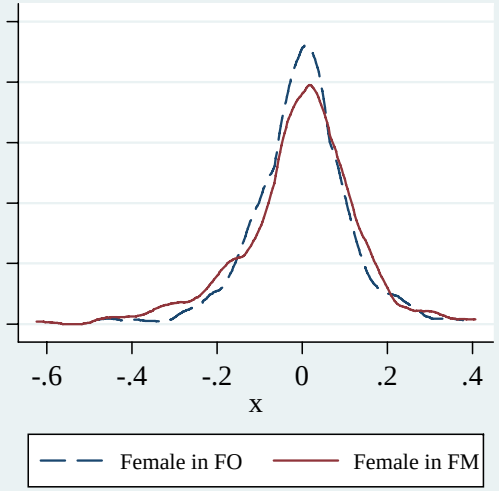

Figure 5.B: Kernel Density of Mean Pass Residual

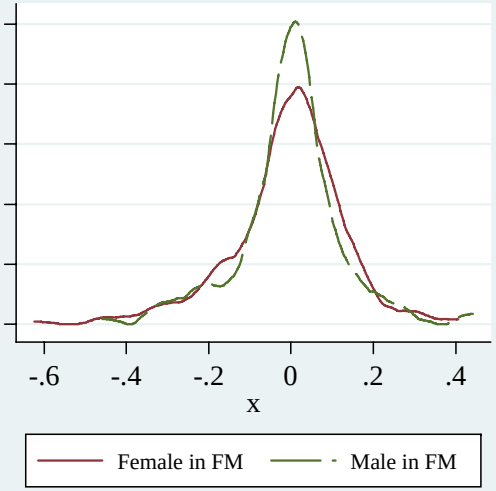

Figure 5.C: Kernel Density of Mean Pass Residual