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

Nagore Iriberri
Pedro Rey-Biel

This version July 2016
(February 2016)

Barcelona GSE Working Paper Series

Working Paper nº 879

Competitive Pressure Widens the Gender Gap in Performance:

Evidence from a Two-Stage Competition in Mathematics*

Nagore Iriberri+ and Pedro Rey-Biel**

July, 2016

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 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 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) Ministerio de Ciencia e Innovación (ECO2011-25295), 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 UPV-EHU, IKERBASQUE, Basque Foundation for Science. E-mail nagore.iriberri@upf.edu.

**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. For instance the gender wage gap, although it has shown a decreasing trend over time, still persists in developed countries. 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. Gneezy, Niederle, and Rustichini (2003) show that although there are no performance differences between women and men under piece-rate incentives women underperform compared to men under competitive incentive schemes. Niederle and Vesterlund (2007) further argue that women have a stronger preference for non-competitive settings, showing that women shy away from competition. In these two seminal papers, and in the extensive follow-up literature, the effects are attributed to reward schemes being competitive or not.

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. Examples of such vertical hierarchies that resemble multi-stage contests abound in labor markets. In the academic world, assistant professors compete for associate professor positions and associate professors compete for full professor positions. In the legal profession contract lawyers compete to become associate lawyers who then compete to become partners. Many companies offer some type of hierarchical structure, with a large base of workers who are directly supervised by a smaller level above them, which is in turn supervised by other levels above it, all the way up to the top ranking officer such as the company President or CEO. Scaling up within company levels is often possible through internal promotion, which consists of some type of contest. Furthermore, the glass-ceiling effect shows that the presence of women significantly decreases as one climbs up the hierarchy. For example women make up the majority of students in American colleges and universities (59% of graduate enrollment), but 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. 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 average ability of participants 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 knocked out. Second, proximity to the highly rewarded prize/position also increases. Moreover, having already qualified in previous stages may increase the pressure. This paper addresses the core question of whether men and women react differently to increases in competitive pressure in a multi-stage elimination contest.

Regional two-stage contests in mathematics for students aged between 10 and 16 offer a unique opportunity to study gender differences in performance as competitive pressure increases. We use data from the 2014 edition of *Concurso de Primavera de Matemáticas*, a regional math competition in Madrid (Spain) where students compete in four independent contests separated by age groups, which we refer to as levels, in which they must complete a 90-minute 25 multiple choice question math test in each of two stages. About 40,000 students compete in the first stage, and only about 2,800 make it to the second stage, 146 of whom get to be recognized as best performers. 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.

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 figures, graphed in Figure 1, clearly show that female representation decreases as we move up through the stages of the contest, evidencing a clear glass-ceiling effect.

[Figure 1 Here]

The dramatic decrease in the presence of women as we move up through the elimination contest 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, we further show that there is *another* cause of the decrease in the female presence as we move up in the elimination contest. In particular, we show that for the set of participants whose performance can be observed in multiple stages, the balanced sample, female participants have a comparable grade in Math at school than male participants; 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. This gap widens up to 7.3 test points when they move to the second and final stage of the contest, 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. The widening of the gender gap in performance as competitive pressure increases is due to an *increase* in the gender gap for the number of omitted questions, with women failing to answer more questions, as well as an *increase* in the gender gap for the number of correct answers, with women providing fewer. This result is robust to alternative specifications and to alternative samples.

We also perform a heterogeneity analysis. We use the variation in age-levels, in academic years within each age-level, and in ability, to test whether gender differences vary with those characteristics as competitive pressure increases. We find that the gender differential in reaction to increases in competitive pressure is stronger among high ability participants than low ability participants. We also exploit the variation in school characteristics. 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 able to rule out three alternative explanations for the increase in the gender gap we observe. First, the idea that male and female participants may differ in their characteristics, such as the type of school they attend, which would pose a selection problem. We indeed find that in the balanced sample of participants, female participants are more likely to come from smaller schools and schools with lower quality in mathematics, although the differences are not large. However, once we control for these characteristics our main result is robust and similar in both magnitude and significance. Furthermore, in the sample where we match male and female participants with the same scores in stage 1, where no significant differences in school characteristics appear, our results are again robust. Second, male and female contestants may have different likelihoods of qualifying for the second stage or of being selected among the winners due to some type of discrimination. Once their performance in stage 1 is controlled for, we find that male and female contestants have the same likelihood of being selected for and/or showing up in stage 2; and once performance in stage 2 is controlled for, male and female contestants have the same likelihood of being selected among the winners. Also, male and female participants do not have different likelihoods of dropping out of stage 2 once they are selected. Thus, the selection is gender neutral and the decrease in the presence of women in stage 2 and among the winners is purely due to gender differences in performance. Third, stage 1 and 2 also vary in difficulty, with stage 2 being more difficult. However, using the variation in difficulty at the question level within each stage, we show that there are no gender differential reactions to increasing difficulty in stage 1 and that gender differences are greater in relatively *easier* questions in stage 2. More importantly, we test for gender differences between stage 1 and stage 2 of the contest while controlling for difficulty and gender differences that depend on difficulty, and show that the increase in the gender gap as competitive pressure increases is robust and sound.

Identifying gender differential reactions to changes in the underlying incentive schemes is not easily done using field data. Three papers are closest to our work:

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. Finally, Azmat, Calsamiglia, and Iriberri (2014) 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. Our contribution 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.

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. Finally, section 5 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.[1] 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

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

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.[2]

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 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

[2] 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). For more information see: http://www.educa2.madrid.org/web/cdi/pruebas-cdi

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.[3] 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.[4]

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.[5] Teachers at the schools, who were able to download the stage 1 test only a few days before it was scheduled to place, used answer keys provided by the organizers to grade stage 1 tests. Those 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, performance in stage 2 determines the winners in the contest, so students are closer to the prize in stage 2 than in stage 1. Second, 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. Also, as shown by the number of competitors and the winning positions, the

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

[4] 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 i-pads, 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.

[5] We ran a survey at school level to gather information on how the stage 1 contest was carried out at schools. Only 4% of schools said that they used criteria other than the stage 1 test in order to select their students to participate in stage 2, so the vast majority of schools do indeed use the stage 1 test to select their students. 56% of the rest said that participation is open to all students who voluntarily want to participate, 21% said that all students participated, 19% said participation was restricted to best performing students who volunteered to participate, and 3% said participation was restricted to only the best performing students. Our main analysis, shown in Table 2, compares gender differences among students who did both the stage 1 and stage 2 tests.

proportion of winning spots is lower in stage 2, making winning harder. These differences appear in all four levels.

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, shown in Figure 2, depict similar patterns for the overall distribution. It can further be observed 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. Table 3 shows the estimation results following an identification strategy based on differences-in-differences at the student level with random effects.

[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, level fixed effects, school characteristics (*Mixed* and *Private*, with *Public* being omitted; location dummies, with *Madrid* being omitted; *Size*, *School_Overall_Quality* and *School_Math_Quality*) and clustered standard errors at school level. 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, which includes level and school fixed effects as well as clustered standard errors at school level.

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 4.9 test points. Moreover, performance in stage 2 is lower than performance in stage 1 (see the analysis on difficulty in Section 4). More importantly, girls underperform more (the gap is more negative) in the second stage than in the first by about 2.4 points, as shown by the coefficient of the interaction term between *Female* and *Stage 2*. This represents an increase of almost 50% on the gender gap in stage 1, showing a gender differential reaction to increases in competitive pressure. 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.[6] Estimation results in column 2 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 2 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

[6] 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, that of the interaction between *Female* and *Stage 2* is always negative, significant, and of similar magnitude as the one in column 3.

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. This suggests that even if the scoring rule did not reward omitted answers in comparison to wrong answers, an increase in female underperformance would still be observed as competitive pressure increases.

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 1-3) and the matched sample (7-9). The matched sample puts together male and female students with non-distinguishable stage 1 scores. We perform a regression analysis on this matched sample, with a total of 1,153 participants (576 boys and 575 girls) using probability score matching (see Rosenbaum and Rubin, 1985, and Caliendo and Kopeinig, 2008 for a practical guide). The increase in female underperformance as competitive pressure increases is always negative, significant and of a similar magnitude. The magnitude is even 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. Second, our results are robust to alternative specifications: we estimate individual fixed effects (columns 2, 5 and 8) and OLS models (columns 3, 6 and 9 in Table A.5) 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. 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.

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 0.8309).

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 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 0.769) 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 in stage 1 of the contest and 0 otherwise.[7] 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

[7] We have also used *Math at school* 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, although the coefficient is not significantly different from zero. 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.

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: 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 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, 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.

The estimation results in Table 6 rule out this 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 passing to stage 2, some 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 estimated negative coefficients of the *Stage 2* dummy in Table 3 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 histograms in Figure 3 show that there is variation in the level of difficulty in both the stage 1 and stage 2 tests.

Table 7 shows the estimation results for regressions with the dependent variable being the performance or score in a particular question in stage 1 in columns 1-3, and in stage 2 in columns 4-6. We construct two alternative controls for difficulty: the *Easy Dummy* takes the value of 1 if the question is among the top-half in Figure 3, 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.[8] From now on we will only use Mean Score variable to control for easiness.

More importantly, we can replicate our main results, shown in Table 3, with the score at question level as the dependent variable instead of the score at stage level, where we

[8] Additionally, as a robustness check we use the mean scores by female participants and by male participants instead of across gender. The results are replicated when using the mean score by male participants but we find that there is no differential reaction to difficulty by gender when we use the mean score by female participants.

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. Interestingly, once we control for the difficulty of the question, the estimated coefficient of *Stage 2* ceases to be significant in any of the columns in Table 8, showing that stage 1 and stage 2 no longer differ in performance. Although it emerges that, consistent with what can be observed in Table 7, female underperformance is higher in relatively easier questions (see footnote 8), gender differential reaction to increasing competitive pressure, the negative and significant interaction between *Female* and *Stage 2,* is 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. The fact that it is indeed in the relatively easier questions that women tend to omit less than men and are more likely to answer wrongly is hard to reconcile with women being more risk averse and less overconfident. This also suggests that the gender differences observed are not driven by differences in risk aversion and overconfidence.

5 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 and are expected to perform better than women (Fryer and Levitt, 2010, Bharadwaj et al., forthcoming, Nollenberger et al., 2016). 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. Second, the underlying mechanism determining why female participants react in this way to increasing competitive pressure remains to be explored. Do they prepare less than men when pressure increases? Are there gender differences in inherent motivation or in the encouragement they receive from their parents and teachers? 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".

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*.

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. "Gender Differences in Response to Big Stakes." Forthcoming in *Journal of the European Economic Association*.

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.

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

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.

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.

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.

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.

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

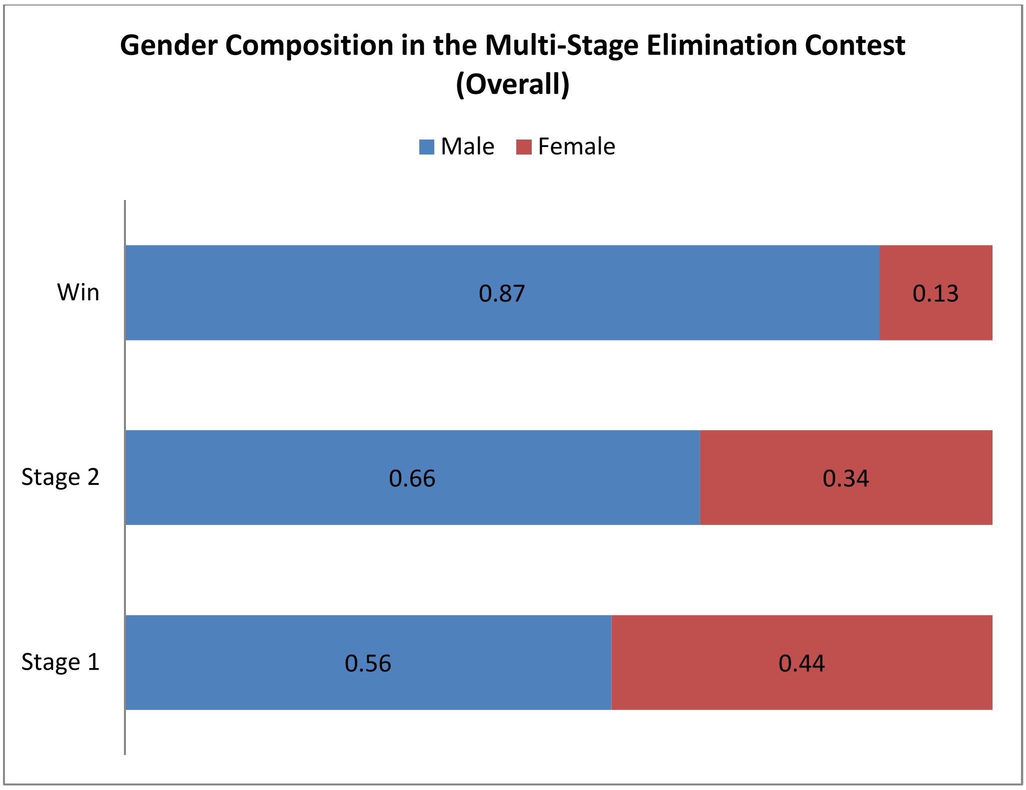
Swineford, F. 1941. “Analysis of a Personality Trait.” *Journal of Educational Psychology*, 45: 81–90.

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



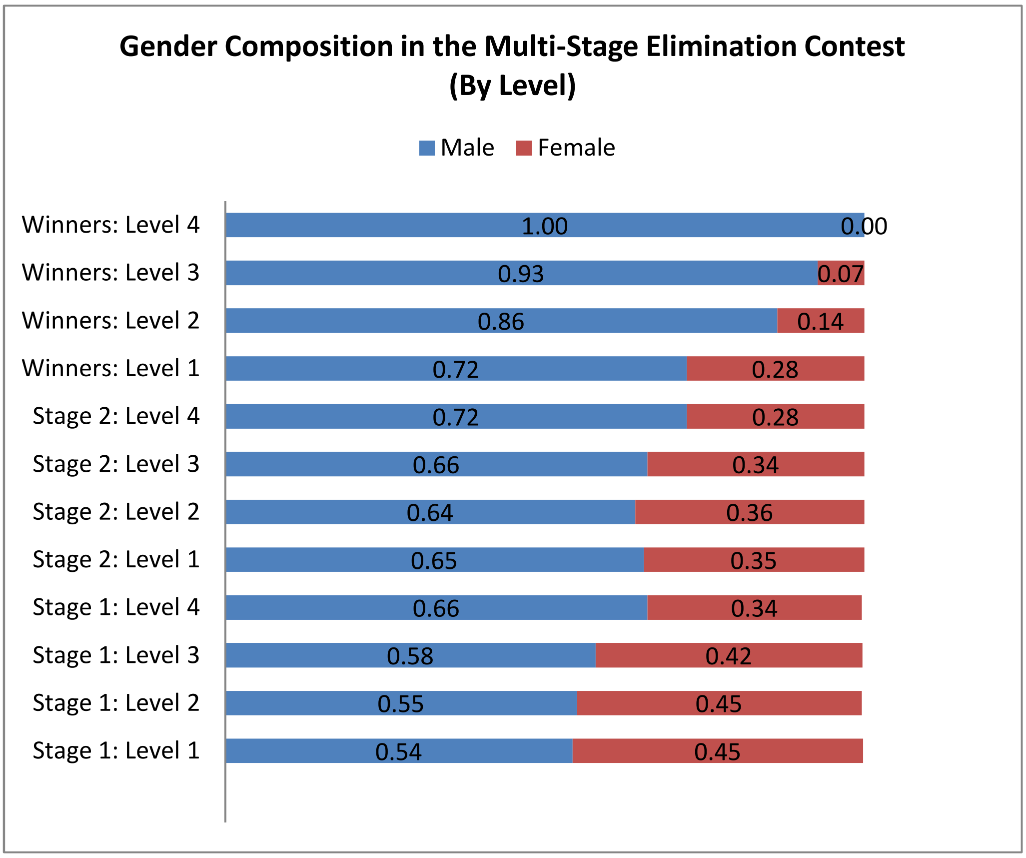


Figure 1. Gender Composition in the 2-stage Math Contest, Overall and by Level

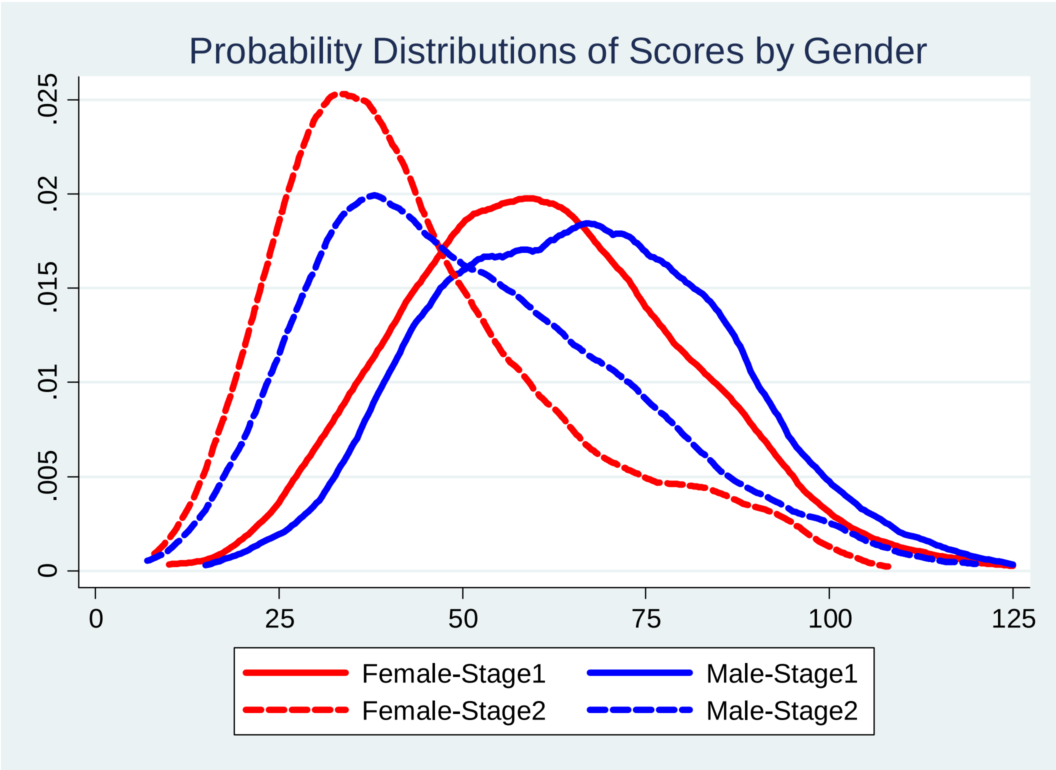


Figure 2. Probability Distributions of Performance by Stage and by Gender

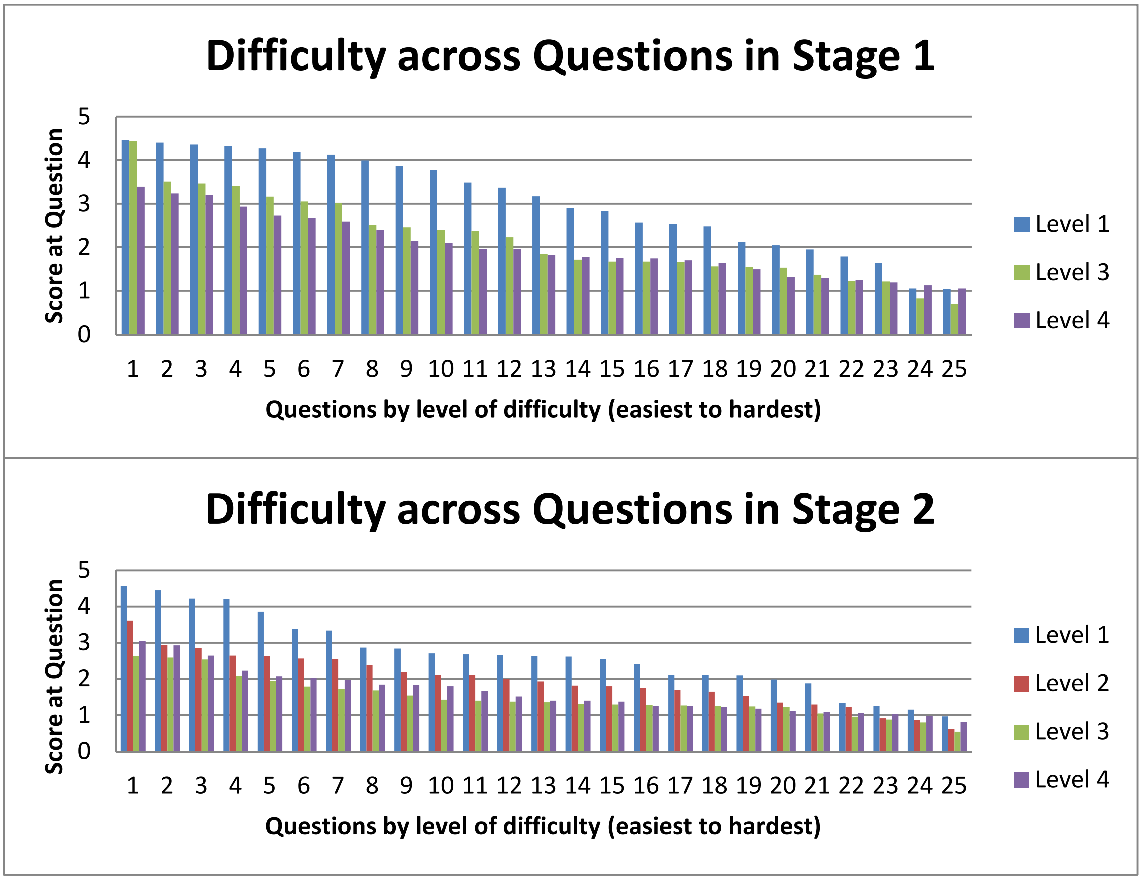


Figure 3. Difficulty of Questions in Stages 1 and 2 of the Contest

Table 1. Competitive Pressure in Stages 1 and 2

		Stage 1			Stage 2		
		Obs.	Mean	St. Dev.	Obs.	Mean	St. Dev.
		(1)	(2)	(3)	(4)	(5)	(6)
Level 1	Performance in Stage 1	443	54.78	14.96	443	77.67	15.17
	Math at School	440	8.15	0.82	434	8.77	0.76
	No. Of Competitors	443	40.62	35.21	443	443	
	No. Of Winning Positions	443	3.92	1.05	443	15	
	Prop. Of Winning Positions	443	0.27	0.32	443	0.03	
Level 2	Performance in Stage 1	593	44.62	13.53	593	69.24	18.84
	Math at School	593	7.62	0.98	592	8.62	1.10
	No. Of Competitors	593	39.42	44.06	593	910	
	No. Of Winning Positions	593	2.79	0.48	593	32	
	Prop. Of Winning Positions	593	0.20	0.26	593	0.05	
Level 3	Performance in Stage 1	508	39.26	10.13	508	55.16	16.48
	Math at School	508	7.26	1.08	508	7.99	1.54
	No. Of Competitors	508	23.98	30.13	508	784	
	No. Of Winning Positions	508	2.72	0.53	508	22	
	Prop. Of Winning Positions	508	0.28	0.26	508	0.04	
Level 4	Performance in Stage 1	269	40.07	11.13	269	51.01	15.36
	Math at School	269	7.42	1.17	269	7.92	1.57
	No. Of Competitors	269	11.94	11.94	269	443	
	No. Of Winning Positions	269	2.59	0.64	269	17	
	Prop. Of Winning Positions	269	0.43	0.32	269	0.06	

Notes: The table reports the number of observations, the mean values and the standard deviations for the main competitive pressure measures for stage 1, columns (1) to (3) and for stage 2, columns (4) to (6). These numbers 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	Obs.	Mean	Obs.	Mean	*p* -value
Performance Data:							
Math at School (0-10)	1803	8.36	1229 (68%)	8.33	574 (32%)	8.42	0.17
Performance in Stage 1 (0-125)	1813	64.65	1232 (68%)	66.22	581 (32%)	61.32	0.00
No. Of Omitted (0-25)		6.62		6.32		7.26	0.00
No. Of Right (0-25)		11.61		11.98		10.81	0.00
No. Of Wrong (0-25)		5.87		5.83		5.95	0.49
Performance in Stage 2 (0-125)	1813	49.32	1232 (68%)	51.66	581 (32%)	44.37	0.00
No. Of Omitted (0-25)		8.39		7.85		9.55	0.00
No. Of Right (0-25)		8.19		8.76		6.96	0.00
No. Of Wrong (0-25)		8.42		8.39		8.48	0.69
Winners	85	0.05	65 (88%)	0.06	10 (12%)	0.02	0.00
Level 1:							
Math at School (0-10)	434	8.77	297 (68%)	8.74	137 (32%)	8.85	0.14
Performance in Stage 1 (0-125)	443	77.67	300 (68%)	78.51	143 (32%)	75.92	0.09
No. Of Omitted (0-25)		3.62		3.24		4.41	0.00
No. Of Right (0-25)		14.81		15.05		14.30	0.03
No. Of Wrong (0-25)		6.27		6.39		6.02	0.22
Performance in Stage 2 (0-125)	443	67.38	300 (68%)	68.98	143 (32%)	64.04	0.01
No. Of Omitted (0-25)		5.07		4.68		5.90	0.00
No. Of Right (0-25)		12.46		12.86		11.63	0.00
No. Of Wrong (0-25)		7.47		7.46		7.48	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	386 (65%)	8.59	206 (35%)	8.58	0.94
Performance in Stage 1 (0-125)	593	69.24	386 (65%)	71.63	207 (35%)	64.77	0.00
No. Of Omitted (0-25)		5.72		5.38		6.37	0.00
No. Of Right (0-25)		12.70		13.25		11.68	0.00
No. Of Wrong (0-25)		5.42		5.32		5.60	0.35
Performance in Stage 2 (0-125)	593	47.21	386 (65%)	50.16	207 (35%)	41.71	0.00
No. Of Omitted (0-25)		7.77		7.16		8.91	0.00
No. Of Right (0-25)		7.89		8.60		6.56	0.00
No. Of Wrong (0-25)		9.34		9.24		9.53	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	344 (68%)	7.92	164 (32%)	8.13	0.14
Performance in Stage 1 (0-125)	508	55.16	344 (68%)	57.33	164 (32%)	50.62	0.00
No. Of Omitted (0-25)		8.22		7.76		9.18	0.00
No. Of Right (0-25)		9.39		9.91		8.29	0.00
No. Of Wrong (0-25)		6.19		6.10		6.38	0.42
Performance in Stage 2 (0-125)	508	39.28	344 (68%)	41.69	164 (32%)	34.22	0.00
No. Of Omitted (0-25)		10.36	344	9.60		11.96	0.00
No. Of Right (0-25)		5.78	344	6.42		4.45	0.00
No. Of Wrong (0-25)		8.85	344	8.98		8.59	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	202 (75%)	7.96	67 (75%)	7.77	0.38
Performance in Stage 1 (0-125)	269	51.01	202 (75%)	52.78	67 (75%)	45.67	0.00
No. Of Omitted (0-25)		10.52		10.23		11.42	0.11
No. Of Right (0-25)		8.10		8.51		6.85	0.00
No. Of Wrong (0-25)		5.58		5.50		5.82	0.56
Performance in Stage 2 (0-125)	269	43.20	202 (75%)	45.77	67 (75%)	35.45	0.00
No. Of Omitted (0-25)		11.52		10.89		13.43	0.00
No. Of Right (0-25)		6.33		6.98		4.40	0.00
No. Of Wrong (0-25)		7.14		7.13		7.16	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

	Performance	Performance	Performance	No. Of Omitted	No. Of Right	No. Of Wrong	Prop. Of Right	Prop. Of Right
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Female	-4.905***	-4.203***	-3.513***	0.648***	-0.832***	0.138	-0.0192**	-0.0249***
	(0.975)	(0.848)	(0.766)	(0.248)	(0.172)	(0.217)	(0.00914)	(0.00852)
Stage 2	-14.56***	-14.55***	-14.59***	1.528***	-3.223***	2.571***	-0.137***	-0.150***
	(0.481)	(0.654)	(0.672)	(0.139)	(0.142)	(0.140)	(0.00761)	(0.00758)
Female*Stage 2	-2.382***	-2.334**	-2.408**	0.772***	-0.636***	-0.0346	-0.0292**	-0.0358***
	(0.869)	(0.989)	(1.020)	(0.237)	(0.218)	(0.230)	(0.0118)	(0.0116)
Math at School		3.299***	3.192***	0.315***	0.575***	-0.874***	0.0414***	0.0387***
		(0.278)	(0.277)	(0.0862)	(0.0616)	(0.0782)	(0.00315)	(0.00309)
No. Of Omitted								0.00880***
								(0.000903)
Mixed		1.016						
		(1.231)						
Private		0.0329						
		(2.290)						
North		-0.190						
		(1.980)						
South		-4.562***						
		(1.240)						
East		-3.148**						
		(1.486)						
West		-0.884						
Size		0.0674***						
		(0.0140)						
School_Overall_Quality		0.0253						
		(0.0377)						
School_Math_Quality		0.210***						
		(0.0680)						
Constant	66.22***	36.35***	58.84***	1.030*	10.08***	10.74***	0.537***	0.421***
	(0.557)	(3.447)	(3.300)	(0.605)	(0.445)	(0.901)	(0.0366)	(0.0295)
Level FE	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School FE	No	No	Yes	Yes	Yes	Yes	Yes	Yes
School Characteristics	No	Yes	No	No	No	No	No	No
Clustered S.E.	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3,626	3,524	3,606	3,606	3,606	3,606	3,605	3,605
Number of Participants	1,813	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). *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 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 (1)	Performance Level 2 (2)	Performance Level 3 (3)	Performance Level 4 (4)	Performance Lower Ac. Year (5)	Performance Higher Ac. Year (6)	Performance Low Ability (7)	Performance High Ability (8)
Female	-0.632	-3.712***	-5.404***	-4.660**	-2.512**	-3.345***	-0.338	-1.326
	(1.560)	(1.427)	(1.677)	(2.108)	(1.196)	(1.190)	(0.839)	(1.107)
Stage 2	-9.579***	-21.47***	-15.63***	-7.010***	-15.22***	-14.08***	-9.098***	-19.18***
	(1.080)	(1.096)	(1.028)	(1.519)	(0.976)	(0.875)	(0.885)	(0.801)
Female*Stage 2	-2.180	-1.681	-0.766	-3.214	-2.726*	-2.167	-2.849**	-6.305***
	(2.346)	(1.784)	(1.774)	(2.430)	(1.504)	(1.404)	(1.275)	(1.599)
Math at School	5.772***	4.165***	2.744***	1.312**	2.997***	3.523***	2.102***	3.003***
	(0.939)	(0.550)	(0.515)	(0.557)	(0.525)	(0.420)	(0.269)	(0.447)
Constant	16.46*	36.22***	41.15***	45.39***	51.51***	54.07***	56.34***	66.96***
	(8.932)	(3.789)	(2.329)	(5.640)	(5.111)	(3.283)	(2.047)	(5.396)
Observations	868	1,184	1,016	538	1,592	2,014	1,840	1,766
No. of Participants	434	592	508	269	796	1,007	920	883

	Performance Size (9)	Performance School_Overall_Quality (10)	Performance School_Math_Quality (11)	Performance Public (12)	Performance Non-Public (13)	Performance Madrid (14)	Performance Outside Madrid (15)
Female	-0.997	-2.841	-1.814	-3.603***	-4.722***	-5.303***	-3.129***
	(1.718)	(1.814)	(2.997)	(1.148)	(1.312)	(1.281)	(1.126)
Stage 2	-14.70***	-13.54***	-12.71***	-13.57***	-15.68***	-15.67***	-13.70***
	(1.324)	(1.255)	(1.910)	(0.836)	(1.023)	(0.907)	(0.922)
Female*Stage 2	-1.064	-3.921**	-5.449*	-2.513**	-2.347	-3.005**	-1.986
	(2.217)	(1.971)	(3.285)	(1.152)	(1.749)	(1.394)	(1.374)
Math at school	3.288***	3.296***	3.297***	3.392***	3.096***	2.991***	3.462***
	(0.280)	(0.279)	(0.279)	(0.349)	(0.442)	(0.439)	(0.357)
Size	0.0819***	0.0674***	0.0673***	0.0519***	0.0818***	0.0928***	0.0473**
	(0.0177)	(0.0140)	(0.0140)	(0.0157)	(0.0278)	(0.0213)	(0.0190)
School_Overall_Quality	0.0277	0.0389	0.0252	0.00428	0.0463	0.0623	0.00284
	(0.0376)	(0.0419)	(0.0378)	(0.0517)	(0.0526)	(0.0547)	(0.0521)
School_Math_Quality	0.206***	0.210***	0.232***	0.264***	0.173*	0.132	0.275***
	(0.0675)	(0.0679)	(0.0718)	(0.0925)	(0.0997)	(0.103)	(0.0912)
Female*Size	-0.0392**						
	(0.0184)						
Stage 2*Size	0.00173						
	(0.0149)						
Female*Size*Stage 2	-0.0159						
	(0.0227)						
Female*School_Overall_Quality		-0.0278					
		(0.0329)					
Stage 2*School_Overall_Quality		-0.0194					
		(0.0234)					
Female*School_Overall_Quality*Stage 2		0.0319					
		(0.0363)					
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)				
Constant	35.22***	35.68***	35.17***	32.75***	44.95***	39.33***	31.59***
	(3.542)	(3.578)	(3.757)	(4.389)	(10.73)	(5.060)	(4.490)
Observations	3,524	3,524	3,524	1,950	1,488	1,492	2,032
Number of Students	1,762	1,762	1,762	975	744	746	1,016

Notes: Dependent variables measure performance or score. *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 level and school fixed effects. Regressions in the bottom panel include level fixed effects and school characteristics. Standard errors, clustered at school 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			Male Participants			Female Participants			
School Characteristics	Obs.	Mean	St. Dev.	Obs.	Mean	St. Dev.	Obs.	Mean	St. Dev.	*p* -value
Public	1772	0.55	0.50	1203	0.54	0.50	569	0.59	0.49	0.04
Mixed	1772	0.34	0.47	1203	0.36	0.48	569	0.31	0.46	0.05
Private	1772	0.10	0.31	1203	0.11	0.31	569	0.10	0.30	0.73
Madrid	1772	0.43	0.49	1203	0.43	0.50	569	0.41	0.49	0.30
North	1772	0.09	0.28	1203	0.09	0.28	569	0.09	0.29	0.83
South	1772	0.20	0.40	1203	0.20	0.40	569	0.21	0.41	0.49
East	1772	0.12	0.32	1203	0.12	0.32	569	0.11	0.32	0.73
West	1772	0.17	0.37	1203	0.16	0.37	569	0.18	0.38	0.44
Size	1772	83.65	40.82	1203	85.95	40.39	569	78.78	41.33	0.00
School_Overall_Quality	1772	50.27	27.21	1203	52.03	27.29	569	46.54	26.67	0.00
School_Math_Quality	1772	53.27	16.18	1203	54.27	16.22	569	51.15	15.89	0.00

Overall Sample										
	Overall			Male Participants			Female Participants			
School Characteristics	Obs.	Mean	St. Dev.	Obs.	Mean	St. Dev.	Obs.	Mean	St. Dev.	*p* -value
Public	21538	0.47	0.50	12149	0.48	0.50	9313	0.46	0.50	0.00
Mixed	21538	0.40	0.49	12149	0.40	0.49	9313	0.42	0.49	0.00
Private	21538	0.12	0.33	12149	0.12	0.33	9313	0.13	0.33	0.19
Madrid	21538	0.44	0.50	12149	0.44	0.50	9313	0.43	0.50	0.40
North	21538	0.13	0.34	12149	0.12	0.33	9313	0.14	0.35	0.00
South	21538	0.15	0.35	12149	0.15	0.36	9313	0.13	0.34	0.00
East	21538	0.12	0.33	12149	0.12	0.33	9313	0.12	0.33	0.99
West	21538	0.16	0.37	12149	0.16	0.37	9313	0.17	0.37	0.11
Size	21522	85.19	37.69	12140	87.29	37.87	9310	82.49	37.26	0.00
School_Overall_Quality	21522	55.38	26.85	12140	55.25	26.99	9310	55.58	26.63	0.37
School_Math_Quality	21522	56.63	16.61	12140	56.60	16.59	9310	56.69	16.57	0.68

Participants in Stage 2										
	Overall			Male Participants			Female Participants			
School Characteristics	Obs.	Mean	St. Dev.	Obs.	Mean	St. Dev.	Obs.	Mean	St. Dev.	*p* -value
Public	2689	0.49	0.50	1791	0.47	0.50	898	0.51	0.50	0.05
Mixed	2689	0.37	0.48	1791	0.38	0.49	898	0.36	0.48	0.25
Private	2689	0.14	0.35	1791	0.15	0.35	898	0.13	0.33	0.21
Madrid	2689	0.43	0.50	1791	0.44	0.50	898	0.41	0.49	0.18
North	2689	0.10	0.30	1791	0.10	0.29	898	0.10	0.30	0.73
South	2689	0.20	0.40	1791	0.19	0.39	898	0.21	0.41	0.27
East	2689	0.11	0.31	1791	0.10	0.31	898	0.11	0.31	0.74
West	2689	0.17	0.37	1791	0.17	0.37	898	0.17	0.37	0.94
Size	2686	83.72	39.85	1788	86.11	40.14	898	78.95	38.86	0.00
School_Overall_Quality	2686	54.42	27.33	1788	55.87	27.32	898	51.53	27.12	0.00
School_Math_Quality	2686	55.95	16.88	1788	56.95	16.96	898	53.96	16.53	0.00

Matched Sample										
	Overall			Male Participants			Female Participants			
School Characteristics	Obs.	Mean	St. Dev.	Obs.	Mean	St. Dev.	Obs.	Mean	St. Dev.	*p* -value
Mixed	1140	0.31	0.46	571	0.32	0.47	569	0.31	0.46	0.78
Private	1140	0.10	0.30	571	0.09	0.29	569	0.10	0.30	0.75
Madrid	1140	0.41	0.49	571	0.42	0.49	569	0.41	0.49	0.62
North	1140	0.09	0.29	571	0.09	0.29	569	0.09	0.29	0.99
South	1140	0.21	0.41	571	0.21	0.41	569	0.21	0.41	0.92
East	1140	0.12	0.32	571	0.12	0.32	569	0.11	0.32	0.73
West	1140	0.17	0.37	571	0.16	0.36	569	0.18	0.38	0.41
Size	1140	81.93	41.28	571	85.07	41.02	569	78.78	41.31	0.01
School_Overall_Quality	1140	47.70	27.11	571	48.85	27.53	569	46.54	26.66	0.15
School_Math_Quality	1140	51.71	16.07	571	52.27	16.25	569	51.15	15.88	0.24

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.	Stage 1 Perf.	Stage 1 Perf.	Stage 2 Perf.	Stage 2 Perf.	Stage 2 Perf.
	(1)	(2)	(3)	(4)	(5)	(6)
Female	-0.113***	-0.125***	-0.125**	-0.253***	-0.170***	-0.0561
	(0.0256)	(0.0336)	(0.0593)	(0.0333)	(0.0355)	(0.0517)
Math at School	0.0954***	0.0954***	0.0954***	0.178***	0.177***	0.177***
	(0.0107)	(0.0107)	(0.0107)	(0.0129)	(0.0129)	(0.0129)
Easy Dummy		1.603***			1.221***	
		(0.0286)			(0.0259)	
Female*Easy Dummy		0.0234			-0.163***	
		(0.0470)			(0.0426)	
Mean Score			1.000***			1.034***
			(0.0146)			(0.0151)
Female*Mean Score			0.00458			-0.0961***
			(0.0209)			(0.0251)
Constant	2.370***	1.537***	-0.699***	1.354***	0.719***	-1.395***
	(0.117)	(0.118)	(0.121)	(0.147)	(0.148)	(0.153)
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 performance or score at the 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 level and school fixed effects. Standard errors, clustered at the school level, in parentheses *** p<0.01,

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

	Performance by Question	Performance by Question	Omitted by Question	Omitted by Question	Correct by Question	Correct by Question
	(1)	(2)	(3)	(4)	(5)	(6)
Female	-0.137***	-0.0404	0.0262***	0.0496***	-0.0175**	-0.0261*
	(0.0293)	(0.0499)	(0.00980)	(0.0179)	(0.00864)	(0.0144)
Stage 2	-0.578***	0.0283	0.0585***	-0.00685	-0.131***	0.00786
	(0.0280)	(0.0262)	(0.00569)	(0.00609)	(0.00758)	(0.00685)
Female*Stage 2	-0.0875**	-0.0859**	0.0272***	0.0188**	-0.0321***	-0.0235**
	(0.0402)	(0.0396)	(0.00913)	(0.00960)	(0.0113)	(0.0108)
Math at school	0.129***	0.129***	0.0113***	0.0113***	0.0415***	0.0390***
	(0.0108)	(0.0108)	(0.00333)	(0.00333)	(0.00306)	(0.00291)
Mean Score		1.016***		-0.109***		0.208***
		(0.0118)		(0.00317)		(0.00265)
Female*Mean Score		-0.0431**		-0.00850*		0.000409
		(0.0185)		(0.00465)		(0.00444)
Constant	2.265***	-0.934***	0.0152	0.361***	0.406***	-0.261***
	(0.123)	(0.126)	(0.0345)	(0.0366)	(0.0341)	(0.0331)
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 the performance or score at the 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. *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 level and school fixed effects. Standard errors, clustered at the school 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 2 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)	(4)	(5)
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.	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)
Constant	72.77***	49.49***	70.38***	73.09***
	(1.786)	(2.397)	(4.497)	(4.653)
Level FE	Yes	Yes	Yes	Yes
School FE	Yes	Yes	Yes	Yes
Clustered S.E. School Level	Yes	Yes	Yes	Yes
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 Whole Sample

	Overal		Male		Female		
Overall:	Obs.	Mean	Obs.	Mean	Obs.	Female	*p* -value
Performance Data:							
Math at School (0-10)	14117	7.13	8092 (57%)	7.10	6021 (43%)	7.17	0.03
Performance in Stage 1 (0-125)	20751	40.27	11638 (56%)	41.72	9038 (44%)	38.44	0.00
No. Of Omitted (0-25)		8.09		7.69		8.58	0.00
No. Of Right (0-25)		6.44		6.80		5.97	0.00
No. Of Wrong (0-25)		9.53		9.54		9.53	0.86
Performance in Stage 2 (0-125)	2792	49.63	1851 (66%)	52.08	941 (34%)	44.81	0.00
No. Of Omitted (0-25)		8.45		7.94		9.45	0.00
No. Of Right (0-25)		8.24		8.83		7.07	0.00
No. Of Wrong (0-25)		8.31		8.23		8.48	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	1608 (56%)	7.72	1248 (44%)	7.65	0.27
Performance in Stage 1 (0-125)	5123	48.10	2767 (54%)	50.07	2345 (46%)	45.79	0.00
No. Of Omitted (0-25)		5.57		5.19		6.02	0.00
No. Of Right (0-25)		8.51		8.98		7.95	0.00
No. Of Wrong (0-25)		10.56		10.46		10.67	0.09
Performance in Stage 2 (0-125)	655	68.80	429 (65%)	70.17	226 (35%)	66.19	0.02
No. Of Omitted (0-25)		5.00		4.58		5.81	0.00
No. Of Right (0-25)		12.76		13.12		12.08	0.00
No. Of Wrong (0-25)		7.24		7.31		7.12	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)	6116	7.10	3358 (55%)	7.09	2755 (45%)	7.12	0.52
Performance in Stage 1 (0-125)	9253	39.15	5069 (55%)	40.46	4143 (36%)	37.59	0.00
No. Of Omitted (0-25)		8.07		7.50		8.76	0.00
No. Of Right (0-25)		6.22		6.59		5.77	0.00
No. Of Wrong (0-25)		9.53		9.67		9.37	0.00
Performance in Stage 2 (0-125)	910	47.69	584 (64%)	50.95	326 (36%)	41.86	0.00
No. Of Omitted (0-25)		7.77		7.16		8.87	0.00
No. Of Right (0-25)		7.98		8.76		6.60	0.00
No. Of Wrong (0-25)		9.25		9.08		9.54	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	2185 (59%)	6.67	1534 (41%)	6.85	0.01
Performance in Stage 1 (0-125)	4844	35.59	2796 (58%)	37.16	2032 (42%)	33.45	0.00
No. Of Omitted (0-25)		9.61		9.17		10.21	0.00
No. Of Right (0-25)		5.19		5.60		4.65	0.00
No. Of Wrong (0-25)		9.03		9.11		8.93	0.21
Performance in Stage 2 (0-125)	784	39.44	518 (66%)	41.97	266 (34%)	34.52	0.00
No. Of Omitted (0-25)		10.41		9.79		11.62	0.00
No. Of Right (0-25)		5.81		6.44		4.58	0.00
No. Of Wrong (0-25)		8.78		8.77		8.80	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	941 (66%)	7.08	484 (34%)	7.21	0.24
Performance in Stage 1 (0-125)	1531	35.64	1006 (66%)	37.76	518 (34%)	31.59	0.00
No. Of Omitted (0-25)		11.81		11.45		12.44	0.00
No. Of Right (0-25)		4.77		5.26		3.83	0.00
No. Of Wrong (0-25)		7.68		7.55		7.94	0.13
Performance in Stage 2 (0-125)	443	43.28	320 (72%)	46.24	123 (28%)	35.59	0.00
No. Of Omitted (0-25)		11.48		10.90		12.99	0.00
No. Of Right (0-25)		6.36		7.07		4.52	0.00
No. Of Wrong (0-25)		7.16		7.03		7.49	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		
	RE	FE	OLS	RE	FE	OLS	RE	FE	OLS
	Performance	Performance	Performance	Performance	Performance	Performance	Performance	Performance	Performance
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Female	-4.466***		-4.493***	-3.513***		-3.513***	-0.405		-0.405
	(0.277)		(0.278)	(0.766)		(0.766)	(0.872)		(0.872)
Stage 2	-1.331**	-14.59***	0.499	-14.59***	-14.59***	-14.59***	-13.16***	-13.16***	-13.16***
	(0.653)	(0.641)	(0.680)	(0.672)	(0.642)	(0.672)	(0.866)	(0.806)	(0.866)
Female*Stage 2	-2.062***	-2.408**	-2.053***	-2.408**	-2.408**	-2.408**	-3.834***	-3.834***	-3.834***
	(0.750)	(0.973)	(0.760)	(1.020)	(0.974)	(1.020)	(1.157)	(1.076)	(1.157)
Math at School	3.673***		3.683***	3.192***		3.192***	2.566***		2.566***
	(0.137)		(0.136)	(0.277)		(0.277)	(0.324)		(0.324)
Constant	35.23***	46.07***	33.47***	58.84***	64.64***	58.84***	47.39***	61.14***	67.98***
	(1.886)	(0.105)	(1.895)	(3.300)	(0.298)	(3.300)	(4.261)	(0.345)	(3.012)
Observations	15,721	15,721	15,721	3,606	3,606	3,606	2,306	2,306	2,306
R-squared		0.449	0.445		0.449	0.610		0.441	0.644
Number of Participants	13,918	13,918		1,803	1,803		1,153	1,153	

Notes: Dependent variables measure performance or score. *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 1-3 include the overall sample, columns 4-6 include the balanced sample, and columns 7-9 include the matched sample. Columns 1, 4 and 7 show random effects model, columns 2, 5 and 8 show the individual fixed effects model, and columns 3, 6 and 9 show OLS estimates. All regressions include level and school fixed effects. Standard errors, clustered at school level, are reported in parentheses with *** p<0.01, ** p<0.05, * p<0.1.