Competitive Pressure Widens the Gender Gap in Performance: Evidence from a Two-Stage Competition in Mathematics[*]

Nagore Iriberri[+] and Pedro Rey-Biel[**]

January 12th, 2018

Abstract

In two-stage elimination math contests participants from four different age groups compete to pass from stage 1 to stage 2 and later to be among the winners. Although female participants have higher Math grades at school the gender gap reverses in the two stages of the contests. More importantly, following the same individual participant across different stages, we find that the gender gap in performance increases from stage 1 to stage 2 of the competition. The increase in female underperformance is attributed to higher competitive pressure and alternative explanations based on selection, discrimination and differences in reaction to increasing difficulty are ruled out.

Keywords: Gender gap, glass-ceiling effect, education, competition, mathematics, field data.

JEL classification: C72; J16; J31.

[*] We thank Sara de la Rica, Bernardo García-Pola, Uri Gneezy, Iñigo Hernández-Arenaz, an editor and three anonymous referees and seminar participants at various institutions and conferences for helpful comments. We are particularly thankful to the organizers of *Concurso de Primavera de Matemáticas*, who collaborated throughout the project. We also thank Adrià Bronchal, Gerard Domènech and Lluis Puig for excellent research assistance. Nagore Iriberri acknowledges financial support from Departamento de Educación, Política Lingüística y Cultura del Gobierno Vasco (IT869-13), Ministerio de Economía y Competitividad (ECO2015-66027-P), and Instituto de la Mujer (2001-0004-INV-00037). Pedro Rey-Biel acknowledges financial support from Programa Ramón y Cajal, Ministerio de Economía y Competitividad (ECO2015-63679-P and the Severo Ochoa Programme for Centres of Excellence in R&D (SEV-2015-0563)), Instituto de la Mujer (2001-0004-INV-00037) and Barcelona GSE.

[+] University of the Basque Country, IKERBASQUE, Basque Foundation for Science. E-mail nagore.iriberri@gmail.com.

[**]Universitat Autònoma de Barcelona and Barcelona Graduate School of Economics. Department d´Economia i d´Historia Econòmica, 08193 Bellaterra. Barcelona (Spain). Tel: (+34) 935812113. E-mail: pedro.rey@uab.es.

1 Introduction

The gender gap in labor market outcomes has long been a major subject for study in economics. In developed countries, a gender wage gap still persists. This presents a challenge to conventional explanations based on differences in human capital, preferences or statistical discrimination (Blau and Kahn, 2000). Also, men hold a larger portion of the highest-ranked occupations even within firms, in what is frequently referred to as "the glass ceiling effect" (Bertrand, 2009; Blau, Farber, and Winkler, 2010; Bertrand and Hallock, 2001; Wolfers, 2006). Crucially, it is hard to identify the causes of these phenomena in labor settings due to the difficulty of observing key variables such as objective and comparable measures of performance and controls for ability.

Experimental studies in which several such variables can be controlled for have proposed gender differences in competitiveness as a complementary behavioral explanation for the observed gender gap in labor market outcomes. In seminal papers, Gneezy, Niederle, and Rustichini (2003) show that women underperform compared to men under competitive incentive schemes, while Niederle and Vesterlund (2007) further show that women prefer incentive schemes rewarding their individual performance than competitive incentive schemes. However, not only do labor market settings differ in the underlying incentive schemes but many relevant labor settings include features of a multi-stage elimination contest in which only the fittest survive to reach the final stages of the competition, where a few highly rewarded positions lie at the top.[1] Furthermore, the glass-ceiling effect shows that the presence of women significantly decreases as one climbs up the hierarchy.[2]

[1] Examples abound in labor markets. In the academic world, while women make up the majority of students in American colleges and universities (59% of graduate enrollment), only 42% of full-time faculty members and only 28% of full professors are women, (Curtis, 2011). Similarly, in the legal profession women make-up 48% of the enrollment in law schools but only 44% of associates in private practice and only 20% of partners are women (American Bar Association, Commission on Women in the Profession, 2014; Wood et al, 1993). Finally, looking at the five highest-paid executives in each of a large number of U.S. firms for 1992-1997, Bertrand and Hallock (2001) find that women represent only 2.5% of the sample (1,134 women out of 46,708 executives). Similar results are found in Wolfers (2006) and in Gayle, Goan and Miller (2011) for the US and in Aher and Dittmar (2012) and Matsa and Miller (2013) for Europe.

[2] At the very top, in the Associated Press list of the ten highest paid CEOs in the US in 2015 there is only one woman, ranked 5th, while the two highest-paid male CEOs make more than all the ten top-paid female CEOs combined.

Competitive pressure, i.e. the stress that one feels when competing, increases as one moves up in a multi-stage elimination contest. First, the participants' average ability increases with each stage due to the selection process inherent in multi-stage elimination contests, as the best performers move ahead and the worst are taken out. Second, the ratio of prizes to the number of participants may be lower in the final stages of a contest, making it more competitive. This paper addresses the core question of whether men and women react differently to increases in competitive pressure in a multi-stage elimination contest.

A regional two-stage math contest, *Concurso de Primavera de Matemáticas*, offers a unique opportunity to study this question. Importantly, the first stage of the competition is done at students' own schools, a familiar environment, while all students qualifying to the second stage compete on the same date and at the same public university location, an unfamiliar environment, which may contribute to increase the competitive pressure. Both stage 1 and stage 2 tests are multiple choice tests, and thus, grading does not depend on who and where the test is administered. Moreover, we use students' grades in Mathematics in their respective schools in the semester prior to the contest as a control for individual ability. Finally, we are able to control for school characteristics, including size, overall school quality, and school quality in mathematics. We find that although female presence is close to being balanced in stage 1, with 56% of contestants being male and 44% female, it is highly unbalanced in stage 2, where 66% are male and 34% female. Furthermore, out of the 146 contestants who are recognized as best performers at the end of stage 2 only 19 (13%) are female. These numbers clearly show that female representation decreases as we move up through the stages of the contest, evidencing a clear glass-ceiling effect.

Our result is partly explained by gender differences increasing among the upper part of the performance distribution, as only the best performing students move up to the second stage (Ellison and Swanson, 2010). More importantly, given that we can compare performance levels *for the same students* in both the first and second stages of the contest, that is, in the balanced sample, we further show that there is *another* cause for our main result. In particular, we show that for the set of participants whose performance can be observed in multiple stages, female and make participants have similar Math grades at school; but once we shift to the contest, a gender gap emerges. Male participants perform better than female participants in the initial stage of the contest, where the gender gap is

4.9 test points (0.34 standard deviations of the mean). This gap widens up to 7.3 test points when they move to the second and final stage of the contest (0.09 standard deviations of the mean) which represents an increase of almost 50% on the gender gap in stage 1, and shows a gender differential reaction to increases in competitive pressure.

We use the variation in age-levels, in academic years within each age-level, and in ability, as well as variation in school characteristics, to perform a heterogeneity analysis. We find that the gender differential in reaction to increases in competitive pressure is stronger among high ability participants than low ability participants. Although some school characteristics, such as size and quality in math, are shown to be significant determinants of performance in these contests (Ellison and Swanson, 2016), we find no evidence of gender differences as competitive pressure increases varying with those characteristics. We are also able to rule out three alternative explanations for the increase in the gender gap we observe: participant selection, participant discrimination and level of difficulty of the questions. Finally, using the responses to the survey, we obtain direct evidence that female participants claim to be more affected by competitive pressure than male participants do.

Identifying gender differential reactions to changes in the underlying incentive schemes is not easily done using field data. Jurajda and Münich (2011) examine multiple university entry exams taken by the same individuals and find that men perform better than women when applying for more competitive institutions, but no such difference exists in entry exams for less competitive schools. Similarly, Örs, Palomino and Peyrache (2013), compare the performances of the same population in the French Baccalaureat, which is non-competitive, and in the highly competitive entrance exam for the Ecole des Hautes Etudes Commerciales in Paris, and find that although female students perform better in the non-competitive setting, the gender gap is reversed in the competitive exam. Buser, Niederle and Oosterbeek (2014) find an important association between a real-world measure of competition (prestige and Math intensity of different academic tracks) and a laboratory experimental measure of competitiveness using school data from the Netherlands. Azmat, Calsamiglia, and Iriberri (2016) use school performance data in a non-competitive setting to show that gender differences increase with the stakes, measured by the weight of a test in the final course grade. In a very different contest, tennis tournaments among professional players, Paserman (2010) uses a clever statistical procedure to identify how much is at stake at its particular point in a tennis match, and finds that women make more unforced mistakes than men in more important points. Brown and Minor (2014) also uses tennis tournament data to test the predictions of a theoretical model that studies past, current and future performance in a multi-stage tournament and identify how future rival's ability might affect current effort and performance. This study does not take a gender perspective.

Our contribution, on top of adding onto the evidence from the field, relies on measuring *changes* in the gender gap in performance in a two-stage elimination contest in which competitive pressure increases from the first stage to the second. Additionally, in our setting the format and the grading of the tests taken in both stages are held constant, so the differences in performance can be directly attributed to increases in competitive pressure. Interestingly, our dataset also offers variation in age, academic year, and ability, enabling us to study whether the gender gap is heterogeneous across these variables as competitive pressure increases.

Admittedly, there are other aspects in our competition that change from stage 1 to stage 2, which may or not be interpreted as changes in competitive pressure. First, the second-stage test is taken in a non-familiar environment, which might increase pressure. Second, in the second-stage test participants are competing against people they have not met before, which might increase or decrease competitive pressure. Third, the gender composition between the two stages also differs, being fewer women in the second stage. Gneezy et al. (2003) find that female underperformance in competitive environments is absent in all-female competitions. See Booth, Cardona-Sosa and Nolen (2013) where they show that female performance is affected by the gender composition of students in a non-competitive environment. Although our setting does not allow us to distinguish between these three additional factors, we do find robust field support that female underperformance increases from stage 1 to stage 2 of an eliminatory competition, comparing the same set of participants who can be observed in both stages.

The rest of the paper is organized as follows. Section 2 describes the two-stage math contest and the data. Section 3 contains the results. Section 4 considers alternative explanations for the observed findings, ruling them out. Section 5 uses the answers to the survey in the 2017 edition of the contest, which we use to interpret our results. Finally, section 6 concludes.

2 The Data

2.1 The Setting: A Two-Stage Contest in Mathematics

We use data from the 2014 edition of *Concurso de Primavera de Matemáticas*, a regional math contest involving about 40,000 students from 439 schools in the Madrid region of Spain. This contest has been organized every year since 1996 by the Mathematics Department of Universidad Complutense de Madrid.[3] As explained on their website, the contest has two main goals: to "motivate a large number of students by showing them that thinking and studying math can be fun," and, "to promote thinking outside the box and textbooks when solving problems, using logical reasoning, class geometry, parity issues, the properties of numbers, and probability."

To enable the external validity of our findings to be checked, Table A.1 in the appendix compares the school characteristics of the 439 schools that take part in the contest with the full sample of 1,578 schools in the region of Madrid. Thus, overall we have roughly 28% of schools in the region of Madrid. Specifically, we have 20% of those primary schools and 48% of secondary schools. Among the school characteristics, the schools taking part in the contest contain a lower proportion of public schools, have larger numbers of students and, as expected, are of higher quality in mathematics. In order to measure school quality over different subjects, and in particular in mathematics, we use average school performances in a test externally designed, administered and evaluated by the Department of Education in the region of Madrid. The fact that all students take the same test enables us to compare and rank different schools in the region of Madrid.[4]

The rules of the contest are clearly set out. First, there are four different contests according to age groups, which we refer to as levels 1 to 4, such that students from two consecutive school years compete within each level. Thus, level 1 includes children in their fifth and sixth academic years of primary school, so contestants are aged 10 and 11. Similarly, level 2 includes 12-13 year-olds, level 3 includes 14-15 year-olds and level 4

[3] For the organization's website see http://www.sociedadpuigadam.es/primavera/index_nuevo11.php

[4] In particular, *School_Overall_Quality* measures schools' centile in the ranking of the "Conocimientos y Destrezas Indispensables" (CDI – "Essential Knowledge & Skills") tests, which include the subjects of Math, Spanish Language and General Culture; and *School_Math_Quality* measures schools' centile in the performance of the Math "Conocimientos y Destrezas Indispensables" (CDI) test. This test is administered to all students in the 6th year of primary school (11 year-olds) and in the 3rd year of secondary school (14 year-olds). Unfortunately, this test scores are anonymized such that we cannot match these test scores at the student level but we use these test scores as a measure of school quality. For more information see: http://www.educa2.madrid.org/web/cdi/pruebas-cdi

includes 16-17 year-olds. Secondly, it is a two-stage elimination contest in which only the students who perform best in the first stage (3 to 5 per level and school) qualify for the second stage. Thirdly, in both stage 1 and stage 2 the contests consist of a test for each level made up of 25 multiple-choice questions, all of them designed by the contest organizers. The questions for each level are designed so that students in the lower school year in each level have already seen the material necessary to answer the questions correctly. Each question has 5 possible answers, only one of which is correct. The grading system awards 0 points for wrong answers, 1 point for questions not answered and 5 points for questions answered correctly, so students' can score from 0 to 125 in each test. The stage 2 test uses the same format but is designed by the contest organizers to be more difficult than stage 1.[5] Fourthly, the top prizes are awarded to the best three contestants in each level in stage 2 of the contest. Additionally, the top 5% contestants in stage 2 are awarded a diploma and a small gift in a public ceremony.[6]

We ran a survey at schools that send participants to stage 2 of the contest to gather information on how the stage 1 contest was carried out at their schools. Only 4% of schools said that they used criteria other than the performance in the stage 1 test in order to select their students to participate in stage 2. Out of those schools that only use stage 1 performance, 56% of them said that participation in stage 1 is open to all students who voluntarily want to participate, 21% said that all students at the school participated, 19% said participation was restricted to best performing students who volunteered to participate, and 4% said participation was restricted to only the best performing students. Notice that even in schools where all students participated, they were allowed to just sit in and put as low effort as they wanted to. Our main analysis, shown in Table 2, compares gender differences among students who did both the stage 1 and stage 2 tests.

The timing of the contest is as follows. In January 2014 schools signed up online to participate in the contest. Stage 1 of the contest took place on a preset day in February, when students took the relevant test at their respective schools. Teachers at the schools, who were able to download the stage 1 test only a few days before it was scheduled to place, used objective answer keys provided by the organizers to grade stage 1 tests. Those

[5] All past exams and correct answers for all stages and levels are available on the contest website.

[6] As can be checked on the website, it is not revealed ex-ante what the main prizes are. In past editions, prizes were scientific calculators or ipads, and the gifts for the top 5% in stage 2 were books. The most important reward is the prestige associated with being among the top 5% of all contestants, which is publicly announced on the website and in a public award ceremony.

teachers then selected at most 5 students from among the top performers within each level in stage 1 to participate in stage 2. The stage 2 test took place on a preset Saturday morning in mid-April on the campus of Universidad Complutense de Madrid, where the grading was done on that same day by the organizers. Finally, prizes were awarded in a public ceremony held a few days after the stage 2 competition.

[Table 1: Competitive Pressure in Stages 1 and 2]

Competitive pressure increases from stage 1 to stage 2. First, the average ability of competing peers in stage 2 is higher than in stage 1. Table 1 shows this comparison for two performance measures for each of the four different contests. On the one hand, by construction the participants in stage 2 have shown higher performance levels in stage 1 than those who do not go on to stage 2. On the other hand, the participants that do go on to stage 2 also show higher Math grades at school than those that do not. These two variables are explained in detail in the next subsection. Second, also, the proportion of winning spots is lower in stage 2, making winning harder, as shown by the number of competitors and the winning positions. These differences appear in all four levels. Moreover, there might be other factors that influence the perceived higher pressure participants feel in stage 2. For example, the second stage test took place at an unfamiliar and large location for all students, where they competed in large classrooms with students, for whom they had much less information about their average performance.

In order to empirically confirm that participants indeed felt more pressured while performing in the second stage, we administered a questionnaire to the participants of a posterior edition (2017) of this contest. We explicitly asked the participants (Question 6 of the Questionnaire, see Figure 2): “Please, say your agreement level (1 referring to *Strongly Disagree* and 5 to *Strongly Agree*) with the statement: “While doing the test I felt more pressure during Stage 2 than in Stage 1”. Note that despite the pool of respondents not being the same, given the selection process in the two editions remained constant, results are informative about the perceived pressure. In Figure 2, we can see participants’ responses, separated by gender. Most participants either *Agree* or *Strongly Agree* with this statement, confirming that participants indeed felt more pressure in the second stage. Interestingly, female participants’ answers are inclined to be more in agreement with this statement (which is discussed in Section 5).

2.2 The Sample

We created a database with three pieces of performance data in mathematics. First we collected the scores and answer sheets of all the approximately 2,800 participants in stage 2, which were provided by the contest organizers. Second, we obtained about 20,700 stage 1 scores and answer sheets, which were voluntarily provided by school teachers (out of about 40,000 participants in stage 1). Third and finally, we also collected Math grades at school in two different ways: First, students were asked to report their Math grade on the answer sheets of their stage 1 and stage 2 tests. Second, for those students who progressed to stage 2, teachers were requested to report students' Math grade at school. This gives us two complementary sources for the Math grade at school, one self-reported (for those who participate only in stage 1 or in both stages) and one reported by teachers (only for those who go on to stage 2). Ideally, we would like to combine the two sources to increase the number of observations, but there may be cause for concern about gender differences when Math grades are self-reported and this could potentially bias the gender differences observed in Math at school. Crucially, we have both types of Math grade for a subset of 2,554 participants who go on to stage 2 and also provide self-reported Math grades for both or one of the stages (91% of stage 2 participants), which means that we can compare them and test for gender differences in reporting. Table A.2 in the appendix shows that no differences between male and female students are found when comparing the self-reported Math grades with those reported by teachers. From now on, we combine these two sources of Math grades and take the average for the two whenever both are available. We call this variable simply “Math at School”. We obtained Math grades for 14,113 students.

This paper measures the gender difference as competitive pressure increases from stage 1 to stage 2 of the contest. For such a comparison, ideally, one would like to have performance in stage 1 as well as the Math grade for *all* the students who reach stage 2 (2,800 participants). Unfortunately, as both the Math grade and the performance in stage 1 were voluntarily provided by either the school teacher or the contestants, this is not the case here and we have all three math performance levels for about 1,800 participants, to which we will refer as the balanced sample. We thus need to test whether and how the subjects whose performance levels across different stages can be observed (1,800 participants) differ from the subjects whose performance levels across different stages cannot be observed (the remaining 1,000 participants out of 2,800). More importantly, as we are focusing on gender gaps we must also test whether the gender composition in the selected sample of 1,800 participants whose performance in different stages will be measured is different from the gender composition in the whole sample of 2,800. Table A.3 in the Appendix shows no change either in the composition of participants (columns 1 and 3) or in their gender (columns 2 and 4). Those participants who provide their stage 1 performance and Math grade perform slightly better than those who do not, as shown by the positive coefficient of the *Stage 1 Dummy* (shown in columns 1 and 3) and by the positive coefficient of the *Math Dummy* (shown in column 3). However, neither the *Stage 1 Dummy* nor the *Math Dummy* are significant. More importantly, regarding changes in gender composition, the coefficients for the interactions of these variables with the *Female* variable are also insignificant, as shown in columns 2 and 4, confirming that there is no change in the composition of male and female contestants. From now on, we use the balanced sample for our main analysis, although we also show the robustness of our results using other samples.

[Table 2: Descriptive Statistics in the Balanced Sample]

Table 2 presents the main outcome variables on performance that we study, overall and by gender, aggregated across all levels and separately for each of the four levels. The last column shows the *p*-values for the F-Test of equality of variable means across gender. Based on performance across all levels, contest participants can be considered as good students of Mathematics, with an average grade of 8.36 out of 10. Also, girls have slightly higher Math grades at school than boys (8.42 vs. 8.33), although the difference is not significantly different from zero. However the gender gap reverses in the two stages of the contest, with male students showing significantly higher scores (66 in stage 1 and 52 in stage 2) than female participants (61 and 44, respectively). Furthermore, the gender gap in stage 2 is wider than in stage 1. Density distribution functions of performance by gender and by stage, for the balanced sample, shown in Figure 1, depict similar patterns. It can further be observed in Table 2 that the advantage of boys over girls in the contest comes from the fact that girls failed to answer more questions and got fewer right answers, but not from the number of wrong answers, in which there are no significant differences by gender. Very similar patterns are observed when looking at each level separately. Descriptive statistics for the whole sample can be found in Table A.4 in the appendix, which shows very similar patterns.

3 Results

3.1 Main Result

In order to test whether the gender gap in performance changes as competitive pressure increases, we measure gender differences in performance from stage 1 to stage 2 of the contest for the set of students who took part in both stages, the balanced sample of 1,800 participants. We will use standardized values with mean zero and standard deviation of 1, at each stage and competition level, for all the dependent variables. Table 3 shows the estimation results using random effects at the student level.

[Table 3 here]

Columns 1 to 3 show the main regressions, where the dependent variable is the score or performance for each participant in the different stages of the contest, and the three main independent variables are a dummy for gender (*Female*), a dummy for stage 2 performance (*Stage 2*), and a term for the interaction between these two variables (*Female*Stage 2).* Column 1 includes no controls or fixed effects. Column 2 includes Math grades at school as a control, as well as school characteristics (*Mixed* and *Private*, with *Public* being omitted; location dummies, with *Madrid* being omitted; *Size*, *School_Overall_Quality* and *School_Math_Quality*). Finally, column 3 replaces school characteristics with school fixed effects. When we compare estimation results in columns 1 to 3, the estimated coefficients change very little and the standard errors increase, as expected. From now on we use the most restrictive specification with school fixed effects (the one in column 3).

Female participants underperform compared to male participants, as shown by the fact that the *Female* coefficient is negative and significant, with a gender gap of 0.34 standard deviations of the mean. More importantly, girls underperform more (the gap is more negative) in the second stage than in the first in 0.09 standard deviations of the mean, as shown by the coefficient of the interaction term between *Female* and *Stage 2*. Therefore, given that we have two observations – performance in stage 1 and in stage 2 – for each contestant, the interpretation of the interaction coefficient is that girls not only perform worse than boys in each stage of the contest but that they underperform *even more* when competitive pressure increases, i.e. in stage 2, showing that there is a gender differential reaction to increasing competitive pressure. Math grades at school are shown to affect performance in the contest positively, as would be expected.[7] Estimation results in column 2 also show that school size and school math quality are positively correlated with performance in the math contest, as reported in Ellison and Swanson (2016).

The rest of the columns in Table 3 show similar estimation results for different dependent variables, such as the number of omitted answers (column 4), answered correctly, referred to as right, (column 5), and answered wrongly, referred to as wrong (column 6). The increase in female underperformance is explained by an increase in the number of omitted answers and a decrease in the number of right answers. Interestingly, there is no change in the number of wrong answers. When we look at the proportion of right answers, defined as the number of questions answered correctly out of the questions actually answered (columns 7 and 8), we again see that female participants show a decrease in the proportion of right answers from stage 1 to stage 2. Note that in column 8, we additionally control for the number of omitted answers, and the results are robust.

Female participants omitting more answers when there is a penalty for wrong answers or equivalently a reward for not answering, as is the case here, has been found previously by Swineford (1941) and Anderson (1989), and more recently by Tannenbaum (2012), Espinosa and Gardeazabal (2013), and Baldiga (2014). Omitting more answers is compatible with two underlying behavioral differences: On the one hand lower confidence in the likelihood that one will know the right answer should lead to more questions not being answered. On the other hand, for the same level of confidence a higher risk aversion should also lead to more questions not being answered. Women are found to be on average more risk averse and less confident than men (Croson and Gneezy, 2009). Baldiga (2014) uses a laboratory design to show that female participants omitting more answers is partly explained by differences in risk aversion, and not by differences in confidence. Furthermore, when we control for the number of answers omitted, as shown in columns 7 and 8, we see that female participants indeed decrease the proportion of right answers as competitive pressure increases. Competitive pressure increases the gender gap in the number of omitted and the number of correct answers, which might be

[7] We have also replicated the regressions without combining different sources for the Math grade at school. When we use only the Math grade reported by the teacher we restrict the sample to 1,767 observations and if we use only the self-reported Math grade we restrict the sample to 1,698 observations. The Math grade at school is always positive and significantly different from zero and the coefficient of interest, the interaction between *Female* and *Stage 2*, is always negative, of similar magnitude as the one in column 3, and becomes significant at the 5% level.

due to either gender differences in risk preferences or in confidence. However, the results we find when controlling for difficulty suggest this may not be straightforward (see section 4).

We perform various robustness checks. The results are reported in Table A.5 in the appendix. First, we replicate the same analysis in two additional samples: the overall sample (columns 1-2) and the matched sample (5-7). Using probability score matching (see Rosenbaum and Rubin, 1985, and Caliendo and Kopeinig, 2008 for a practical guide), we match male and female participants without replacement based on their score in stage 1 test once we control for level, shown in columns 5 and 6. The matched sample consists of a total of 1,162 participants, 581 boys and 581 girls. We perform a regression analysis on this matched sample. Alternatively, we have also matched male and female participants based on their Math grade at school after we control for level and school. Results are shown in columns 7 and 8 in Table A.5. This last matched sample consists of 1,006 participants, with 503 boys and 503 girls. Results are qualitatively similar. However, the Math grade at school is an internal grade based on several different evaluations (which may or may not include standardized tests in all cases). Thus, we believe it is more objective to match participants based on performance in the same test, such as test 1 in the competition. The increase in female underperformance as competitive pressure increases is always negative, significant and of a similar magnitude. The magnitude is substantially greater when the analysis is restricted to comparable male and female participants based on their performance in stage 1 of the contest, i.e. on the matched sample. Note that in the matched sample, male and female participants' performance distributions in stage 1 and have the same variance, as shown in Figure 1. Second, our results are robust to alternative specifications: we estimate individual fixed effects (columns 2, 4 and 6) for all three samples: overall, balanced and matched. The coefficient of the variable of interest, the interaction term between *Female* and *Stage 2*, is negative and of about the same magnitude and significance in all specifications. Note that when using individual fixed effects will identify the interaction between *Female* and *Stage 2* coefficient only based on the sample with both stage 1 and stage 2 scores. The results for other dependent variables such as number of omitted, right and wrong answers and proportion of right answers are also robust to additional samples and specifications.[8]

[8] Note that competitive pressure in our setting includes two main factors: higher ability of participants and a lower ratio of winning positions. In fact, we have variation in these two factors across

3.2. Heterogeneity: age-levels, school year, ability, and school characteristics

We now exploit the structure of the contest to test for heterogeneity in the gender differential reaction to increasing competitive pressure with respect to four different dimensions: age-levels, school year within a contest level, ability, and school characteristics.

[Table 4: Heterogeneity Analysis]

We first test whether the increase in female underperformance as competitive pressure increases as estimated in the previous sub-sections differs according to age-level. This may be related to the hypothesis of whether gender differences under competition are due to nature or nurture. If the female negative reaction to competitive pressure is due to cultural reasons then the effect may be expected to increase as age rises, i.e. with longer exposure to culture and socialization. Table 2 shows that gender differences are larger at higher levels, but this gap remains the same in both stages, which suggests that the age effect on the gender gap is independent of competitive pressure. Columns 1-4 in Table 4 show the estimation results separately by age-levels which confirm the result. We cannot reject the null hypotheses that the increase in the gender gap from stage 1 to stage 2 is the same across the different age-levels (*p*-value of 0.8898).

There is also variation in academic years within each level in the contest. Participants within each level come from two consecutive academic years. We define *Lower Academic Year*, which takes the value of 1 when students are 10, 12, 14 or 16 years old and 0 when they are 11, 13, 15 or 17 years old in Levels 1, 2, 3, and 4, respectively. Although both the lower and upper academic years within each level should be familiar with the material required for them to do well in each test level, the lower/higher academic year students may feel less/more pressure to do well given they have had less/more exposure to the

different schools-levels using only stage 1 data. We can exploit this variation across schools-levels in both Math ability, measured by *School_Math_Quality*, and number of participants, measured by *Size*, using just stage 1 data, to test whether female participants perform relatively worse when competing against higher ability peers or in larger schools. We find that the interaction terms between gender and the measures of ability in Math (*Female***School_Math_Quality*) and between the gender and the measure of size (*Female***Size*) are both negative suggesting that female participants in higher ability and larger schools perform relatively worse than male participants. However, the interaction with higher ability is not statistically significant, although the interaction with size is. We interpret this as additional suggestive evidence that these two factors contribute to the increase in competitive pressure and indeed affect female participants' performance negatively.

knowledge of mathematics. We can thus estimate the interaction between the variables *Female* and *Lower Academic Year*. Columns 5-6 in Table 4 present the results. Again, the null hypothesis that the increase in female underperformance with competitive pressure is the same between the lower and the higher academic years (*p*-value 0.616) cannot be rejected.

Third, using performance data from stage 1, we can define a proxy for participants' ability. We define *Low Ability*, which takes a value of 1 when students perform below the median within our sample in stage 1 of the contest and 0 otherwise.[9] Columns 7-8 in Table 4 show the estimation results for low and high ability separately. We find that for low ability participants the gender differential reaction to increases in competitive pressure is lower than for high ability participants, which is significant at the 1%. This shows that the high ability participants are more affected by the differential gender reaction to competitive pressure.

Finally, we also exploit the variation in the school characteristics to test for heterogeneity effects. The bottom columns of Table 4 show the results for school size in column 9, school overall quality in column 10 and school math quality in column 11. We find that girls in larger schools do significantly worse. However, we find no heterogeneity effects on the gender gap as competitive pressure increases. Columns 12-13 show the heterogeneity across public and non-public schools and finally columns 14-15 show the heterogeneity across school locations. We find no evidence for heterogeneous effects based on these two school characteristics.

4. Alternative Explanations: Selection, Discrimination and Difficulty

We have shown that the amount by which girls underperform boys increases when we move from stage 1 to stage 2. This identifies a gender difference in reaction to competitive pressure. In this section we rule out three alternative explanations based on selection, discrimination and difficulty.

Regarding selection, the idea that male and female participants could come from different types of school could be a concern. Table 5 shows the mean values of the school

[9] We have also used school Math grade in order to identify low and high ability participants. We also find that the gender differential in reaction to increasing competitive pressure is lower among low ability participants, significant at 5%. Nevertheless, we believe the performance in stage 1 of the competition is a better proxy for ability in performing in stage 2 than the Math grade at school.

characteristics by gender for four different samples. The top panel shows the school characteristics by gender for the balanced sample, which shows that girls are more likely to come from public, smaller, and lower quality schools. In order to understand whether these are general differences or differences coming from participants who make it to the second stage, the same means by gender can be compared in the overall sample and the sample of participants who make it to the second stage. It turns out that the main differences come from the participants who make it to the second stage. For example, the differences in the quality of schools are not significant in the overall sample but become significant in the stage 2 sample. However, this cannot be the explanation for our main result in Table 3, as the interaction between *Female* and *Stage 2* is negative and significant when we control for all these school characteristics (shown in column 2). That is, the gender gap increases even when we control for school characteristics. Furthermore, in the bottom panel we show that there are no significant differences in school characteristics by gender for the matched sample, except for the size of schools. We have already shown that although there are fewer observations, in this matched sample our main result is negative, significant and in fact higher in magnitude than in the balanced sample. All this shows that the differences between male and female participants as regards school characteristics cannot be the main explanation for the result observed.

Regarding discrimination, it must be noted that the students who go through from stage 1 to stage 2 are selected by teachers within schools. This raises concern that schools may be selecting on the basis of criteria other than performance in stage 1 of the contest. For example, Lavy (2008) and Cornwell, Mustard, and Van Parys (2011) conclude that school teachers discriminate in favor of girls when grading. Similarly, if school teachers discriminate in favor of or against female participants when selecting them for stage 2 of the contest, we would be comparing male and female participants with different ability levels. In other words, given that we set out here to study gender, there is cause for concern if equally well performing male and female participants have different likelihoods of being selected for stage 2. Similarly, it could also be that male and female participants have different dropout rates, that is, once being selected they can still decide not to show up the day of the stage 2 test.

Regarding selection from stage 1 to stage 2, for those participants for whom we can indeed observe their stage 1 score, 87% of the participants selected into stage 2 are indeed the top performers in their school and level. Furthermore, half of the remaining 13% are just below the top performers. By gender, as shown by the probability distribution in Figure 1, male participants show higher performance than female participants in stage 1, which is confirmed both within the top performers (87%) and those just below the top performers (17%). This shows that teachers indeed do a good job in selecting their participants. Regarding dropout rates, from a total of 3,233 participants initially signed up to participate in stage 2,441 of them do not show up the day of the stage 2 test (13.64%). By gender, male and female participants show similar dropout rates (12.8 and 14.8%, respectively).

The estimation results in Table 6 rule out discrimination and gender differential dropout rates as alternative explanation. Columns 1 and 2 show that female contestants do not have a different likelihood of being selected for stage 2 once performance in stage 1 is controlled for. As expected, performance in stage 1 of the contest is positively and highly significant in predicting qualifying for stage 2. Therefore, although fewer female participants get to stage 2 of the competition this is due to their lower performance in stage 1 and not because they are discriminated against. Also in regard to the likelihood of showing up to stage 2, only 13.64% of participants who are eligible for stage 2 voluntarily drop out of the contest. In columns 3 and 4 of Table 6, we test whether male and female participants have different likelihoods of dropping out of the contest but we find no evidence of this. We also perform the same analysis to see if the winners are selected correctly by the organizers. Estimation results are shown in columns 5 and 6 in Table 6. Again, it can be seen that once performance in stage 2 is controlled for female participants do not show a different likelihood of being selected as prize winners.

In summary, it emerges that the high proportion of men over women in stage 2, with 66% male and 34% female participants, as well as among winners, with only 13% of women, are explained by female participants performing significantly worse than male participants in both stage 1 and stage 2. It also emerges that our main finding – the increase in female underperformance when moving from stage 1 to stage 2 – is not due to gender differences in the selection process by the institution or by themselves.

Regarding differences in difficulty, the contest organizers privately informed us that they designed the stage 2 test to be more difficult than the stage 1 test with the goal of preventing ties when selecting the final winners. The mean values for the score in stages 1 and 2 in Table 2 already confirm that participants find the stage 2 test harder than the stage 1 test. Since the stage 2 test is more difficult than the stage 1 test, one may argue that the increasing gender gap in performance from stage 1 to stage 2 may be due to women underperforming when facing harder questions. We test this alternative explanation using variation in the level of difficulty at question level. Each test consists of 25 different questions in each of which participants can score 0 for a wrong answer, 1 for not answering, and 5 for a correct answer. For the restricted sample of participants who do both stage 1 and 2 tests, the average score per question varies from 0.85 to 3.81 in all levels and in both stages 1 and 2.[10] Regarding the content of questions with higher (easier) and lower (harder) average score, although we could not identify a clear pattern among the highly scored easiest questions, we did find that the questions with the lowest average score, that is, the hardest questions in most tests (6 out of 8, 2 stages by 4 levels in each) contained trigonometry problems.

Table 7 shows the estimation results for regressions with the dependent variable being the standardized performance or score in a particular question by level in stage 1 in columns 1-3, and in stage 2 in columns 4-6, at each of the competition level. We construct two alternative controls for difficulty: the *Easy Dummy* takes the value of 1 if the question is above the median score in a particular level, i.e., among the easiest questions, and 0 otherwise. *Mean Score*, on the other hand, measures easiness continuously as the mean score across all participants within a level and stage. For stage 1 regressions the interaction between *Female* and the control for the relative easiness of the question shows that female underperformance is independent of the difficulty of the question. On the other hand, and contrary to intuition, for stage 2 regressions the estimation results in columns 5 and 6 show that female underperformance is *greater* in the relatively *easier* questions.

More importantly, we can replicate our main results, shown in Table 3, with the standardized score at question level as the dependent variable instead of the standardized score at stage level, where we can now control for the difficulty of the question at hand. The estimation results, shown in Table 8, rule out differences in difficulty as an explanation. The negative and significant interaction between *Female* and *Stage 2,* is

[10] In particular, in stage 1, in level 1 the average score per question varies from 1.05 to 4.46, in level 2 from 0.93 to 4.35, in level 3 from 0.69 to 4.44 and in level 4 from 1.05 to 3.39. In stage 2, in level 1 the average score per question varies from 0.97 to 4.58, in level 2 from 0.62 to 3.61, in level 3 from 0.62 to 2.63 and in level 4 from 0.84 to 3.04.

found to be robust even after difficulty is controlled for. This is the case for the main outcome variable - performance - shown in columns 1-2, as well as for the omitted answers variable in columns 3-4, and right answers dependent variables in columns 5-6.

5 Interpreting Results Using Evidence from a Questionnaire

In order to help us interpret the main results, we conduct a survey-type questionnaire using the 2017 edition of *Concurso de Primavera de Matemáticas,* since participants in the 2014 edition were not available to us. The questionnaire was administered during the stage 2 test to all participants taking the test. It was in the backside of the answer-sheet, which had to be handed in after the test. During instructions, participants were asked to respond the questionnaire before handing in the answer-sheet. Most stage 2 participants responded to the full questionnaire, with an average response rate of 90%. However, since there was no obligation to answer, the number of observations varies with the question.[11] The questionnaire included 10 different questions, which can be found in Figure 2, as well as the distribution of answers, separated by gender.[12]

Questions 1 to 6 ask participants about how they value different stages. Participants reveal that it is more important (for them and for their parents) to be selected for stage 2 than winning stage 2 (Questions 1 and 2). They also reveal that it is more important for them to do well in stage 2 than in stage 1 (Question 3). Note that this is not necessarily in contradiction with their answers to Questions 1 and 2, since in order to do well in stage 2 one needs to qualify first to pass to stage 2. Accordingly, participants say they studied more to prepare for Stage 2 than for Stage 1 (Question 4). Question 5 asks participants directly how many hours they devoted to prepare stage 2 test, with an average of almost 5 hours. Most importantly for our interpretation of results, participants claimed to feel more pressure during stage 2 of the competition than in stage 1 (Question 6). Question 7 and 8 ask them about their confidence in answering the questions. The majority of participants answered a stage 2 test question only when they are absolutely sure or sure about the answer. Finally, Questions 9 and 10 ask them to self-assess their own Math

[11]There are 3,406 participants taking the stage 2 test in the 2017 edition. In particular, the response rates for Questions 1 to 10 were given by 96%, 96%, 96%, 96%, 91%, 92%, 76%, 83%, 94% and 94%, respectively.

[12] Although we have only been able to obtain performance data from both stages 1 and 2 of the competition in the 2014 edition of the contest, we do have access to stage 2 data from subsequent years, where the main structure of the competition remained the same. In particular, in the 2017 edition, all features remain the same, except for the fact that in the stage 2 test, the test questions 1-13, omitted and incorrect answers yield 0 points, while test questions 14-25 omitted answers yield 1 point while incorrect answers yield 0 points. The difference in scoring rule will be further studied in a companion paper.

ability, where as expected all participants believe they are good at Math, and their perception about whether gender differences in Math ability exist, or whether Math is perceived to be gender neutral. A large majority of participants perceive that Math is gender neutral, which is in sharp contrast to previous findings (Fryer and Levitt, 2010, Bharadwaj et al., forthcoming, Nollenberger et al., 2016), where Math is perceived to be a male task.

We now look at gender differences in the responses to questions 1 to 6 in the questionnaire, which are most related to understanding why men and women might differ in their performance between stages 1 and 2. Female participants (as well as their parents) agree stronger than male participants with the statement that being in stage 2 is more important than winning stage 2. Also, male participants agree stronger than female participants that doing well in stage 2 is more important than doing well stage 1. These gender differences go in the expected direction that female participants may value less stage 2 than stage 1. However, women claim to put in more effort than men in preparing stage 2 (Question 4), while in fact they do not differ in the actual number of hours devoted to stage 2 preparation (Question 5). Finally, in Question 6, female participants indeed claimed that they felt more pressure than male participants.

To sum up, we find no evidence for differences in effort provision to explain the results. Moreover, we do find evidence that female participants indeed value more being selected for stage 2 than winning stage 2 than men do and also value less doing well in stage 2 than men do. However, at the same time, they claim to feel more pressure while doing stage 2 test than stage 1 test. This apparent contradiction may be related to gender differences in how much they may care for their projected self-image regarding performance in competitions; while women say it is more important to participate, men say it is more important to win. Most importantly, and consistent with our main results, the perceived pressure is higher for women than men.

6 Discussion

Field data from two-stage elimination math contests in which individual performance of the same subject can be traced as students qualify for further stages of the competition offers a unique opportunity to test for and measure gender differences in performance as competitive pressure increases. Our setting resembles many of the features found in hierarchical organizations, so we identify an important source for the diminishing female presence as one moves up in multi-stage elimination contest-like hierarchical organizations in the labor market. We find that the gender gap in performance does indeed increase from stage 1 to stage 2, which contributes to the lower presence of women in later stages. We attribute this to changes in competitive pressure and rule out alternative explanations based on selection, discrimination and gender differential reactions to difficulty. The increase in female underperformance is not explained only by female participants being more risk averse or less confident (as shown by the fact that the number of answers omitted increases as competitive pressure increases): even when we control for the number of answers omitted the number of right answers given by female contestants *decreases*.

Two important questions remain open regarding the mechanism underlying our result. The first concerns the task used in our contests, i.e. mathematics tests, a task in which men regularly perform better. Further research should be conducted to determine whether increases in competitive pressure have similar differential effects in gender neutral or even female favoring tasks, where stereotypes should not be a threat (Iriberri and Rey-Biel, 2017). Second, the underlying mechanism determining why female participants react in this way to increasing competitive pressure remains to be explored. In the questionnaire we found that female participants seem to value less to win and to do well in the second stage than in the first stage but yet reveal to feel more pressure. We intend to explore these questions in future research, which may potentially help in the design of optimal policies aimed at eliminating the glass ceiling effect.

References

Ahern, K. R. and A K Dittmar. 2012. "The changing of the boards: The Impact on firm valuation of mandated female board representation," *The Quarterly Journal of Economics*, 127 (1), 137–197.

American Bar Association, Commission on Women in the Profession. "A Current Glance at Women in the Law 2014".

Anderson, J. 1989. "Sex-related Differences on Objective Tests among Undergraduates." *Educational Studies in Mathematics*, 20: 165–177.

Azmat, G., C. Calsamiglia, and N. Iriberri. 2016. "Gender Differences in Response to Big Stakes." *Journal of the European Economic Association* 14(6).

Baldiga, K., 2014. "Gender Differerences in Willingness to Guess." *Management Science* 60(2): 434–448.

Bertrand, M., 2009. "CEOs." *Annual Review of Economics* 1(1): 121–150.

Bertrand, M., and K. Hallock. 2001. "The gender gap in top corporate jobs." *Industrial and Labor Relations Review*, 55: 3–21.

Bharadwaj P., G. Di Giorgi, D. Hansen and C. Neilson. "The Gender Gap in Mathematics: Evidence from Chile". Forthcoming at the *Economic Development and Cultural Change*.

Booth, A. Cardona-Sosa, L.M. and P. J. Nolen. 2013. "Do Single-Sex Classes Affect Exam Scores? An Experiment in a Coeducational University"

Blau, F., M. Ferber, and A. Winkler. 2010. "The Economics of Women, Men and Work." Engelwood Cliffs, NJ: Prentice Hall.

Brown, J., 2011. "Quitters Never Win: The (Adverse) Incentive Effects of Competing with Superstars*.*" *Journal of Political Economy*, 119(5): 982-1013.

Brown, J. and D.B. Minor., 2014. "Selecting the Best? Spillover and Shadows in Elimination Tournaments." Management Science, 60(12): 3087-3102.

Buser, T., M. Niederle and H. Oosterbeek, 2014. "Gender, Competitiveness and Career Choices". *The Quarterly Journal of Economics*, 129(3): 1409-1447.

Caliendo, M. and S. Kopeinig. 2008. "Some Practical Guidance for the Implementation of Propensity Score Matching." *Journal of Economic Surveys*, 22(1): 31–72.

Croson, R. and U. Gneezy, 2009. "Gender Differences in Preferences," *Journal of Economic Literature*, 47(2): 1–27.

Curtis, J. W., 2011. "Persistent Inequity: Gender and Academic Employment." American Association of University Professors.

Ellison, G. and A. Swanson. 2010. "The Gender Gap in Secondary School Mathematics at High Achievement Levels: Evidence from the American Mathematics Competitions." *Journal of Economic Perspectives*, 24(2): 109–28.

Ellison, G. and A. Swanson. 2016. “Do Schools Matter for High Math Achievement? Evidence from the American Mathematics Competitions” *American Economic Review*, 106(6): 1244–1277.

Espinosa, M. P., and J. Gardeazabal. 2013. “Do Students Behave Rationally in Multiple Choice Tests? Evidence from a Field Experiment.” *Journal of Economics and Management*, 9(2): 107–135.

Fryer, R. and Levitt, S., 2010. “An Empirical Analysis of the Gender Gap in Mathematics”. *American Economic Journal: Applied Economics*, 2(2): 210–240.

Gayle, G. L., L. Golan and R. A. Miller (2012) “Gender differences in executive compensation and job mobility”, *Journal of Labor Economics*, 30 (4): 829–872.

Gneezy, U., M. Niederle and A. Rustichini. 2003. “Performance in Competitive Environments: Gender Differences.” *The Quarterly Journal of Economics*, 118 (3):1049–1074.

Iriberri, N. and P. Rey-Biel. 2017. “Stereotypes are Only a Threat when Beliefs are Reinforced: On the Sensitivity of Gender Differences in Performance under Competition to Information Provision.” *Journal of Economic Behavior and Organization*, 135: 99–111.

Jurajda, Š., and D. Münich. 2011. “Gender Gap in Performance under Competitive Pressure: Admissions to Czech Universities.” *American Economic Review Papers and Proceedings*, 101(3): 514–18.

Matsa, D. A., and A. R. Miller. 2013. "A Female Style in Corporate Leadership? Evidence from Quotas." *American Economic Journal: Applied Economics*, 5(3): 136–69.

Niederle, M., and L. Vesterlund. 2007. “Do Women Shy away from Competition? Do Men Compete too Much?” *The Quarterly Journal of Economics*, 122(3): 1067–1101.

Nollenberger, N., Sevilla, A., Rodriguez-Planas N. 2016. “The Math Gender Gap: The Role of Culture.” American Economic Review 106(5): 257–261.

Örs, E., F. Palomino and E. Peyrache. 2013. “Performance Gender Gap: Does Competition Matter?” *Journal of Labor Economics*, 31(3): 443–499.

Paserman, D. 2010. “Gender Differences in Performance in Competitive Environments? Evidence from Professional Tennis Players”. Mimeo.

Rosenbaum, P., and D. Rubin. 1985. “Constructing a Control Group Using Multivariate Matched Sampling Methods that Incorporate the Propensity Score.” *The American Statistician*, 39: 33–38.

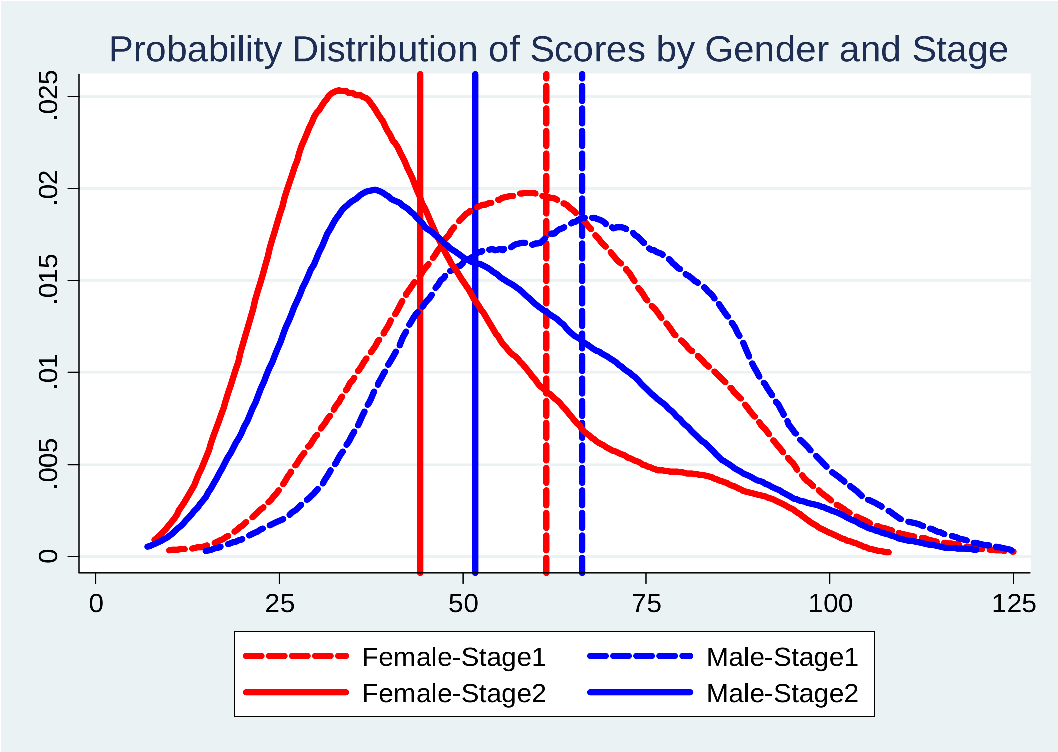
Swineford, F. 1941. “Analysis of a Personality Trait.” *Journal of Educational Psychology*, 45: 81–90.

Tannenbaum, D. 2012. “Do Gender Differences in Risk Aversion Explain the Gender Gap in SAT Scores? Uncovering Risk Attitudes and the Test Score Gap.” Mimeo.

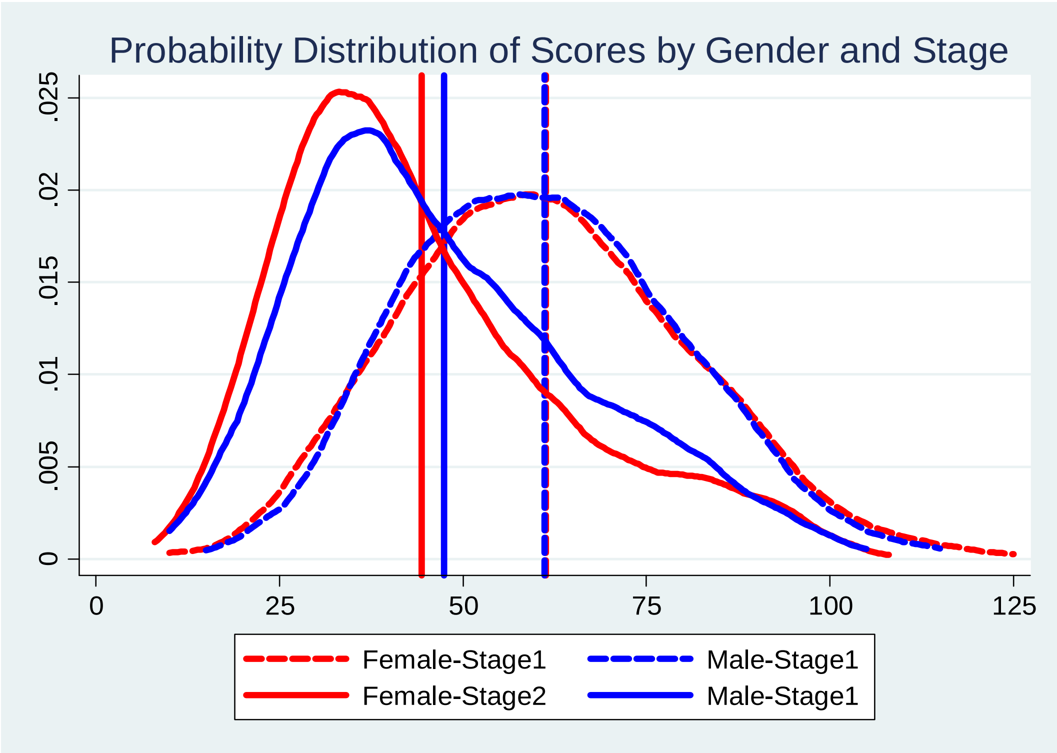
Wolfers, J. 2006. “Diagnosing Discrimination: Stock Returns and CEO Gender.” *The Journal of the European Economic Association*, 4(2/3): 531–541.

Wood, R. G., Corcoran, M.E. and Courant, P.N., 1993. “Pay Differences among the Highly Paid: The Male-Female Earnings Gap in Lawyers' Salaries.” *Journal of Labor Economics*, 11(3):417–441.

Figures and Tables

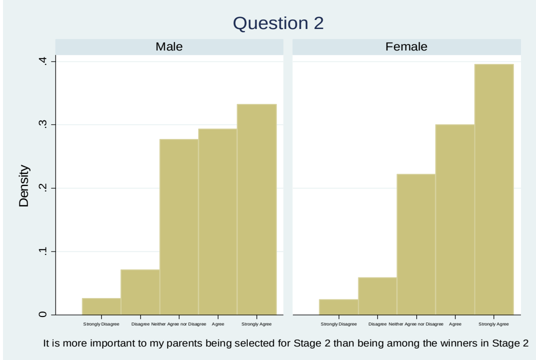


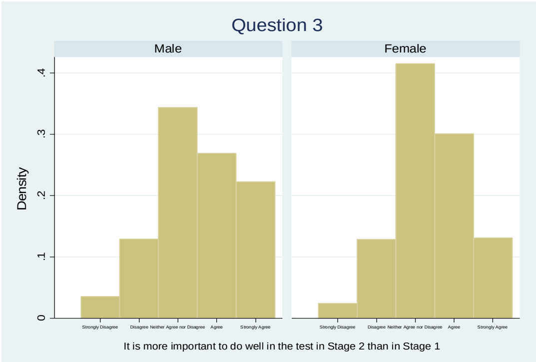
(a) Balanced Sample

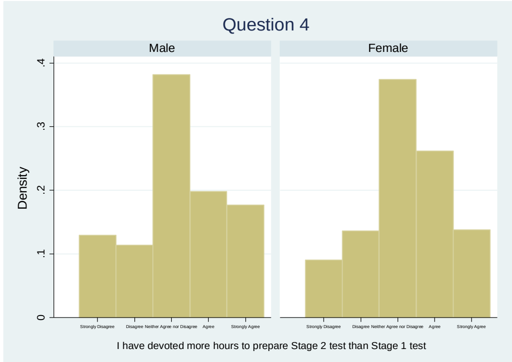


(b) Matched Sample

Figure 1. Probability Distributions of Performance by Gender and Stage

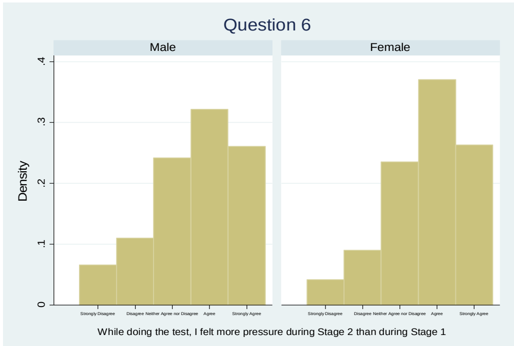






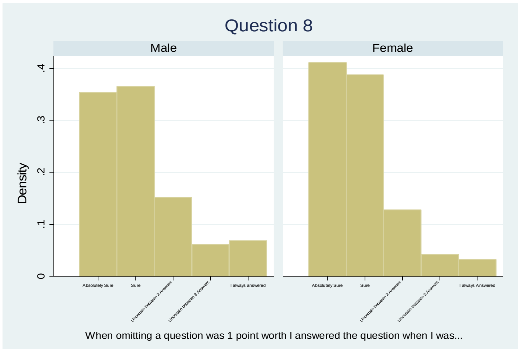
Question 5: No. Of Hours Devoted to Prepare Stage 2 Test

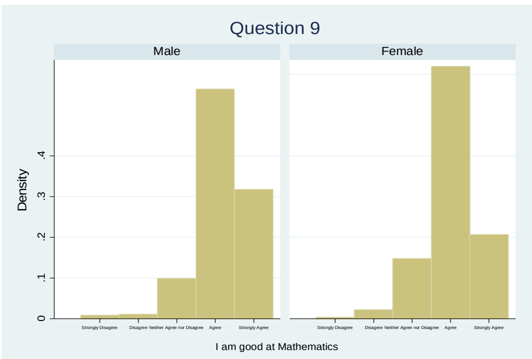
	Overall	Male	Female
Mean	4,77	4,91	4,51
St. Deviation	11,62	12,93	8,54
No. Of Participants	3090	2033	1057



Question 7: Guessed No. Of Correct Answers

	Overall	Male	Female
Mean	12,66	13,06	11,92
St. Deviation	4,97	4,78	5,21
No. Of Participants	2580	1674	906





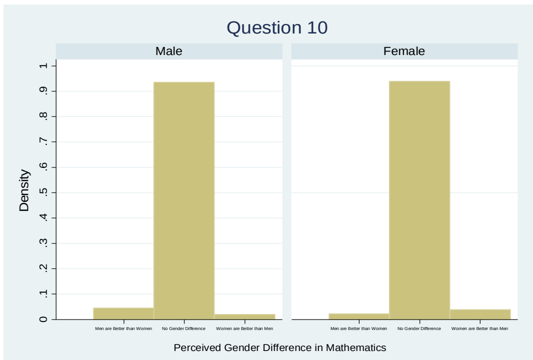


Figure 2. Answers to the Survey Type Questionnaire by Gender

Table 1. Competition in Stages 1 and 2

			Stage 1		Stage 2	
		Obs.	Mean	St. Dev.	Mean	St. Dev.
		(1)	(2)	(3)	(5)	(6)
Level 1	Performance in Stage 1	443	54.78	14.96	77.7	15.17
	Math at School	440	8.15	0.82	8.77	0.76
	No. Of Competitors	443	40.62	35.21	655	
	No. Of Winning Positions	443	3.92	1.05	36	
	Prop. Of Winning Positions	443	0.27	0.32	0.05	
Level 2	Performance in Stage 1	593	44.62	13.53	69.24	18.84
	Math at School	593	7.62	0.98	8.62	1.10
	No. Of Competitors	593	39.42	44.06	910	
	No. Of Winning Positions	593	2.79	0.48	45	
	Prop. Of Winning Positions	593	0.20	0.26	0.05	
Level 3	Performance in Stage 1	508	39.26	10.13	55.16	16.48
	Math at School	508	7.26	1.08	7.99	1.54
	No. Of Competitors	508	23.98	30.13	784	
	No. Of Winning Positions	508	2.72	0.53	41	
	Prop. Of Winning Positions	508	0.28	0.26	0.05	
Level 4	Performance in Stage 1	269	40.07	11.13	51.01	15.36
	Math at School	269	7.42	1.17	7.92	1.57
	No. Of Competitors	269	11.94	11.94	443	
	No. Of Winning Positions	269	2.59	0.64	25	
	Prop. Of Winning Positions	269	0.43	0.32	0.06	

Notes: The table reports the number of observations, the mean values and the standard deviations for the main competition measures for stage 1, columns (2) to (3) and for stage 2, columns (4) to (5). These numbers, except for the *No. of Competitors*, *No. Of Winning Positions* and *Prop. of Winning Positions* in stage 2 (which use the overall sample for stage 2), are calculated using the sample of students for whom we have both stage 1 and 2 performance (balanced sample).

Table 2. Descriptive Statistics for the Balanced Sample

	Overal		Male		Female		
Overall:	Obs.	Mean (St. Dev)	Obs. (Percent)	Mean (St. Dev)	Obs. (Percent)	Mean (St. Dev)	*p* -value
Performance Data:							
Math at School (0-10)	1803	8.36 (1.30)	1229 (68%)	8.33 (1.36)	574 (32%)	8.42 (1.18)	0.17
Performance in Stage 1 (0-125)	1813	64.65 (19.60)	1232 (68%)	66.22 (19.55)	581 (32%)	61.32 (19.30)	0.00
No. Of Omitted (0-25)		6.62 (4.81)		6.32 (4.82)		7.26 (4.73)	0.00
No. Of Right (0-25)		11.61 (4.45)		11.98 (4.46)		10.81 (4.34)	0.00
No. Of Wrong (0-25)		5.87 (3.49)		5.83 (3.37)		5.95 (3.72)	0.49
Performance in Stage 2 (0-125)	1813	49.32 (20.78)	1232 (68%)	51.66 (21.11)	581 (32%)	44.37 (19.14)	0.00
No. Of Omitted (0-25)		8.39 (5.27)		7.85 (4.74)		9.55 (4.31)	0.00
No. Of Right (0-25)		8.19 (4.68)		8.76 (4.74)		6.96 (4.31)	0.00
No. Of Wrong (0-25)		8.42 (4.57)		8.39 (4.62)		8.48 (4.48)	0.69
Winners	85	0.05 (0.21)	75 (88%)	0.06	10 (12%)	0.02	0.00
Level 1:							
Math at School (0-10)	434	8.77 (0.76)	297 (68%)	8.74 (0.78)	137 (32%)	8.85 (0.71)	0.14
Performance in Stage 1 (0-125)	443	77.67 (15.17)	300 (68%)	78.51 (15.12)	143 (32%)	75.92 (15.20)	0.09
No. Of Omitted (0-25)		3.62 (3.25)		3.24 (3.05)		4.41 (3.50)	0.00
No. Of Right (0-25)		14.81 (3.35)		15.05 (3.31)		14.30 (3.39)	0.03
No. Of Wrong (0-25)		6.27 (2.93)		6.39 (2.85)		6.02 (3.09)	0.22
Performance in Stage 2 (0-125)	443	67.38 (19.34)	300 (68%)	68.98 (19.29)	143 (32%)	64.04 (19.06)	0.01
No. Of Omitted (0-25)		5.07 (3.76)		4.68 (3.71)		5.90 (3.75)	0.00
No. Of Right (0-25)		12.46 (4.19)		12.86 (4.19)		11.63 (4.07)	0.00
No. Of Wrong (0-25)		7.47 (3.99)		7.46 (3.89)		7.48 (4.21)	0.98
Win Prize (0-1)	15	0.03	13 (86%)	0.04	2 (14%)	0.01	0.11
Level 2:							
Math at School (0-10)	592	8.59 (1.10)	386 (65%)	8.59 (1.12)	206 (35%)	8.58 (1.06)	0.94
Performance in Stage 1 (0-125)	593	69.24 (18.83)	386 (65%)	71.63 (18.62)	207 (35%)	64.77 (18.46)	0.00
No. Of Omitted (0-25)		5.72 (3.94)		5.38 (3.95)		6.37 (3.87)	0.00
No. Of Right (0-25)		12.70 (4.07)		13.25 (4.06)		11.68 (3.87)	0.00
No. Of Wrong (0-25)		5.42 (3.41)		5.32 (3.22)		5.60 (3.74)	0.35
Performance in Stage 2 (0-125)	593	47.21 (19.29)	386 (65%)	50.16 (20.25)	207 (35%)	41.71 (16.00)	0.00
No. Of Omitted (0-25)		7.77 (4.45)		7.16 (4.31)		8.91 (4.49)	0.00
No. Of Right (0-25)		7.89 (4.21)		8.60 (4.40)		6.56 (3.47)	0.00
No. Of Wrong (0-25)		9.34 (4.40)		9.24 (4.41)		9.53 (4.39)	0.46
Win Prize (0-1)	31	0.05	25 (81%)	0.06	6 (19%)	0.03	0.06
Level 3:							
Math at School (0-10)	508	7.99 (1.54)	344 (68%)	7.92 (1.64)	164 (32%)	8.13 (1.32)	0.14
Performance in Stage 1 (0-125)	508	55.16 (16.48)	344 (68%)	57.33 (16.90)	164 (32%)	50.62 (14.63)	0.00
No. Of Omitted (0-25)		8.22 (4.66)		7.76 (4.69)		9.18 (4.44)	0.00
No. Of Right (0-25)		9.39 (3.66)		9.91 (3.76)		8.29 (3.18)	0.00
No. Of Wrong (0-25)		6.19 (3.68)		6.10 (3.53)		6.38 (3.98)	0.42
Performance in Stage 2 (0-125)	508	39.28 (15.53)	344 (68%)	41.69 (16.45)	164 (32%)	34.22 (11.94)	0.00
No. Of Omitted (0-25)		10.36 (5.36)	344	9.60 (5.43)		11.96 (4.86)	0.00
No. Of Right (0-25)		5.78 (3.43)	344	6.42 (3.59)		4.45 (2.62)	0.00
No. Of Wrong (0-25)		8.85 (4.92)	344	8.98 (5.13)		8.59 (4.46)	0.40
Win Prize (0-1)	22	0.04	20 (91%)	0.06	2 (9%)	0.01	0.02
Level 4:							
Math at School (0-10)	269	7.92 (1.57)	202 (75%)	7.96 (1.60)	67 (75%)	7.77 (1.46)	0.38
Performance in Stage 1 (0-125)	269	51.01 (15.36)	202 (75%)	52.78 (15.88)	67 (75%)	45.67 (12.31)	0.00
No. Of Omitted (0-25)		10.52 (5.21)		10.23 (5.15)		11.42 (5.31)	0.11
No. Of Right (0-25)		8.10 (3.50)		8.51 (3.61)		6.85 (2.82)	0.00
No. Of Wrong (0-25)		5.58 (3.95)		5.50 (3.90)		5.82 (4.13)	0.56
Performance in Stage 2 (0-125)	269	43.20 (15.67)	202 (75%)	45.77 (16.49)	67 (75%)	35.45 (9.38)	0.00
No. Of Omitted (0-25)		11.52 (5.46)		10.89 (5.56)		13.43 (4.68)	0.00
No. Of Right (0-25)		6.33 (3.60)		6.98 (3.81)		4.40 (1.83)	0.00
No. Of Wrong (0-25)		7.14 (4.61)		7.13 (4.60)		7.16 (4.66)	0.96
Win Prize (0-1)	17	0.06	17 (100%)	0.08	0 (0%)	0.00	0.01

Notes : This table reports the number of observations and the mean values for the main performance variables: *Performance* or *Score* , *No. Of Omitted* , *Right* and *Wrong* questions, as well as a dummy variable that takes the value of 1 if the participant wins the competition. The *p* -value are for the F-Test of equality of variable means across gender.

Table 3. Gender Differential in Performance to Competitive Pressure (Balanced Sample)

	Performance (1)	Performance (2)	Performance (3)	No. Of Omitted (4)	No. Of Right (5)	No. Of Wrong (6)	Prop. Of Right (7)	Prop. Of Right (8)
Female	-0.338***	-0.266***	-0.229***	0.171***	-0.245***	0.0389	-0.105**	-0.134***
	(0.0488)	(0.0459)	(0.0417)	(0.0500)	(0.0429)	(0.0506)	(0.0446)	(0.0431)
Stage 2	0.0283	0.0300	0.0283	-0.0351	0.0368	0.0108	0.0337	-0.0314
	(0.0264)	(0.0267)	(0.0276)	(0.0271)	(0.0268)	(0.0285)	(0.0296)	(0.0289)
Female*Stage 2	-0.0920*	-0.0862*	-0.0920*	0.111**	-0.119**	-0.0321	-0.110**	-0.143***
	(0.0480)	(0.0487)	(0.0502)	(0.0510)	(0.0493)	(0.0518)	(0.0556)	(0.0544)
Math at School		0.185***	0.182***	0.0517***	0.148***	-0.200***	0.203***	0.207***
		(0.0143)	(0.0141)	(0.0167)	(0.0144)	(0.0169)	(0.0138)	(0.0132)
No. Of Omitted								0.0426***
								(0.00371)
School FE	No	No	Yes	Yes	Yes	Yes	Yes	Yes
School Characteristics	No	Yes	No	No	No	No	No	No
Observations	3,606	3,524	3,606	3,606	3,606	3,606	3,606	3,606
Number of Participants	1,803	1,762	1,803	1,803	1,803	1,803	1,803	1,803

Notes: Dependent variables measure performance or score (columns 1 to 3), the number of omitted/right/wrong answers (columns 4, 5, 6), and the number of right answers divided by the number of non-omitted questions (columns 7 and 8). All dependent variables show standardized values at each stage and competition level. *Female* takes the value of 1 if the participant is female and 0 otherwise. *Stage 2* takes the value of 1 if the score refers to the second stage and 0 otherwise and *Math at School* measures the school grade in Math. Standard errors, clustered at participant level, are reported in parentheses with *** p<0.01, ** p<0.05, * p<0.1.

Table 4. Gender Differential in Performance to Competitive Pressure: Heterogeneity Analysis

	Performance Level 1	Performance Level 2	Performance Level 3	Performance Level 4	Performance Lower Ac. Year	Performance Higher Ac. Year	Performance Low Ability	Performance High Ability
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Female	-0.0412	-0.198***	-0.325***	-0.325***	-0.185***	-0.178***	-0.0737*	-0.0779
	(0.0874)	(0.0658)	(0.0876)	(0.118)	(0.0710)	(0.0657)	(0.0406)	(0.0591)
Stage 2	0.0228	0.0257	0.0204	0.00540	0.0118	0.0368	0.368***	-0.274***
	(0.0533)	(0.0507)	(0.0564)	(0.0843)	(0.0501)	(0.0418)	(0.0379)	(0.0375)
Female*Stage 2	-0.0816	-0.0786	-0.0774	-0.157	-0.115	-0.0534	-0.124**	-0.322***
	(0.111)	(0.0860)	(0.102)	(0.162)	(0.0894)	(0.0801)	(0.0600)	(0.0801)
Math at School	0.326***	0.218***	0.173***	0.0841***	0.137***	0.213***	0.108***	0.164***
	(0.0441)	(0.0227)	(0.0250)	(0.0268)	(0.0305)	(0.0231)	(0.0136)	(0.0225)
Observations	866	1,184	1,010	516	1,138	1,684	1,886	1,720
No. of Participants	433	592	505	258	569	842	943	860

	Performance Size	Performance School_Overall_Quality	Performance School_Math_Quality	Performance Public	Performance Non-Public	Performance Madrid	Performance Outside Madrid
	(9)	(10)	(11)	(12)	(13)	(14)	(15)
Female	-0.0564	-0.164*	-0.0959	-0.209***	-0.318***	-0.341***	-0.198***
	(0.102)	(0.0932)	(0.152)	(0.0582)	(0.0747)	(0.0689)	(0.0614)
Stage 2	-0.0391	0.151***	0.257***	0.125***	-0.0803**	-0.0511	0.0921**
	(0.0592)	(0.0553)	(0.0845)	(0.0354)	(0.0402)	(0.0377)	(0.0372)
Female*Stage 2	-0.0434	-0.186*	-0.252	-0.125**	-0.0571	-0.111	-0.0773
	(0.108)	(0.0957)	(0.157)	0.182***	0.186***	(0.0706)	(0.0661)
Math at school	0.185***	0.185***	0.185***	(0.0187)	(0.0225)	0.167***	0.203***
	(0.0143)	(0.0143)	(0.0143)	(0.000591)	(0.000963)	(0.0239)	(0.0192)
Size	0.00458***	0.00406***	0.00406***	0.00404***	0.00462***	0.00554***	0.00277***
	(0.000693)	(0.000512)	(0.000513)	(0.000591)	(0.000963)	(0.000784)	(0.000706)
School_Overall_Quality	0.00153	0.00290*	0.00137	0.000295	0.00237	0.00355	3.25e-05
	(0.00148)	(0.00165)	(0.00148)	(0.00200)	(0.00219)	(0.00235)	(0.00190)
School_Math_Quality	0.0122***	0.0124***	0.0151***	0.0143***	0.00992***	0.00750*	0.0162***
	(0.00253)	(0.00254)	(0.00278)	(0.00352)	(0.00359)	(0.00407)	(0.00321)
Female*Size	-0.00260**						
	(0.00112)						
Stage 2*Size	0.000804						
	(0.000645)						
Female*Size*Stage 2	-0.000471						
	(0.00120)						
Female*School_Overall_Quality		-0.00203					
		(0.00177)					
Stage 2*School_Overall_Quality		-0.00232**					
		(0.000976)					
Female*School_Overall_Quality*Stage 2		0.00188					
		(0.00182)					
Female*School_Math_Quality			-0.0455				
			(0.0564)				
Stage 2*School_Math_Quality			-0.0340				
			(0.0362)				
Female*School_Math_Quality*Stage 2			0.0589				
			(0.0624)				
Observations	3,524	3,524	3,524	1,950	1,574	1,492	2,032
Number of Students	1,762	1,762	1,762	975	787	746	1,016

Notes : Dependent variables measure standardized performance or score at stage and competition level. *Female* takes the value of 1 if the participant is female and 0 otherwise. *Stage 2* takes the value of 1 if the score refers to the second stage and 0 otherwise and *Math at School* measures the school grade in Math. All regressions in the top panel include school fixed effects. Regressions in the bottom panel include school characteristics. Standard errors, clustered at participant level, are reported in parenthesis, with *** p<0.01, ** p<0.05, * p<0.1.

Table 5. School Characteristics by Gender in Different Samples

	Balanced Sample					Overall Sample				
	Male Participants		Female Participants			Male Participants		Female Participants		
School Characteristics	Mean	St. Dev.	Mean	St. Dev.	*p* -value	Mean	St. Dev.	Mean	St. Dev.	*p* -value
Public	0.54		0.59		0.04	0.48		0.46		0.00
Mixed	0.36		0.31		0.05	0.40		0.42		0.00
Private	0.11		0.10		0.73	0.12		0.13		0.19
Madrid	0.43		0.41		0.30	0.44		0.43		0.40
North	0.09		0.09		0.83	0.12		0.14		0.00
South	0.20		0.21		0.49	0.15		0.13		0.00
East	0.12		0.11		0.73	0.12		0.12		0.99
West	0.16		0.18		0.44	0.16		0.17		0.11
Size	85.95	40.39	78.78	41.33	0.00	87.29	37.87	82.49	37.26	0.00
School_Overall_Quality	52.03	27.29	46.54	26.67	0.00	55.25	26.99	55.58	26.63	0.37
School_Math_Quality	54.27	16.22	51.15	15.89	0.00	56.60	16.59	56.69	16.57	0.68
Obs.	1203		569			12140		9310		

	Participants in Stage 2					Matched Sample				
	Male Participants		Female Participants			Male Participants		Female Participants		
School Characteristics	Mean	St. Dev.	Mean	St. Dev.	*p* -value	Mean	St. Dev.	Mean	St. Dev.	*p* -value
Public	0.47		0.51		0.05	0.57		0.59		0.50
Mixed	0.38		0.36		0.25	0.32		0.31		0.78
Private	0.15		0.13		0.21	0.09		0.10		0.75
Madrid	0.44		0.41		0.18	0.42		0.41		0.62
North	0.10		0.10		0.73	0.09		0.09		0.99
South	0.19		0.21		0.27	0.21		0.21		0.92
East	0.10		0.11		0.74	0.12		0.11		0.73
West	0.17		0.17		0.94	0.16		0.18		0.41
Size	86.11	40.14	78.95	38.86	0.00	85.07	41.02	78.78	41.31	0.01
School_Overall_Quality	55.87	27.32	51.53	27.12	0.00	48.85	27.53	46.54	26.66	0.15
School_Math_Quality	56.95	16.96	53.96	16.53	0.00	52.27	16.25	51.15	15.88	0.24
Obs.	1788		898			571		569		

Notes : this table shows the observations, mean values and standard deviations for the school characteristics for the different samples.

Table 6. Selection into Stage 2, Dropping Out and Winning a Prize

VARIABLES	Prob(Stage 2) (1)	Prob(Stage 2) (2)	Prob(Drop Out) (3)	Prob(Drop Out) (4)	Prob(Win Prize) (5)	Prob(Win Prize) (6)
Female	-0.0322***	-0.00558	0.0213	0.0122	-0.0330***	0.00499
	(0.00430)	(0.00385)	(0.0160)	(0.0195)	(0.00932)	(0.00795)
Performance in Stage 1		0.00959***		-0.00101		
		(0.000426)		(0.000676)		
Performance in Stage 2						0.00418***
						(0.000141)
Constant	0.142***	-0.374***	0.0501***	0.117**	0.0716***	-0.436***
	(0.00841)	(0.0233)	(0.0180)	(0.0564)	(0.0162)	(0.0409)
Observations	21,480	20,270	3,233	2,026	2,791	2,791

Notes : Dependent variable in columns 1-2, *Prob(Stage 2)* , takes the value of 1 if the student is selected to participate in stage 2 of the contest, and 0 otherwise; in columns 3-4, *Prob(Drop Out)* , takes the value of 1 if the participant was selected to participate in stage 2 but does not show up and 0 otherwise; and the dependent variable in columns 5-6, *Prob(Win Prize)* , takes the value of 1 if the student wins a prize in the final stage of the contest, and 0 otherwise. We estimate LPM. All regressions include school and level fixed effects. Standard errors, clustered at the school level, are reported in parentheses, *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 7. Female Underperformance Within Stages with Difficulty

	Stage 1 Perf. By Question			Stage 2 Perf. By Question		
	(1)	(2)	(3)	(4)	(5)	(6)
Female	-0.0525***	-0.0522***	-0.0589**	-0.129***	-0.0968***	-0.0611**
	(0.0120)	(0.0163)	(0.0297)	(0.0149)	(0.0169)	(0.0282)
Math at School	0.0449***	0.0449***	0.0450***	0.0844***	0.0844***	0.0842***
	(0.00468)	(0.00468)	(0.00472)	(0.00578)	(0.00578)	(0.00580)
Easy Dummy		0.00119			0.0211*	
		(0.0121)			(0.0115)	
Female*Easy Dummy		-0.000598			-0.0629***	
		(0.0203)			(0.0190)	
Mean Score			-0.000787			0.0104
			(0.00647)			(0.00759)
Female*Mean Score			0.00253			-0.0330***
			(0.0104)			(0.0126)
Observations	49,500	49,500	49,500	51,650	51,650	51,650
Number of Participants	1,980	1,980	1,980	2,066	2,066	2,066

Notes: dependent variables measure the standardized performance or score, at the stage, level and question level, in stages 1 (columns 1-3) and 2 (columns 4-6). *Female* takes the value of 1 if the participant is female and 0 otherwise. *Easy Dummy* takes the value of 1 if the question is among the easiest questions and 0 otherwise. *Mean Score* measures the mean value in the score in the participant population. All regressions include school fixed effects. Standard errors, clustered at the partricipant level, in parentheses *** p<0.01, ** p<0.05, * p<0.1

Table 8. Gender Differential in Performance to Competitive Pressure, Controlling for Difficulty

	Performance by Question		Omitted by Question		Correct by Question	
	(1)	(2)	(3)	(4)	(5)	(6)
Female	-0.0732***	-0.0356	0.0621***	0.112***	-0.0518***	-0.0536
	(0.0138)	(0.0270)	(0.0197)	(0.0346)	(0.0157)	(0.0332)
Stage 2	-0.0132	-0.00872	-0.00949	-0.00817	-0.0112	-0.0161
	(0.00962)	(0.00999)	(0.0113)	(0.0116)	(0.0115)	(0.0119)
Female*Stage 2	-0.0305*	-0.0394**	0.0483**	0.0363*	-0.0490**	-0.0488**
	(0.0168)	(0.0176)	(0.0208)	(0.0215)	(0.0206)	(0.0215)
Math at school	0.0617***	0.0614***	0.0198***	0.0199***	0.0802***	0.0807***
	(0.00481)	(0.00485)	(0.00701)	(0.00701)	(0.00546)	(0.00550)
Mean Score		0.00757		0.00221		-0.00733
		(0.00564)		(0.00563)		(0.00650)
Female*Mean Score		-0.0147		-0.0193**		0.000685
		(0.00923)		(0.00973)		(0.0110)
Observations	86,650	86,650	86,650	86,650	60,227	60,227
No. Of Participants	1,733	1,733	1,733	1,733	1,733	1,733

Notes: dependent variables measure standardized performance or score, at each stage, level and question level. *Female* takes the value of 1 if the participant is female and 0 otherwise. *Stage 2* takes the value of 1 if the score refers to the second stage and 0 otherwise and *Math at School* measures the school grade in Math. *Mean Score* measures the mean value in the score in the participant population. All regressions include school fixed effects. Standard errors, clustered at the participant level, are shown in parentheses, with *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table A.1 Comparing Schools in and out of the Contest

	All Schools in Madrid			Schools Participating in the Contest			
	Obs.	Mean	Stand. Dev.	Obs.	Mean	Stand. Dev.	*p* -value
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Public	1578	0.66	0.47	439	0.57	0.50	0.00
Mixed	1578	0.28	0.45	439	0.31	0.46	0.04
Private	1578	0.07	0.25	439	0.12	0.32	0.00
Madrid	1578	0.41	0.49	439	0.40	0.49	0.88
North	1578	0.08	0.28	439	0.09	0.29	0.35
South	1578	0.25	0.44	439	0.22	0.42	0.06
East	1578	0.14	0.35	439	0.12	0.33	0.12
West	1578	0.11	0.32	439	0.16	0.37	0.00
CDI Test Primary School							
Size	1272	48.05	25.86	250	63.23	30.04	0.00
School_Overall_Quality	1272	49.96	28.88	250	58.49	27.16	0.00
School_Math_Quality	1271	54.46	15.12	250	58.57	13.50	0.00
CDI Test Secondary School							
Size	772	72.64	39.95	369	86.85	39.30	0.00
School_Overall_Quality	772	50.34	28.76	369	51.43	27.75	0.31
School_Math_Quality	771	52.31	18.71	369	53.74	18.22	0.04

Notes : The table reports the number of observations, the mean values and the standard deviations for the main school characteristics for the whole sample of schools in Madrid, columns (1) to (3) and for the sample of participating schools in the contest, columns (4) to (6). The final column reports the *p* -value for the F-Test of equality of variable means across the two samples. Public, Mixed and Private take the value of 1 when the school is public, with mixed funding, or privately owned, and 0 otherwise. Location variables take the value of 1 when the school is located in that particular area and 0 otherwise. *Size* reports the number of students in the 6th year of primary school (11 years old) and in the third year of secondary school (14 years old). *School_Overall_Quality* reports the centile at the normalized ranking in the multiple CDI tests. *School_Math_Quality* reports the centile at the normalized performance in the Math CDI test. See footnote 4 in the paper for more information on the CDI.

Table A.2. Self-reported Math Grade and Math Grade Reported by the Teacher

	Diff(Self-Teacher)	Diff(Self-Teacher)	Abs.Diff(Self-Teacher)	Abs.Diff(Self-Teacher)
	(1)	(2)	(3)	(4)
Female	0.0506	0.0578	-0.00745	-0.0462
	(0.0352)	(0.0451)	(0.0281)	(0.0331)
Constant	0.0265	0.0650	0.509***	0.569***
	(0.0202)	(0.0615)	(0.0162)	(0.0414)
Level FE	No	Yes	No	Yes
School FE	No	Yes	No	Yes
Clustered S.E. at school level	No	Yes	No	Yes
Observations	2,554	2,554	2,554	2,554
R-squared	0.001	0.229	0.000	0.294

Notes : The dependent values measures the difference/absolute difference between the self reported math grade and the math grade reported by the teacher in columns (1),(2), and (3),(4), respectively. *Female* is a dummy variable that takes vlaue 1 if the participant is a female, and 0 otherwise. Standard errors are reported in parentheses, with ***p<0.01, ** p<0.05, * p<0.1

Table A.3. Testing for Changes in the Composition of Male and Female Students

	Stage 2 performance (1)	Stage 2 performance (2)	Stage 2 performance (3)	Stage 2 performance (4)
Female	-5.457***	-3.403**	-5.440***	-10.27
	(0.805)	(1.709)	(0.806)	(7.502)
Stage 1 Dummy	0.112	2.783	0.0883	0.270
	(2.063)	(3.674)	(2.061)	(2.303)
Female*Stage 1 Dummy			2.439	-0.411
			(4.165)	(4.325)
Math Dummy		-2.528		-0.482
		(2.057)		(1.725)
Female*Math Dummy				5.204
				(7.484)
Observations	2,792	2,792	2,792	2,792
R-squared	0.533	0.374	0.533	0.533

Notes : The dependent variables refer to the performance in stage 2. All regressions include level and school fixed effects. *Female* is a dummy variable that takes vlaue 1 if the participant is a female, and 0 otherwise.Standard errors, clustered at school level, are reported in parentheses, with *** p<0.01, ** p<0.05, * p<0.1

Table A.4 Descriptive Statistics for the Overall Sample

		Overal	Male		Female		
Overall:	Obs.	Mean (St. Dev)	Obs. (Percent)	Mean (St. Dev)	Obs. (Percent)	Mean (St. Dev)	*p*-value
Performance Data:							
Math at School (0-10)	13925	7.15 (1.88)	8006 (57%)	7.12 (1.91)	5915 (43%)	7.20 (1.83)	0.02
Performance in Stage 1 (0-125)	20345	40.49 (17.46)	11445 (56%)	41.93 (18.17)	8825 (44%)	38.67 (16.32)	0.00
No. Of Omitted (0-25)		8.03 (5.49)		7.66 (5.53)		8.50 (5.39)	0.00
No. Of Right (0-25)		6.49 (3.84)		6.86 (3.99)		6.03 (3.57)	0.00
No. Of Wrong (0-25)		9.51 (4.90)		9.51 (4.95)		9.53 (4.84)	0.83
Performance in Stage 2 (0-125)	2792	49.63 (21.34)	1851 (66%)	52.08 (21.50)	941 (34%)	44.81 (20.20)	0.00
No. Of Omitted (0-25)		8.45 (5.29)		7.94 (5.27)		9.45 (5.18)	0.00
No. Of Right (0-25)		8.23 (4.81)		8.83 (4.84)		7.07 (5.54)	0.00
No. Of Wrong (0-25)		8.31 (4.57)		8.23 (4.58)		8.48 (4.54)	0.17
Winners	146	0.05	127 (87%)	0.07	19 (13%)	0.02	0.00
Level 1:							
Math at School (0-10)	2856	7.69 (1.62)	1608 (56%)	7.72 (1.63)	1248 (44%)	7.65 (1.60)	0.27
Performance in Stage 1 (0-125)	5123	48.10 (18.63)	2767 (54%)	50.07 (19.32)	2345 (46%)	45.79 (17.52)	0.00
No. Of Omitted (0-25)		5.57 (4.49)		5.19 (4.42)		6.02 (4.58)	0.00
No. Of Right (0-25)		8.51 (4.05)		8.98 (4.19)		7.95 (3.80)	0.00
No. Of Wrong (0-25)		10.56 (4.48)		10.46 (4.46)		10.67 (4.52)	0.09
Performance in Stage 2 (0-125)	655	68.80 (20.36)	429 (65%)	70.17 (20.12)	226 (35%)	66.19 (20.61)	0.02
No. Of Omitted (0-25)		5.00 (3.87)		4.58 (3.78)		5.81 (3.92)	0.00
No. Of Right (0-25)		12.76 (4.46)		13.12 (4.40)		12.08 (4.49)	0.00
No. Of Wrong (0-25)		7.24 (3.93)		7.31 (3.87)		7.12 (4.05)	0.56
Win Prize (0-1)	36	0.05	26 (72%)	0.06	10 (28%)	0.04	0.38
Level 2:							
Math at School (0-10)	5924	7.16 (1.84)	3272 (55%)	7.14 (1.86)	2649 (45%)	7.18 (1.83)	0.34
Performance in Stage 1 (0-125)	8847	39.62 (17.28)	4876 (55%)	40.91 (18.21)	3930 (45%)	38.07 (15.94)	0.00
No. Of Omitted (0-25)		7.94 (5.39)		7.41 (5.38)		8.58 (5.33)	0.00
No. Of Right (0-25)		6.34 (3.67)		6.70 (3.88)		5.90 (3.35)	0.00
No. Of Wrong (0-25)		9.49 (5.05)		9.61 (5.09)		9.36 (4.99)	0.02
Performance in Stage 2 (0-125)	910	47.69 (19.26)	584 (64%)	50.95 (19.86)	326 (36%)	41.86 (16.66)	0.00
No. Of Omitted (0-25)		7.77 (4.50)		7.16 (4.39)		8.86 (4.49)	0.00
No. Of Right (0-25)		7.98 (4.51)		8.76 (4.32)		6.60 (3.54)	0.00
No. Of Wrong (0-25)		9.25 (4.51)		9.08 (4.43)		9.54 (4.63)	0.15
Win Prize (0-1)	44	0.05	38 (86%)	0.07	6 (14%)	0.02	0.00
Level 3:							
Math at School (0-10)	3720	6.74 (2.00)	2185 (59%)	6.67 (2.02)	1534 (41%)	6.85 (1.96)	0.01
Performance in Stage 1 (0-125)	4844	35.59 (14.39)	2796 (58%)	37.16 (15.04)	2032 (42%)	33.45 (13.13)	0.00
No. Of Omitted (0-25)		9.61 (5.42)		9.17 (5.54)		10.21 (5.14)	0.00
No. Of Right (0-25)		5.19 (3.11)		5.60 (3.26)		4.65 (2.79)	0.00
No. Of Wrong (0-25)		9.03 (4.86)		9.11 (4.99)		8.93 (4.67)	0.21
Performance in Stage 2 (0-125)	784	39.44 (16.05)	518 (66%)	41.97 (17.14)	266 (34%)	34.52 (12.27)	0.00
No. Of Omitted (0-25)		10.41 (5.24)		9.79 (5.30)		11.62 (4.91)	0.00
No. Of Right (0-25)		5.81 (3.54)		6.44 (3.75)		4.58 (2.73)	0.00
No. Of Wrong (0-25)		8.78 (4.82)		8.77 (5.01)		8.80 (4.44)	0.93
Win Prize (0-1)	41	0.05	38 (93%)	0.07	3 (7%)	0.01	0.00
Level 4:							
Math at School (0-10)	1425	7.12 (1.93)	941 (66%)	7.08 (1.99)	484 (34%)	7.21 (1.81)	0.24
Performance in Stage 1 (0-125)	1531	35.64 (14.58)	1006 (66%)	37.76 (15.15)	518 (34%)	31.59 (12.53)	0.00
No. Of Omitted (0-25)		11.81 (5.49)		11.45 (5.55)		12.44 (5.32)	0.00
No. Of Right (0-25)		4.77 (3.23)		5.26 (3.39)		3.83 (2.68)	0.00
No. Of Wrong (0-25)		7.68 (4.72)		7.55 (4.75)		7.94 (4.67)	0.13
Performance in Stage 2 (0-125)	443	43.28 (16.35)	320 (72%)	46.24 (17.25)	123 (28%)	35.59 (10.39)	0.00
No. Of Omitted (0-25)		11.48 (5.44)		10.90 (5.53)		12.99 (4.89)	0.00
No. Of Right (0-25)		6.36 (3.75)		7.07 (3.97)		4.52 (2.22)	0.00
No. Of Wrong (0-25)		7.16 (4.56)		7.03 (4.55)		7.49 (4.58)	0.35
Win Prize (0-1)	25	0.06	25 (100%)	0.08	0 (0%)	0.00	0.00

Notes: This table reports the number of observations and the mean values for the main performance variables: *Performance* or *Score*, *No. Of Omitted*, *Right* and *Wrong* questions, as well as a dummy variable that takes the value of 1 if the participant wins the competition. The *p*-value are for the F-Test of equality of variable means across gender.

Table A.5 Gender Differential in Performance to Competitive Pressure: Alternative Samples and Specifications

	Overall Sample		Balanced Sample		Matched Sample using Stage 1 Score		Matched Sample using Math at School	
	RE	FE	RE	FE	RE	FE	RE	FE
	Performance	Performance	Performance	Performance	Performance	Performance	Performance	Performance
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Female	-0.270***		-0.229***		-0.00925		-0.212***	
	(0.0153)		(0.0417)		(0.0503)		(0.0504)	
Stage 2	-0.712***	-1.486***	0.0283	0.0283	0.107**	0.107**	0.0836*	0.0836*
	(0.0247)	(0.0276)	(0.0276)	(0.0264)	(0.0448)	(0.0418)	(0.0485)	(0.0454)
Female*Stage 2	-0.0982***	-0.0880*	-0.0920*	-0.0920*	-0.219***	-0.219***	-0.156**	-0.156**
	(0.0372)	(0.0501)	(0.0502)	(0.0480)	(0.0656)	(0.0611)	(0.0657)	(0.0616)
Math at School	0.222***		0.182***		0.170***		0.175***	
	(0.00465)		(0.0141)		(0.0172)		(0.0213)	
Observations	15,529	15,529	3,606	3,606	2,306	2,306	2,012	2,012
R-squared		0.705		0.002		0.011		0.006
Number of Participants	13,726	13,726	1,803	1,803	1,153	1,153	1,006	1,006

Notes: Dependent variables measure standardized performance or score at each stage and competition level. *Female* takes the value of 1 if the participant is female and 0 otherwise. *Stage 2* takes the value of 1 if the score refers to the second stage and 0 otherwise and *Math at School* measures the school grade in Math. Columns 1-2 include the overall sample, columns 3-4 include the balanced sample, and columns 5-6 include the matched sample. Columns 1, 3, 5 and 7 show random effects model, columns 2, 4, 6 and 8 show the individual fixed effects model. All regressions include school fixed effects. Standard errors, clustered at participant level, are reported in parentheses with *** p<0.01, ** p<0.05, * p<0.1.