

Mitigating the Gender Gap in the Willingness to Compete: Evidence from a Randomized Field Experiment*

Sule Alan, University of Essex

Seda Ertac, Koc University

January 2016

Abstract

The lower willingness of females to compete is extensively documented, and has a wide range of implications including gender gaps in occupational choice, achievement and labor market outcomes. In this paper we evaluate the impact on competitiveness of two randomized interventions that involve (1) targeted education to foster grit, a non-cognitive skill that has been shown to be highly predictive of achievement, (2) exposure to successful female role models. The interventions are implemented in a large sample of elementary schools, and we measure their impact using a dynamic competition task with interim performance feedback. We find that competitiveness is malleable in both girls and boys. Specifically, when children are exposed to a worldview that encourages goal-setting and perseverance, the gender gap in the willingness to compete disappears. Introducing successful female role models to all children, however, does not replicate this effect, as it leads both boys and girls to compete more relative to control group, leaving the gender gap intact. We explore the effect of treatments on self-confidence and the response to performance feedback as a potential mechanism to explain these results.

JEL Categories: C91, C93, D03, I28

Keywords: competition, gender, grit, role models, randomized interventions, experiments

*We would like to thank the ING Bank of Turkey for providing funding. Ertac thanks the Turkish Academy of the Sciences (TUBA-GEBIP program) and Alan thanks the ESRC Research Center on Micro-Social Change (MISOC) for financial support. We also thank Lale Orta (soccer referee), Ece Ozluk (Turkish Airlines pilot), Aylin Coskun (sea captain), Seda Keskin (chemical engineer), Selda Goktas (medical scientist) for agreeing to be our role models and participate in the interviews that were compiled into short video clips by the great effort of Nergis Zaim and the UNDP media team. We would also like to thank Elif Kubilay, Banu Donmez and Enes Duysak, as well as numerous other students who provided excellent research assistance. All errors are our own.

1 Introduction

It is well-known that fewer women than men occupy top leadership positions in politics and the corporate world, and fewer women choose science-and-technology-related occupations¹. Gender differences in attitudes and preferences towards competition have been put forward as an explanation for these findings; since if fewer women choose to compete, there will be less female winners in the competition for top positions, or in ambitious careers that usually involve competitive paths (see Flory et al. (2014) for field evidence). From an economic standpoint, an important concern here is efficiency: if males and females are equally able in tackling a task and if females shy away from competition involving such a task, then winners of tournaments will on average be less able than if males and females had similar entry rates. Efficiency aside, this result might have far-reaching implications for the society at large, since sustainable economic growth and social cohesion whereby women are financially and socially empowered are difficult to achieve when perfectly able women do not participate in every aspect of economic, social and political life.

In the economics literature, a number of papers use laboratory experiments to explore the effectiveness of policies that aim to mitigate gender differences in competitiveness. For example, one strand of the literature considers affirmative action and preferential treatment through changes in tournament rules favoring women (e.g. Balafoutas and Sutter (2012), Niederle et al. (2013), Sutter et al. (2015)). These studies find that women enter tournaments more frequently with such policies, without sacrificing efficiency. Booth and Nolen (2012) show that single-sex schooling might eliminate gender differences in competitiveness, while Petrie and Segal (2014) find that the gender difference disappears if tournament prizes are high enough.² There is also some evidence that shows that role models in administration, in education and at home can be effective in influencing girls' decisions and aspirations, and thereby reduce gender differences in choices. For example, using a randomized natural experiment in India, Beaman et al. (2012) show that having a female leader as a role model raises girls' career aspirations and eliminates gender differences in adolescent educational attainment. Olivetti et al. (2013) show that mothers as well as friends' mothers can affect the gender identity and subsequent work decisions of girls. In education, there is evidence that female teachers can affect gender gaps in entry into STEM majors, as well as grades and graduation (Carrell et al. (2010) and Bettinger and Long (2005)).

¹Female representation of CEO positions at S&P 500 companies is about 4.4%. In the G20 Summit 2014 in Australia, among the 58 political leaders who represented over 5 billion people around the world, only 5 were women.

²See also Tacsir et al. (2014) for a review of policies aimed at improving women's participation in STEM careers.

The aforementioned studies support the idea that preferences and attitudes, once thought to be fixed and innate, can be influenced by targeted interventions. This idea can be further substantiated by another growing literature that shows that personality traits, often referred to as “non-cognitive skills”, not only are predictive of educational and labor market outcomes, but also may be malleable through educational interventions³. To the extent that these skills are malleable, there may be large benefits to implementing targeted educational interventions early in the life cycle to eliminate inefficient gender differences across a number of domains. That such targeted educational interventions may also offer cost-effective ways of eliminating the gender gap in competitiveness- one of the most debated inefficiency issues in economics- is the focus of this paper.

We explore whether the gender gap in the willingness to compete can be mitigated in childhood through interventions that foster grit or alternatively, through exposing children, in the classroom environment, to female role models who have been successful in male-dominated fields. Grit, as a non cognitive skill, has been shown to be particularly important for educational achievement (see Duckworth et al. (2007), Duckworth and Quinn (2009), Maddie et al. (2012) and Eskreis-Winkler et al. (2014)). This particular skill involves passion for long-term goals, and perseverance and resilience in response to negative feedback. In determining gritty behavior, beliefs regarding the role of effort in the performance process are of utmost importance. An individual will set an ambitious performance goal, will not be easily discouraged by setbacks and will attribute her success to hard work if she believes that the goal is attainable through productive effort⁴. Choosing to compete essentially means setting an ambitious performance goal, and sustained competitiveness in many cases requires perseverance in the competitive path by interpreting failure and success constructively rather than immediately attributing failures to a lack of ability and successes to ability or luck. We therefore conjecture that grit is one of the driving forces of competitive behavior⁵.

The interventions we evaluate in this paper are part of a large-scale randomized educational intervention program conducted in Turkey. The first intervention arm that aims to foster grit was implemented by the teachers for an entire semester (14 weeks, at least 2 hours per week) with the help of a curriculum consisting of video clips, stories and in-class activities that build on the aforementioned ideas. The second intervention arm is inspired by the literature on the importance of role models and

³See Almlund et al., (2011), Kautz et al., (2014), Sutter et al. (2013), Alan and Ertac (2014), Alan, Boneva and Ertac (2015)

⁴Recent research in psychology shows that believing that skills are not fixed but malleable through effort, i.e. having a “growth mindset” can increase motivation, perseverance and achievement (Dweck (2006)).

⁵Grit has also been shown to be malleable in young ages. See Alan, Boneva and Ertac (2015), Blackwell et al. (2007), Aronson et al. (2002) and Good et al. (2003).

highlights the idea that women can excel in any profession they choose. Here, children watch video clips containing testimonies from 5 Turkish women who have become successful in male-dominated professions. After hearing about their success stories, children engage in classroom discussions under the supervision of their own teachers. Due to its nature, this intervention was shorter (two weeks)⁶. While the first arm allows us to test whether a change in mindset about the role of effort in achieving goals influences competitive behavior in girls, the latter allows us to test whether exposure to successful female role models inspires girls to compete more. Using a two-by-two randomized design, we evaluate the impact of each arm independently, as well as the impact of their interaction on the willingness to compete.

Our experimental outcome measures come from an incentivized mathematical real-effort task, whereby children choose to compete, receive performance feedback, then make a choice again. The dynamic nature of our experimental task suits our purposes well, as in real-life performance settings individuals usually do not make one-time, permanent choices but rather can observe how they fared and revise their decisions. Choosing whether to stick with or quit a difficult degree program or a competitive career after receiving negative performance feedback, or choosing whether to pursue more ambitious and competitive paths after doing well in less ambitious ones are important decisions. The use of a competition task with interim feedback given to everyone, allows us to assess heterogenous treatment effects with respect to feedback type and helps us explore potential mechanisms through which treatments work.⁷

We first show that in our control sample, while there is no statistically significant gender gap either in the willingness to compete or in actual performance in the first stage of the competition task, a gender gap of about 11 percentage points emerges in the second stage competition, after feedback. It appears that girls' willingness to compete plummets after receiving negative performance feedback and does not rise as much as that of boys after receiving positive feedback: Conditional on choosing to compete and receiving negative feedback in the first stage, the gender gap significantly widens and becomes 19 percentage points. Conditional on choosing to compete and receiving positive feedback in the first stage, an 8 percentage point gender gap remains. We then analyze how our treatments affect this gap. We find that fostering grit in the classroom environment goes a long way in mitigating the

⁶All materials were developed by a multidisciplinary team that also includes the authors. While the main concepts and ideas were provided by the authors, the end product was the joint effort of a team of six volunteer teachers, two contracted education psychologists, two children's story writes and several media artists. Role models are: a famous FIFA soccer referee, a Turkish Airlines pilot, a high-ranking sea captain, an accomplished chemical engineer and a medical scientist. More information on the materials is given in the appendix and more details can be given upon request.

⁷Andersen et al. (2014) consider a setting with dynamic competition in a matrilineal and patriarchal society, and find that women in the matrilineal society are more motivated by positive performance feedback.

gender gap in the willingness to compete. This particular treatment leads girls to compete as much as boys in the 2nd stage, regardless of the nature of the performance feedback they receive in the first stage. Interestingly, both boys and girls seem to be inspired by the female role models, as we find that this intervention leads to more competitive choices by both genders. The latter effect is quite sizable for boys when the two treatments are interacted. Overall, boys as well as girls who were exposed to female role models exhibit higher willingness to compete relative to the baseline, leaving the gender gap intact. This shows that policies that aim to raise girls' aspirations, such as making female role models visible, can also have inspirational effects on boys. Our rich data allow us to contemplate a potential mechanism that highlights the response of beliefs (self-confidence) to treatment, to explain these results.

The paper contributes to several distinct strands of the literature. First by unraveling the link between grit, an important but understudied non-cognitive skill, and the willingness to compete, it contributes to the growing literature that strives to understand the role of non-cognitive skills in achievement outcomes. Second, by using a randomized-controlled design and showing for the first time the causal impact of an educational intervention on the gender gap in competitiveness, it contributes to the large literature on gender and competition⁸. Besides opening new and fruitful research avenues, the evidence documented in this paper also provides crucial input into policy actions targeting gender gaps in achievement, and offers specific cost-effective recommendations that can be implemented in the classroom environment. Finally, our two-stage design and related results highlight the importance of considering dynamic choices when analyzing gender gaps in competition, and makes novel contributions to the recent literature on performance feedback and its effects on choices and performance (Azmat and Iriberry (2010), Barankay (2011), Eriksson et al. (2009), Ertac and Szentes (2011), Wozniak et al. (2014)).

The rest of the paper is structured as follows: Section 2 provides background information on the program and presents the evaluation design, Section 3 describes the interventions, Section 4 contains the experimental design and outcome measures, Section 5 presents the results, and Section 6 provides a discussion and concluding remarks.

⁸See Croson and Gneezy (2009) and Niederle and Vesterlund (2011) for reviews.

2 Program Background and Evaluation Design

The Turkish Ministry of Education encourages schools and teachers to participate in socially useful extra-curricular programs offered by the private sector, NGOs, the government and international organizations. All elementary school teachers are given a maximum of 5 hours per week to be involved in these programs. Their participation is voluntary and if they choose not to participate in any program there is no restriction on the way in which these hours are used. The interventions we evaluate in this paper are implemented in state-run primary schools in Istanbul as an extra-curricular program under the oversight of the Education Directorate of Istanbul. In the last few decades middle class families in Turkey have been inclined to prefer private schools over under-resourced state schools for their children. Therefore, the program mainly reaches students from lower socio-economic backgrounds.

The main objective of the program is to improve key non-cognitive skills in elementary school children in the classroom environment, under the supervision of their own teachers. The program has three main arms, each of which has a specific behavioral target. The first arm aims to improve the ability to make decisions in a forward-looking manner and encourage patience. The second arm aims to improve the ability to set long term goals and persevere toward those goals through sustained effort, i.e. improve grit . Finally, the third arm aims to overcome stereotype threats faced by girls in terms of occupational choice and to free their aspirations from limitations imposed by societal beliefs and norms. This program initiative required the production of a rich set of educational materials that involve a broad interdisciplinary endeavor. While the target concepts of these materials were determined by the scientific team (the authors of this paper), specific contents (e.g. scripts) were shaped with input from an interdisciplinary team of education psychologists, a group of voluntary elementary school teachers, children’s story writers and media animation artists, according to the age and cognitive capacity of the students.

The program was implemented in the following manner: After official documents were sent to all elementary schools in designated districts of Istanbul by the Istanbul Directorate of Education, 3rd grade teachers in these schools were contacted in random sequence and were offered to participate in the program. When this first contact took place, teachers were informed that the program aims to improve behavior generally related to saving and consumption. In order to allow us to do a random assignment using a version of a phase-in design, they were not promised an immediate involvement, but rather a sure involvement sometime within the following two academic years. Upon random assignment to treatment groups among the teachers who agreed to participate, teachers who were assigned to the

first group to receive the training were re-contacted and invited to full-day “teacher training seminars”. Before teacher training seminars took place, all classrooms were visited and information on a large set of baseline variables were collected from the entire sample (control and treatment) through student and teacher surveys. Note that the unit of randomization in this design is school, not classroom. This is because the physical proximity of classrooms and teachers would have been likely to generate significant spillover effects, if the unit of assignment were classrooms.

After the baseline data collection was complete, teachers in the first treatment group (15 schools) were trained to implement the first arm of the program: “forward looking behavior”. The implementation of this arm lasted for 8 weeks (Spring 2013). While the same teachers (and the same students) received training for the second arm of the program (grit) in the following semester (Fall 2013), the second group of teachers (9 schools) received the first arm (forward-looking behavior). The remaining 12 schools were kept as control for our evaluation⁹. The third arm of the program (female role models) was given to teachers based on a stratified randomized sampling: a randomly selected sample of 9 out of 15 schools from the first group (grit and forward-looking behavior), 3 out of 9 schools from the second group (forward-looking behavior) and 4 out of 12 schools from the control group were assigned to the female role models treatment.

With this design, using the full sample of 36 schools, we can evaluate the impact of i) arm 1+arm 2 (forward-looking behavior+grit), ii) arm 1 (forward-looking behavior), iii) arm 3 (female role models), iv) arm 1+arm 3 (forward-looking behavior +female role model) and v) arm 1+arm 2+arm 3 (forward-looking behavior+grit+female role model). Table A gives a detailed picture of the entire evaluation design. Note that the design does not allow us to evaluate the impact of arm 2 (grit) in isolation, because students who received the grit training had already received the training on forward looking behavior (arm 1) when we collected the follow-up data. While we can still establish the independent impact of arm 1, we cannot rule out the possibility that our results reflect complementarities between the two arms. Incidentally, we do not find a statistically significant independent impact of forward-looking behavior training (arm 1) on any of our outcome variables. Therefore, in order to circumvent the small cluster problem, we consider children who received only the training on forward looking behavior as part of our control group. All related analyses regarding this choice are presented in the Appendix (Table 1 and 2).

We now provide more information about the nature of the two interventions, grit and exposure to

⁹All teachers eventually received all training materials as promised, albeit at different times within the two academic years.

female role models, which we focus on in this paper.

3 Interventions

In this section we provide some details regarding the nature of the two intervention arms, arm 2 (grit) and arm 3 (female role models)¹⁰. It is important to note here that neither of the interventions contained any material that mentioned or encouraged competitiveness or risk-taking per se. Therefore, the interventions had no direct link to the experimental task we use for measuring competitive behavior.

3.1 An Educational Intervention on Grit

This intervention involves providing animated videos, mini case studies and classroom activities that highlight i) the plasticity of the human brain against the notion of innate ability, ii) the role of effort in enhancing skills and achieving goals, iii) the importance of a constructive interpretation of setbacks and failures, and iv) the importance of goal setting. The aim of the training is to expose students to a worldview in which any one of them can set goals in an area of their interest and can work towards these goals by exerting effort. The materials highlight the idea that in order to achieve these goals, it is imperative to avoid interpreting immediate failures as a lack of innate ability or intelligence. This worldview encompasses any productive area of interest, whether it be music, art, science or sports. The specific educational materials include animated videos, stories and classroom activities highlighting these ideas¹¹. Visual materials and stories are supplemented by classroom activities created and supervised by teachers, based on general suggestions and guidelines put forward in teacher training¹².

3.2 Exposure to Female Role Models

This is a two-week intervention that aims to impart to students the idea that women can excel at any occupation they set their minds to, and should not be discouraged by disbelief on the part of

¹⁰Details regarding the content of arm1 and the results of its evaluation can be found in Alan and Ertac (2014).

¹¹To give an example, in an animated video, two students who hold opposite views on the malleability of ability engage in a dialog. The student who believes that ability is innate and therefore there is no scope for enhancing it through effort, points out that the setbacks she experiences are reminders of the fact that she is not intelligent. Following this remark, the student who holds the opposite view replies that setbacks are usually inevitable on the way to success; she interprets them as opportunities to learn, and therefore, they do not discourage her.

¹²In Alan, Boneva and Ertac (2015), we evaluate the impact of this particular intervention on grit before the second treatment considered in this paper, exposure to female role models, was implemented. We find a significant treatment effect with respect to perseverance, goal setting and resilience to negative feedback within the context of an experimental real effort task, without any significant gender difference.

others. Specifically, students watch a series of video clips about five successful Turkish women who have excelled at male-dominated occupations such as scientist, engineer, sea captain, pilot, and soccer referee. The video clips contain excerpts from interviews with these women, where they share anecdotes about their childhood aspirations related to their chosen professions, overlaid by pictures from their childhood years in order to make it easier for children to form a link between themselves and the role model. Teachers also conduct class activities related to these video clips, such as discussions of the women’s stories, drawings related to career aspirations etc. It is important to note that while the intervention is aimed at girls, boys also participate in the activities.

4 The Outcome Measure: A Two-Stage Competition Task

Our outcome measure is designed to explore the effects of the two randomized interventions, grit and female role-model, on (1) initial competitiveness, (2) competitiveness in response to performance feedback, (3) absolute and relative self-confidence, and (4) gender stereotypes. The task used in the experiment is an addition task, which involves adding two 2-digit numbers and one single-digit number. Children are given 2 minutes per each performance period, and more addition questions than they can possibly complete during this time. The experiment consists of three periods. One of these periods is selected randomly at the end, and rewards are determined based on the performance and decisions in the selected period. For all three periods, each student is matched with another student from a different school (whom we will call “opponent” hereafter), who had done the same tasks before and whose performance was recorded. In the first period, students perform the addition task under a piece-rate incentive scheme, whereby they receive 1 token for every addition they are able to do correctly. In the 2nd period (1st competition choice stage), students have a choice between piece-rate and competing with their matched opponent. If they choose the piece-rate, they are rewarded with one token per correct answer. If they choose the tournament, they are rewarded with 3 tokens per correct answer, but only if their performance exceeds that of the opponent. In the case of having a lower performance, a student that chooses to compete receives zero tokens, whereas in case of a tie, she gets 1 token per correct answer.

After choosing the incentive scheme, children are asked to state their beliefs about (1) the number of correct answers they will have, (2) the number of correct answers of their opponent. These beliefs are incentivized in the following way: Children are told that three people in their class will be randomly

chosen after the experiment ends, and these three will get an extra small gift for each correct guess¹³. After performing the 2nd period task, they receive feedback about (1) the number of additions they were able to do correctly, (2) whether their performance was better than, worse than, or equal to that of the opponent. In the 3rd period (the 2nd competition choice stage), children make a choice again between piece-rate and competition, to be implemented if that period is chosen for payment. After making this choice, they state beliefs again about their own and opponent’s performance, in the same way as in the 2nd period. If the child chooses to compete in any period, her results are compared with the performance of the opponent in the corresponding period, and rewards are determined based on this comparison.

Children learn the gender of their (randomly assigned) opponent before the making their initial competition choice. That is, they know whether they are matched with a boy or a girl. Before this, there is also an incentivized question that elicits gender stereotypes: children are told about one girl and boy, randomly selected from a different school/class, and are asked to guess whether the girl or boy has done better in this same task. If they guess who actually did better correctly, they get a small extra gift. Children were given workbooks, and instructed to turn over pages only at specified times . Each child’s workbook had a number that matched it to a set of actual performances (in the 2nd and 3rd periods) that came from a student in another class. The workbooks were randomly distributed, allowing a random match in terms of opponent gender and opponent performance. In addition to the main decisions as explained above, we have access to (1) baseline risk preferences, collected pre-treatment through an elicitation task based on Gneezy and Potters (1997), (2) detailed information from the teacher about each student, on a variety of pre-treatment traits and behavior, as well as on family characteristics, (3) a student questionnaire.

5 Results

We have data from over 1900 students in 69 classes in 36 schools. About 52% percent of our sample is male, and the average age is about 10 years, as is typically the case for 4th graders in Turkey. In what follows, we conduct our empirical analyses based on a 2-by-2 factorial design whereby we denote our grit intervention by “Grit”, female role model intervention by “FRM”, and their interaction by “Grit+FRM”.

¹³Children are also reminded that there is always an incentive to do as many additions as they can in the actual task and it does not make sense to stop just to be consistent with their guess.

5.1 Internal Validity

We first check whether our data are balanced across these treatment groups with respect to a number of student characteristics, collected at the baseline stage. Table 1 shows results from ordinary least squares regressions of baseline variables on three treatment dummies, leaving the control group as the omitted category. Therefore, whilst the first column gives the mean of the control for the respective variable, subsequent columns present the difference from the control’s mean. Panel 1 presents variables obtained from baseline surveys and Panel 2 presents the balance of three within-task variables. As can be seen from the table, differences are not statistically different from zero with the exception of the percentage of males in the FRM arm and the percentage of male teachers in the Grit arm. As the randomization was conducted at the school level and given the fact that we check a large number of variables here, having two significant coefficients is not unusual.

An important finding to note in this table (Panel 2) is that there is no difference in piece-rate performance across our treatment groups. Since it is balanced across treatment groups, and as we present below, highly predictive of our outcome variables (competition choices in stage 1 and stage 2), we choose to use this variable as a covariate in our treatment effect analyses in addition to our baseline variables. Note, however, that since this is a randomized design, the use of covariates is not necessary to obtain unbiased estimates of the average treatment effect on the treated.¹⁴

5.2 Willingness to Compete and Gender Gaps in the Control Group

Before moving on to estimating the effect of our interventions, we study the gender gap in the control group. Focusing on elementary school children and using a familiar experimental task, this initial analysis provides new evidence on the prevalence of the gender gap in the willingness to compete and gives us the baseline gender gap figures upon which our treatment operates. Table 2 presents marginal effects from logit regressions where the dependent variables are the binary competition choices in stage 1 and stage 2. The first point to note in this table is that there is no evidence of a gender gap in the first stage competition choice (see column 1) . However, the picture is quite different for the second stage: here we find a statistically significant and sizable gender gap in the willingness to compete. More specifically, we estimate that girls’ willingness to compete is about 11 percentage-points lower

¹⁴Since randomization ensures that all covariates are independent of the treatment status, omitting them does not cause omitted variable bias, nor including them is expected to change the size of the estimated treatment effect. However, including highly predictive covariates is likely to increase the precision of the estimated treatment effects via reducing the variation in regression errors.

than boys, after feedback.

A major question that the gender-competition literature has focused on is the determinants of competitiveness in girls and boys, which also provides insights about the nature of the gender gap. Columns 2 and 4 of Table 2 include the main potential determinants of competition choice as explanatory variables: risk attitude, past (piece-rate) performance, and beliefs (relative self-confidence, defined as the expected number of correct answers for oneself minus that for the opponent), in addition to opponent's gender, teacher's gender and cognitive ability score. As expected, higher expected performance relative to the opponent and higher past performance both increase the propensity to compete in the 1st stage. In the second stage, relative performance in the first stage tournament also emerges as a very significant predictor of the competition choice. Note that the gender of the opponent has no effect in either stage, a finding we will revisit later in the text.

Why do we see a gender gap emerging in the second stage? One explanation is that if girls performed generally worse than their opponents in the first stage, we could observe girls having a higher tendency to shy away from competition (perhaps rationally) in the second stage. However, there is absolutely no gender gap in performance in the first stage. As can be seen in Table 3, the performance of boys and girls is not statistically different either in the first or in the second stage, and naturally, their propensity to receive negative/positive feedback after the first stage is also the same (46.4% of girls and 46.8% of boys receive negative feedback). Note also that the second stage gender gap remains when we control for relative performance (feedback type) in the first stage (Table 2, Column 4). A plausible explanation here could be that girls are less resilient to negative performance feedback than boys, and therefore have a higher tendency to give up after a failure. In order to explore this, we first look at the second stage competition choice conditional on feedback, then disaggregate further and condition on competition choice as well as feedback type. Table 4 shows that boys are more likely to keep competing than similar girls after receiving negative or positive feedback, although the results on negative feedback appear stronger. While males always have a higher likelihood of competing in the second stage, the gender difference is largest in the case of initial choice of competition plus negative feedback. Table 5 documents this clearly: Conditional on choosing to compete in the first stage and receiving negative feedback, the gender gap significantly widens and becomes 19 percentage points. Conditional on choosing to compete in the first stage and receiving positive feedback, a gender gap of 8 percentage points remains. A similar gender gap also exists among boys and girls that choose piece-rate in the first stage and receive negative feedback. These results suggest that girls' willingness

to compete declines more than boys’ after receiving negative performance feedback and does not rise as much as that of boys after receiving positive feedback. We will explore the feedback channel further when we contemplate the mechanisms through which the treatments might be working.

5.3 Treatment Effects on Competition Choice

We now turn to estimating treatment effects on the willingness to compete for boys and girls separately. In order to test the null hypothesis that a given intervention had no impact on the experimental outcome y , competition choice, we estimate the average treatment effect by conditioning on baseline covariates:

$$y_{ij} = \alpha_0 + \alpha_1 Grit_j + \alpha_2 GritFRM_j + \alpha_3 FRM_j + X_{ij}\gamma + \varepsilon_{ij}$$

where the dependent variable y_{ij} is a dummy variable which equals 1 if student i in school j chose to compete in stage 1 (stage 2), $Grit_j$, $GritFRM_j$ and FRM_j are treatment dummies, and X_{ij} is a vector of observables for student i in school j that are potentially predictive of the outcome measures we use. These observables include opponent gender, cognitive function (measured by Raven’s progressive matrices), piece-rate performance, risk preferences and teacher gender. The estimated vector of $\hat{\alpha}_k$ is the average treatment effect on the treated. Estimates are obtained via logit regressions since the outcome considered here is binary¹⁵.

Table 6 presents the estimated treatment effects on the willingness to compete in the first stage for girls and boys. The first finding to note here is that children’s 1st-stage competition choices are very responsive to both treatments as well as their combination (column 1). We cannot reject the equality of all three treatment coefficients (p-value=0.42). This finding shows that willingness to compete is malleable in children. The second column in the same table shows that grit education is effective in increasing competitive behavior among girls: the grit intervention and its combination with female role models reach statistical significance in the first stage. Boys’ competitiveness also responds positively to treatment, but this effect, interestingly, is crucially dependent on role models. Boys who were exposed to successful female role models exhibit higher willingness to compete relative to the boys who are in the control group. Moreover, while the grit intervention does not have any significant impact on boys’ competitiveness by itself, when it is complemented with exposure to female role models the impact is quite large and significant.

¹⁵In all empirical analyses where we estimate the treatment effect, standard errors are bootstrapped and clustered at the level of the school, which is the unit of randomization.

5.4 Treatment Effects on Competition Choices after Feedback

The two-stage nature of our experimental task provides us with an outcome measure that is useful for assessing the impact of the interventions on behavior after receiving performance feedback. Recall that before making the second competition decision, students receive feedback on how well they did relative to their opponents in the first stage. This feedback is given to everyone, regardless of the choice of incentive scheme in the first stage, allowing us to condition on feedback when estimating the treatment effects on the second stage choices. Table 7 presents the estimated treatment effects on the propensity to compete in the second stage. Here, we see similar results but much stronger: All three treatment arms are equally effective for making girls more competitive (we cannot reject the equality of all three treatment coefficients, $p\text{-value}=0.41$), while in the case of boys, the role models treatment seems to be effective, especially when it is interacted with the grit treatment. Here, boys who received both treatments are much more likely to compete than boys who received only the FRM treatment, with an effect size that is more than double ($p\text{-value}=0.002$). Recall that the grit intervention exposes the treated children to ideas about the malleability of ability and a more constructive interpretation of negative performance feedback. It appears that these ideas were very effective for girls, while by themselves, they do not have a significant impact on the competitiveness of boys. For boys, competitive drive seems to be invoked by successful role models, especially when ideas that promote grit were already present in children’s minds.

An important expected aspect of the treatments (especially the grit treatment) is influencing the response to feedback. Figure 1 presents competition choices by girls and boys after receiving non-negative and negative feedback, across treatment. It is important to recall here that performance in the first competition stage and the feedback received based on this performance are balanced across treatment groups. The figure shows that for girls, all three treatments induce a similar and strong increase in competitive response after non-negative feedback, while treatments involving grit education seem to have the highest positive effect on competitiveness after negative feedback. For boys, two immediate points to note based on the figure is the motivational effect of the treatments that involve role models in response to non-negative feedback, which is absent in the grit-only treatment, and the strong effect of the combined treatment after both positive and negative feedback.

In order to explore these observations in more detail, Table 8 provides treatment effect estimates according to the type of feedback received (negative or non-negative), for the overall sample as well as for each gender separately. The first two columns of Table 8 show that after non-negative feedback

(receiving information that performance was the same as or better than opponent), treated children end up competing more than untreated children, and this is true for all three treatment groups. However, after receiving negative feedback, only the children that received grit training in some form (either coupled with FRM or by itself) compete more than control, while the role models treatment does not have an effect by itself. Columns 3 to 6 break this down by gender as well as feedback type. Consistently with the overall picture, girls that received any of the treatments are more likely to compete after non-negative feedback. However, after negative feedback, only girls that received grit training (either coupled with FRM or by itself) compete more than control, while FRM does not have an effect by itself (the coefficient of Grit+FRM has a p-value of 0.101). The picture is different for boys: After non-negative feedback, boys that have been exposed to the role models (either by itself or coupled with grit) compete more than control. However, after negative feedback, only the Grit+FRM boys compete more than the control group.

These results suggest that for girls, exposure to role models on its own does not influence the response to negative feedback, and grit training is both necessary and sufficient. The role models treatment on its own does not have an effect on boys' response to negative feedback either, but this treatment has a more crucial role for boys' competitiveness than in the case of girls. Role model exposure is necessary for increasing boys' competitiveness both after positive and negative feedback, as grit training does not seem to influence boys at all on its own. But while exposure to role models is sufficient to motivate boys to compete more after positive feedback, it still needs to be complemented with grit training for it to lead to resilience after negative feedback.

5.5 Treatment Effects on the Gender Gap

Recall that we observe no significant gender gap in the willingness to compete in the first stage, while a sizable gap appears in the second stage in the control sample (about 11 percentage points without covariates and 9 percentage points with covariates, see Table 2). Since the FRM treatment increases both girls' and boys' competitiveness whereas the grit treatment affects only girls, the natural question to ask is what these treatments do to the gender gap (or the lack thereof) compared to control. Table 9 presents regressions of the first and the second stage competition choices on gender, treatment, and gender-treatment interactions. None of the gender-treatment interaction terms are statistically significant for stage 1, showing that the gender gap in competitiveness in the first stage, which was insignificant in the baseline as well, is not affected significantly by any of our treatments.

The second column in Table 9 refers to the second stage. Here we see that the intervention on grit is the only effective treatment in mitigating the gender gap (see the negative and significant interaction term, GritXMale). A Wald test shows that there is no significant difference in competitiveness between girls and boys that received the grit treatment only, i.e. the gender gap observed in the baseline is closed by this particular treatment. Figure 2 shows this in visual clarity: it depicts the coefficient estimates and confidence bands on the male dummy in the logit regressions which were carried out separately for each treatment cell (controlling for all relevant covariates). As can be seen here, the role models treatment (FRM and Grit+FRM) which significantly increases boys' competitiveness as well as girls', does not reduce the gender gap. The particularly strong effect on boys' competitiveness coming from the complementarity of the two interventions is worth noting. It is clear that in an environment where ideas on grit and the malleability of abilities are already in place, seeing female role models can strongly affect not only girls but also boys. These results call for further research to understand how educational interventions targeting gender gaps are perceived by both genders. Policies that attempt to neutralize gender gaps should take into account their potential effects on boys' perceptions and behavior.

An important question here is whether treatment leads to higher payoffs for the children. Based on performance, for some children it is payoff-maximizing to compete, while for others it is better to stay out. One can be concerned that interventions such as the ones we evaluate in this paper may lead to unintended inferior outcomes for some children, by inducing decisions that turn out to be bad for payoffs ex-post. Analyzing how children's decisions fare in terms of expected material payoffs can shed light on these issues. For this, we first calculate the probability of winning for each performance level in the first competition stage, using the empirical distribution of opponents' performances¹⁶. Using these probabilities and the realized first-stage performance, we calculate the expected payoff from competition. We then estimate treatment effects on the payoff from stage 1, payoff from stage 2, and the total payoffs (from Stage 1 and Stage 2). We find that none of the treatments reduce payoffs, either in the first or the second stage. In fact, all three treatments lead to significantly higher total payoffs for participating children (see Table 10)¹⁷. A Wald test shows that the estimated treatment effects are not significantly different from each other ($p=0.47$).

¹⁶Here we use performances from three classes in different schools, where opponent data was collected. We do a Monte-Carlo simulation by drawing 1000 times from the opponent distribution for each performance, and calculating the empirical win, lose and tie probabilities.

¹⁷Presented coefficient estimates are obtained by controlling for initial performance, opponent's gender, risk attitude, cognitive ability score and teacher's gender.

5.6 Mechanisms of Impact

Which aspects of the treatments are likely to drive girls and boys into competition, and what is the reason for the differential impact of the grit and FRM treatments across gender? It may be that multiple factors are at work in driving the response to treatments, and the relative weight of these may be different for girls and boys. One major factor that underlies competition choice is beliefs. We elicit beliefs by asking children about their absolute and relative expected performance, in both choice stages. More specifically, children are asked how many questions they expect to answer correctly and how many they expect their opponent to answer correctly. We construct a relative self-confidence measure using (and normalizing) these variables. As a backdrop to what follows, we first note that in the control group, boys are more self-confident than girls in both the 1st stage and the 2nd stage, controlling for actual performance, which is consistent with the literature on gender gaps in confidence in adults¹⁸. We first analyze the treatment effect on the number of questions the subject expects to solve correctly minus the expectation about the opponent’s performance (which can be a positive or negative number), among girls and boys.

Table 11 shows that girls’ and boys’ first-stage beliefs respond to treatment differently. The first two columns show that girls’ initial confidence levels are positively affected by all treatments, that is, treated girls expect that they will do better relative to the opponent compared to untreated girls. The results imply effect sizes between 20 to 30 percent of a standard deviation. Boys’ relative beliefs, on the other hand, are not significantly affected by any of the treatments, although all coefficients are positive. Given that the piece-rate performance of girls is not different across treatment, this increase in relative beliefs is not driven by differences in ability but is likely to indicate an increase in self-confidence. This self-confidence improvement can explain why treated girls opt to compete more¹⁹. While beliefs are one of the main potential mechanisms through which the interventions affect competition choices, there are other possibilities. One such possibility is that the treatments might have affected risk tolerance levels. While we do not have data on post-treatment risk tolerance for all treatment groups, we have data to compare risk tolerance across the groups that received the grit treatment versus not (this second risk measure was taken after the grit treatment was implemented, but before the FRM treatment). In unreported regressions, we find that the grit treatment does not

¹⁸See Croson and Gneezy (2009) for a review. Regression results are available upon request.

¹⁹Estimated treatment effects are also positive for second-stage beliefs, i.e. all treatments lead to an increase in self-confidence, controlling for feedback type. The effects do not reach statistical significance, which is likely because 2nd-stage beliefs are driven largely by factual data (on own performance and performance relative to opponent). In this sense, first-stage beliefs are probably a better measure to capture self-confidence, since they are elicited in the absence of feedback on actual performance.

have a significant effect on risk tolerance ($p=0.36$). This suggests that the impact of grit training on competitive behavior does not work through increasing the tolerance for risk.

An interesting finding that our data reveal is that like girls, and in some cases more strongly than girls, boys' competitiveness is positively influenced by female role models. This may be because boys get motivated to compete with women (in an "if she can do it, I can definitely do it/do better than her" kind of way), or they are just generally inspired by their stories. In order to shed some light on this question, we look at whether there are differences in the response to treatment by opponent gender. We find no evidence that boys' competition response to the female role model treatment (or any other treatment, for that matter) depends on whether their opponent is male or female. This suggests that the role model treatment does not invoke an urge in boys to compete more with women to show that they can do better. More evidence that supports this comes from post-treatment gender stereotypes. Recall that we have an incentivized measure of stereotypes through a question that asks subjects to guess whether a boy or a girl would do better in the addition task, or whether they would perform similarly. Table 12 shows that all three treatments reduce the likelihood that boys state a boy would do better (FRM does not reach statistical significance). Survey responses collected before and after the interventions about gender stereotypes also point to a similar effect. Table 13 shows very strongly that both treatments reduced boys' gender biases significantly.

Overall, beliefs (improved self-confidence) are likely to play the biggest role in generating the treatment effects for girls, while in the case of boys, treatments might be invoking a stronger "pure preference" for competition rather than operating through beliefs. Given the reduction in the gender biases of boys, we are inclined to believe that boys were inspired by the role models as much as girls. This is also evident from the strong positive response to positive feedback in boys that were exposed to female role models, which is consistent with a motivational/inspirational story (See Table 8).

6 Discussion and Concluding Remarks

Documenting, understanding the reasons for, and exploring policies to mitigate gender gaps in competitiveness have been a very active area of research in economics in recent years. Our paper views the willingness to compete as an individual trait that is linked to other non-cognitive skills such as grit, and explores whether it can be malleable through (1) an educational intervention that aims to promote gritty behavior, i.e. goal-setting and perseverance through sustained effort, and (2) exposure to successful female role models, who can provide inspiration to set and work towards ambitious goals. While

the competition propensity of boys and girls are similar in the first choice stage, a significant gender gap emerges in the second stage competition choice in the control group, after receiving performance feedback.

The absence of an initial gender difference is consistent with other studies that find no gender differences in competitiveness in pre-pubescent children in developing countries (e.g. Andersen et al. (2013), Cardenas et al. (2012), Khachatryan (2012))²⁰. The emergence of a significant gap after feedback suggests, however, that the initial gender balance easily unravels in the 2nd stage, which likely stems from the differential response of girls and boys to performance feedback. Indeed, we find that in the absence of any intervention, girls' competitiveness falls more after negative and rises less after positive feedback than boys.

We find that our interventions can have significant effects on both girls' and boys' competitiveness, which naturally has implications for the gender gap as well. First of all, the data support the hypothesis that educating children on the importance of perseverance and the value of effort in achieving success can have effects on competitiveness. Grit training leads to a higher propensity to compete both in the first and the second stage, but the effects are especially pronounced in influencing the second period competition choice, after feedback. Specifically, grit education is necessary for inducing higher competition levels after receiving negative performance feedback. Female role models are also effective in increasing competitiveness, for both girls and boys. However, there seem to be interesting interactions between gender, whether negative or positive feedback was received, and the response to the two treatments. While grit education is necessary and sufficient by itself to induce girls to compete more in the second stage, in the case of boys, it only works if complemented with exposure to role models. Grit training is still a precondition for resilience to negative performance feedback, but role models are also needed for boys. Seeing successful female role models does not affect self-confidence negatively for boys, improves their gender stereotypes, and probably creates an inspirational effect that increases the pure preference for competition. Coupled with the fact that the treatments are overall payoff-improving for both girls and boys, these results show that making female role models salient do not lead to inferior outcomes for boys, addressing a concern that policymakers might have for motivational interventions targeted towards girls²¹.

While our treatments improve both girls and boys' competitiveness, they may work through differ-

²⁰Sutter and Glatzle-Ruetzler (2015), on the other hand, find that gender differences in a developed country (Austria) start very early and persist.

²¹Beaman et al (2012) also find no effect of female role models in administration on boys' inspirations.

ent mechanisms for the two groups. An important point that is highlighted by the results is the need to consider the effects on boys of policies targeted to girls, since some policy combinations (as in the case of grit education supplemented with female role models) can have large effects on boys also, and can leave the gender gap intact.

Finally, our data highlight the importance of focusing on dynamic frameworks when analyzing competitive choice. Performance feedback is ubiquitous in economic life and in education, and competitiveness interacts with performance feedback in ways that can contribute to widening gender gaps. Our results show that such gender gaps can be mitigated by educational and role model interventions, which have effects on self-confidence and the response to performance feedback. In this sense, our paper provides a complement to the interventions in the literature that close the gender gap by changing tournament rules and thereby the material incentives for women. Incorporating the type of interventions proposed in this paper into the classroom environment (or even the home environment) can be an easy and cost-effective alternative from a policy perspective.

References

- [1] Alan, S. and Ertac, S. (2014), “Good Things Come to Those Who (Are Taught How to) Wait: Results from a Randomized Educational Intervention on Time Preference”, HCEO Working Paper.
- [2] Alan, S., Boneva T. and Ertac S. (2015), “Ever Failed, Try Again, Succeed Better: Results from a Randomized Educational Intervention on Grit”, HCEO Working Paper.
- [3] Almlund, M., Duckworth, A.L., Heckman, J.J. and Kautz, T.D. (2011), “Personality Psychology and Economics”, *Handbook of the Economics of Education*, pp. 1-181
- [4] Andersen, S., Ertac, S., Gneezy, U., List, J. A., & Maximiano, S. (2013). Gender, competitiveness, and socialization at a young age: Evidence from a matrilineal and a patriarchal society. *Review of Economics and Statistics*, 95(4), 1438-1443.
- [5] Andersen, S., Ertac, S., Gneezy, U., List, J. A., & Maximiano, S. (2014). Gender, competitiveness, and the response to performance feedback: Evidence from a matrilineal and a patriarchal society. Mimeo.
- [6] Aronson, J, Fried, C.B. and Good, C. (2002), “Reducing the Effects of Stereotype Threat on African American College Students by Shaping Theories of Intelligence”, *Journal of Experimental Social Psychology*, 38, pp. 113-125
- [7] Azmat, G. and Iriberry, N. (2010), “The Importance of Relative Performance Feedback Information: Evidence from a Natural Experiment using High School Students”, *Journal of Public Economics*, 94, pp. 435-452
- [8] Balafoutas, L., & Sutter, M. (2012). Affirmative action policies promote women and do not harm efficiency in the laboratory. *Science*, 335(6068), 579-582.
- [9] Barankay, I. (2011). Rankings and social tournaments: Evidence from a crowd-sourcing experiment. In Wharton School of Business, University of Pennsylvania Working Paper.
- [10] Beaman, L., Duflo, E., Pande, R., & Topalova, P. (2012). Female leadership raises aspirations and educational attainment for girls: A policy experiment in India. *Science*, 335(6068), 582-586.
- [11] Blackwell, L., Trzesniewski, K.H. and Dweck, C.S. (2007), “Implicit Theories of Intelligence Predict Achievement Across an Adolescent Transition: A Longitudinal Study and an Intervention”, *Child Development*, 78(1), pp. 246-263

- [12] Bettinger, E. P., & Long, B. T. (2005). Do faculty serve as role models? The impact of instructor gender on female students. *American Economic Review*, 152-157.
- [13] Booth, A., & Nolen, P. (2012). Choosing to compete: How different are girls and boys?. *Journal of Economic Behavior & Organization*, 81(2), 542-555.
- [14] Borghans, L., Duckworth, A.L., Heckman, J.J. and ter Weel, B. (2008), “The Economics and Psychology of Personality Traits”, *Journal of Human Resources*, 43(4), pp. 972-1059
- [15] Cárdenas, J. C., Dreber, A., Von Essen, E., & Ranehill, E. (2012). Gender differences in competitiveness and risk taking: Comparing children in Colombia and Sweden. *Journal of Economic Behavior & Organization*, 83(1), 11-23.
- [16] Carrell, S., Page, M. and J. West (2010), “Sex and science: How professor gender affects the gender gap,” *Quarterly Journal of Economics* 125, 1101-1144.
- [17] Chetty, R., Friedman, J.N., Hilger, N., Saez, E., Diane, S.W., Yagan, D. (2010), How Does Your Kindergarten Classroom Affect Your Earnings? Evidence from Project STAR, NBER Working Paper 16381
- [18] Conti, G., Heckman, J.J., Moon, S. and Pinto, R. (2014), The Long-term Health Effects of Early Childhood Interventions, unpublished manuscript.
- [19] Croson, R., & Gneezy, U. (2009). Gender differences in preferences. *Journal of Economic literature*, 448-474.
- [20] Cunha, F., Heckman, J.J., Lochner, L., Masterov, D.V. (2006), “Interpreting the Evidence on Life Cycle Skill Formation”, *Handbook of the Economics of Education*, 1, pp. 679-812
- [21] Cunha, F., Heckman, J.J. and Shennach, S.M. (2010), “Estimating the Technology of Cognitive and Noncognitive Skill Formation”, *Econometrica*, 78(3), pp. 883-931
- [22] Cunha F., Elo I, Culhane J. (2013), “Eliciting Maternal Expectations about the Technology of Cognitive Skill Formation”, NBER Working Paper 19144
- [23] Dee, T. and West, M. (2008), “The Non-Cognitive Returns to Class Size”, NBER Working Paper 13994

- [24] Duckworth, A.L., Peterson, C., Matthews, M.D., and Kelly, D.R. (2007), “Grit: Perseverance and Passion for Long-Term Goals”, *Journal of Personality and Social Psychology*, 92(6), pp. 1087-1101
- [25] Duckworth, A. L. and Quinn, P. D. (2009), “Development and validation of the Short Grit Scale (Grit-S)”, *Journal of Personality Assessment*, 91, pp. 166-174
- [26] Dweck, C. *Mindset: The New Psychology of Success*. Random House Digital, Inc., 2006
- [27] Eriksson, T., Poulsen, A. and Villeval, M.C. (2009), “Feedback and Incentives: Experimental Evidence”, *Labour Economics*, 16(6), pp. 679-688
- [28] Ertac, S., & Szentes, B. (2011). The effect of information on gender differences in competitiveness: Experimental evidence. TÜSİAD-Koç University Economic Research Forum working paper series.
- [29] Flory, J. A., Leibbrandt, A., & List, J. A. (2014). Do Competitive Workplaces Deter Female Workers? A Large-Scale Natural Field Experiment on Job-Entry Decisions. *The Review of Economic Studies*, rdu030.
- [30] Eskreis-Winkler, L., Shulman, E.P., Beal, S. and Duckworth, A.L. (2014), “Survivor mission: Why those who survive have a drive to thrive at work”, *Journal of Positive Psychology*, 9(3), pp. 209-218
- [31] Gneezy, U. and Potters, J. (1997), “An experiment on risk taking and evaluation periods”, *Quarterly Journal of Economics*, 112(2), pp. 631-645
- [32] Golsteyn, B.H.H., Gronqvist, H. and Lindahl, L. (2013), “Adolescent Time Preferences Predict Lifetime Outcomes”, *Economic Journal*, 124, pp. 739-761
- [33] Good, C., Aronson, J. and Inzlicht, M. (2003), “Improving Adolescents’ Standardized Test Performance: An Intervention to Reduce the Effects of Stereotype Threat”, *Applied Developmental Psychology*, 24(6), pp. 645-662
- [34] Heckman, J.J., Stixrud, J. and Urzua, S. (2006), “The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behaviour”, *Journal of Labor economics*, 24, pp. 411-482
- [35] Heckman, J.J., Moon, S.H., Pinto, R., Savelyev, P.A. and Yavitz, A.Q. (2010), “The Rate of Return to the High Scope Perry Preschool Program”, *Journal of Public Economics*, 94(1-2), pp. 114-128

- [36] Heckman, J.J., Pinto, R. and Savelyev, P.A. (2013), “Understanding the Mechanisms Through Which an Influential Early Childhood Program Boosted Adult Outcomes”, *American Economic Review*, 103(6), pp. 1-35
- [37] Heckman, J.J., Moon, S.H. and Pinto, R. (2014), The Effects of Early Intervention on Abilities and Social Outcomes: Evidence from the Carolina Abecedarian Study, unpublished manuscript
- [38] Kautz, T., Heckman, J.J., Diris, R., ter Weel, B., Borghans, L. (2014), Fostering and Measuring Skills: Improving Cognitive and Non-cognitive Skills to Promote Lifetime Success, NBER Working Paper 20749
- [39] Khachatryan, K. (2012). Gender differences in preferences at a young age? Experimental evidence from Armenia. Stockholm School of Economics.
- [40] Maddie, S.R., Matthews, M.D., Kelly, D.R., Villarreal, B. and White, M. (2012), “The Role of Hardiness and Grit in Predicting Performance and Retention of USMA Cadets”, *Military Psychology*, 24, pp. 19-28
- [41] Moffit, T. E., Arseneault, L., Belsky, D., Dickson, N., Hancox, R.J., Harrington, H., Houts, R., Poulton, R., Roberts, B.W., Ross, S., Sears, N. R., Thomsom, W. M. and Caspi, A. (2011), “A Gradient of Childhood Self-control Predicts Health, Wealth, and Public Safety”, *Proceedings of the National Academy of Sciences*, 108 (7), 269-398
- [42] Niederle, M., Segal, C., & Vesterlund, L. (2013). How costly is diversity? Affirmative action in light of gender differences in competitiveness. *Management Science*, 59(1), 1-16.
- [43] Niederle, M., & Vesterlund, L. (2011). Gender and competition. *Annu. Rev. Econ.*, 3(1), 601-630.
- [44] Olivetti, C., Patacchini, E., & Zenou, Y. (2013). Mothers, friends and gender identity (No. w19610). National Bureau of Economic Research.
- [45] Petrie, R., & Segal, C. (2014). Gender differences in competitiveness: The role of prizes. Available at SSRN 2520052.
- [46] Raven, J., Raven, J.C., & Court, J.H. (2004), “Manual for Raven’s Progressive Matrices and Vocabulary Scales”, San Antonio, TX: Harcourt Assessment

- [47] Roberts, B.W., Kuncel, N.R., Shiner, R.L., Caspi, A. and Goldberg, L.R. (2007), “The Power of Personality: The Comparative Validity of Personality Traits, Socioeconomic Status, and Cognitive Ability for Predicting Important Life Outcomes”, *Perspectives in Psychological Science*, 2(4), pp. 313-345
- [48] Sutter, M., Kocher, M.G., Ruetzler, D. and Trautmann, S.T. (2013), “Impatience and Uncertainty: Experimental Decisions Predict Adolescents’ Field Behavior”, *American Economic Review*, 103(1), pp. 510-531
- [49] Sutter, M., & Glätzle-Rützler, D. (2015). Gender differences in the willingness to compete emerge early in life and persist. *Management Science*, forthcoming.
- [50] Sutter, M., Glätzle-Rützler, D., Balafoutas, L., & Czermak, S. (2015). Canceling out early age gender differences in competition—an analysis of policy interventions. *Experimental Economics*, forthcoming.
- [51] Tacsir, E., Grazi, M., & Castillo, R. (2014). Women in Science and Technology: What Does the Literature Say?. Inter-American Development Bank.
- [52] Wozniak, D., Harbaugh, W. T., & Mayr, U. (2014). The menstrual cycle and performance feedback alter gender differences in competitive choices. *Journal of Labor Economics*, 32(1), 161-198.

Tables

TABLE A: EVALUTION DESIGN

		Group 1 (15 schools)		Group 2 (9 Schools)		Group 3 (12 schools)	
March 2013	Measurement	Baseline Data Collection					
Phase 1 Spring 2013	Training	Forward Looking B.		-		-	
Phase 2 Fall 2013	Training	Grit		Forward Looking B.		-	
Phase 3 End of May 2014	Training	-	FRM (9 schools)	-	FRM (3 schools)	-	FRM (4 schools)
June 2014	Measurement (Competition)	Follow-up Data Collection (2 Stage Competition Game)					

Note: 1 school from group 1, 2 schools from group 2 and 1 school from group 3 dropped out after phase 2 due to teachers' unforeseen transfers. The number of schools in this table reflects post-attrition.

Table 1: Randomization Balance

PANEL 1: Baseline Variables				
	Control	Grit	Grit+FRM	FRM
Male	0.54	-0.04 (0.02)	-0.03 (0.03)	-0.06* (0.04)
Raven	0.09	-0.06 (0.16)	0.06 (0.12)	-0.04 (0.18)
Math Grades	68.0	-2.35 (4.55)	0.42 (3.13)	0.41 (2.23)
Math Confidence	4.10	0.06 (0.10)	0.08 (0.10)	0.08 (0.10)
Risk Attitude	2.41	0.18 (0.19)	0.16 (0.21)	0.41 (0.26)
Male Teacher	0.33	-0.22* (0.11)	0.11 (0.21)	-0.13 (0.18)
Birth Order	2.18	0.11 (0.16)	-0.15 (0.13)	-0.06 (0.19)
Wealth	2.84	-0.09 (0.16)	0.12 (0.18)	0.06 (0.27)
Academic Success	3.44	-0.05 (0.11)	-0.04 (0.13)	0.10 (0.08)
Competitive	3.52	0.07 (0.05)	0.07 (0.07)	0.07 (0.07)
Easily Demoralized	0.25	0.05 (0.06)	0.01 (0.06)	0.00 (0.03)
PANEL 2: Within Task Variables				
	Control	Grit	Grit+FRM	FRM
Correct Answers in Part 1	5.35	-0.25 (0.60)	-0.49 (0.72)	-0.14 (0.42)
Performed Worse in Stage 1	0.47	0.06 (0.08)	0.04 (0.08)	0.02 (0.07)
Male Opponent	0.51	-0.00 (0.01)	-0.01 (0.01)	-0.01 (0.02)

Note: Each row reports coefficients from a regression of the variable shown in the first column on three treatment dummies, Grit, Grit+FMR and FRM. The first column reports the mean of the control, the second one reports the difference between the grit-only treatment and control, the third one between Grit+FRM and control, and the last one FRM only and control. Panel 1 presents the balance for demographic variables and baseline attitudes either reported by the child or the teacher. Panel 2 presents the balance for i) the performance in the first period, ii) the proportion of negative feedback in Stage 1 and iii) the porportion of male opponents. Standard errors, obtained via clustering at the school level, are reported in parentheses.

Table 2: Gender Gap in Competition Choice (Control Sample)

	1st Stage		2nd Stage	
	(1) W/O Controls	(2) With Controls	(3) W/O Controls	(4) With Controls
Male	0.046 (0.03)	0.024 (0.03)	0.105*** (0.03)	0.088** (0.04)
Male opponent		0.007 (0.04)		-0.004 (0.03)
Correct answers in part 1		0.019*** (0.01)		0.024*** (0.01)
Risk tolerance		0.047*** (0.02)		0.018 (0.02)
Raven Score		0.022 (0.02)		0.009 (0.02)
Male Teacher		-0.052 (0.07)		-0.014 (0.05)
Relative confidence S1		0.024*** (0.01)		
Performed Worse in S1				-0.146*** (0.03)
Relative confidence S2				0.035*** (0.01)
Observations	665	644	669	650

Standard errors bootstrapped and clustered at the classroom level. * p<0.10, ** p<0.05, *** p<0.01
 Reported estimates are marginal effects from logistic regressions.

Table 3: Gender Difference in Performance (Control Sample)

	1st Stage		2nd Stage	
	(1) W/O Controls	(2) With Controls	(3) W/O Controls	(4) With Controls
Male	0.087 (0.27)	-0.041 (0.16)	-0.156 (0.27)	-0.297 (0.24)
Male opponent		0.107 (0.20)		0.215 (0.14)
Correct answers in part 1		0.861*** (0.03)		0.670*** (0.03)
Risk tolerance		0.022 (0.05)		-0.080 (0.06)
Raven Score		0.272*** (0.08)		0.044 (0.11)
Male Teacher		0.212 (0.20)		-0.049 (0.34)
Relative confidence S1		0.094*** (0.04)		
Performed Worse in S1				-1.260*** (0.17)
Relative confidence S2				0.108*** (0.04)
Constant	6.736*** (0.58)	1.571*** (0.30)	6.334*** (0.57)	3.135*** (0.67)
Observations	666	647	665	648

Standard errors bootstrapped and clustered at the classroom level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
 Reported estimates are from OLS regressions.

Table 4: Feedback Type and Gender Gap in Competition Choice (Control Sample)

	Non-negative Feedback		Negative Feedback	
	(1) W/O Controls	(2) With Controls	(3) W/O Controls	(4) With Controls
Male	0.101** (0.04)	0.067 (0.04)	0.123** (0.05)	0.115*** (0.04)
Correct answers in part 1		0.049*** (0.01)		-0.002 (0.01)
Male opponent		0.003 (0.04)		-0.040 (0.04)
Risk tolerance		0.053*** (0.02)		-0.009 (0.02)
Raven Score		0.055** (0.02)		-0.031 (0.04)
Male Teacher		0.074 (0.05)		-0.079 (0.10)
Relative confidence S2		0.028*** (0.01)		0.041*** (0.01)
Observations	357	350	312	300

Standard errors clustered at the classroom level. * p<0.10, ** p<0.05, *** p<0.01

Reported estimates are marginal effects from logistic regressions of second stage choice conditional on feedback in Stage 1.

Table 5: Feedback Type and Gender Gap in Competition Choice (Control Sample)

	(1)	(2)	(3)	(4)
	NonNeg-PieceRate	Neg-PieceRate	NonNeg-Comp	Neg-Comp
Male	0.049 (0.06)	0.086** (0.04)	0.079* (0.04)	0.186* (0.11)
Observations	186	201	170	108

Standard errors clustered at the classroom level. * p<0.10, ** p<0.05, *** p<0.01

Reported estimates are marginal effects from logistic regressions of second stage choice conditional on choice and feedback in Stage 1.

Table 6: Treatment Effects on Competition Choice: Stage 1

	(1)	(2)	(3)
	All	Females	Males
Grit	0.065** (0.03)	0.094** (0.05)	0.039 (0.03)
Grit+FRM	0.150** (0.06)	0.142* (0.08)	0.159*** (0.05)
FRM	0.083* (0.05)	0.087 (0.06)	0.078* (0.05)
Male	0.029 (0.02)		
Male opponent	0.013 (0.02)	-0.001 (0.03)	0.031 (0.02)
Correct answers in part 1	0.018*** (0.00)	0.014*** (0.00)	0.021*** (0.00)
Risk tolerance	0.036*** (0.01)	0.045*** (0.01)	0.027** (0.01)
Raven Score	0.029** (0.01)	0.063*** (0.02)	-0.002 (0.02)
Male Teacher	-0.051 (0.04)	-0.008 (0.04)	-0.090** (0.04)
Control Mean	0.42	0.39	0.44
N	1920	933	987

Standard errors bootstrapped and clustered at the school level. * $p < 0.10$
 ** $p < 0.05$, *** $p < 0.01$. Reported estimates are marginal effects from
 logistic regressions

Table 7: Treatment Effects on Competition Choice: Stage 2

	(1)	(2)	(3)
	All	Females	Males
Grit	0.081** (0.03)	0.130** (0.05)	0.056 (0.05)
Grit+FRM	0.191*** (0.04)	0.154** (0.06)	0.249*** (0.03)
FRM	0.084** (0.04)	0.097** (0.05)	0.098* (0.06)
Male	0.065** (0.03)		
Male opponent	-0.008 (0.02)	-0.038 (0.03)	0.023 (0.04)
Correct answers in part 1	0.020*** (0.01)	0.023*** (0.01)	0.016*** (0.01)
Performed Worse in S1	-0.210*** (0.02)	-0.209*** (0.02)	-0.208*** (0.03)
Risk tolerance	0.018*** (0.01)	0.023** (0.01)	0.014 (0.01)
Raven Score	0.032** (0.01)	0.048** (0.02)	0.013 (0.02)
Male Teacher	-0.046 (0.03)	-0.019 (0.05)	-0.063** (0.03)
Control Mean	0.44	0.38	0.48
N	1926	934	992

Standard errors bootstrapped and clustered at the school level. * p<0.10
 ** p<0.05, *** p<0.01. Reported estimates are marginal effects from
 logistic regressions

Table 8: Treatment Effects Conditional on Feedback Type: Stage 2

	Non-negative Feedback			Negative Feedback		
	(1) All	(2) Females	(3) Males	(4) All	(5) Females	(6) Males
Grit	0.048* (0.03)	0.129*** (0.05)	-0.036 (0.04)	0.099** (0.05)	0.128* (0.07)	0.068 (0.07)
Grit+FRM	0.148*** (0.04)	0.147*** (0.05)	0.140*** (0.04)	0.220*** (0.06)	0.147 (0.12)	0.289*** (0.06)
FRM	0.116*** (0.04)	0.125*** (0.04)	0.121** (0.06)	0.032 (0.06)	0.059 (0.07)	-0.002 (0.08)
Male opponent	0.011 (0.02)	-0.014 (0.04)	0.040 (0.03)	-0.029 (0.04)	-0.065 (0.05)	0.002 (0.07)
Risk tolerance	0.024* (0.01)	0.026 (0.02)	0.022 (0.02)	0.014 (0.01)	0.021 (0.02)	0.009 (0.01)
Raven Score	0.067*** (0.02)	0.079** (0.03)	0.054*** (0.02)	0.005 (0.02)	0.024 (0.02)	-0.014 (0.02)
Correct answers in part 1	0.034*** (0.01)	0.036*** (0.01)	0.032*** (0.01)	0.003 (0.01)	0.007 (0.01)	0.000 (0.01)
Male Teacher	0.003 (0.03)	0.051 (0.04)	-0.039 (0.04)	-0.083* (0.05)	-0.069 (0.09)	-0.088* (0.05)
Control Mean	0.59	0.53	0.64	0.26	0.20	0.31
N	977	485	492	949	449	500

Standard errors clustered at the classroom level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
 Reported estimates are marginal effects from logistic regressions.

Table 9: Mitigating the Gender Gap

	(1) Stage 1	(2) Stage 2
Grit	0.087** (0.04)	0.138*** (0.04)
Grit+FRM	0.143 (0.09)	0.167*** (0.06)
FRM	0.079 (0.06)	0.111*** (0.04)
Male	0.039 (0.03)	0.099** (0.05)
Male opponent	0.014 (0.02)	-0.007 (0.02)
Risk tolerance	0.036*** (0.01)	0.018*** (0.01)
Raven Score	0.029** (0.01)	0.032** (0.01)
Correct answers in part 1	0.017*** (0.00)	0.020*** (0.01)
Male Teacher	-0.051 (0.04)	-0.044 (0.03)
Grit X Male	-0.042 (0.05)	-0.109* (0.07)
Grit+FRM X Male	0.015 (0.07)	0.052 (0.06)
FRM X Male	0.011 (0.07)	-0.052 (0.06)
Performed Worse in S1		-0.205*** (0.02)
Control Mean	0.42	0.44
N	1920	1926

Standard errors bootstrapped and clustered at the school level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Reported estimates are marginal effects from logistic regressions

Table 10: Treatment Effects on Payoffs

	(1)	(2)	(3)
	1st Stage	2nd Stage	Total
Grit	1.036** (0.41)	0.570 (0.56)	1.618* (0.89)
Grit+FRM	1.263** (0.62)	1.387 (0.94)	2.634* (1.47)
FRM	1.425* (0.77)	1.767* (0.90)	3.212** (1.55)
Male	-0.040 (0.29)	0.150 (0.28)	0.075 (0.53)
Male opponent	0.130 (0.23)	-0.049 (0.36)	0.083 (0.52)
Correct answers in part 1	2.103*** (0.09)	2.070*** (0.12)	4.174*** (0.20)
Risk tolerance	0.228* (0.12)	-0.007 (0.14)	0.214 (0.22)
Raven Score	0.952*** (0.22)	0.596*** (0.22)	1.571*** (0.42)
Male Teacher	0.389 (0.51)	-0.186 (0.35)	0.242 (0.77)
Constant	-2.996*** (0.67)	-2.572*** (0.69)	-5.588*** (1.13)
Observations	1917	1918	1907

Standard errors bootstrapped and clustered at the classroom level.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Reported estimates are from OLS regressions.

Table 11: Treatment Effects on Relative Confidence

	(1)	(2)	(3)
	All	Females	Males
Grit	0.182** (0.08)	0.299*** (0.10)	0.073 (0.10)
Grit+FRM	0.162* (0.10)	0.185* (0.11)	0.156 (0.13)
FRM	0.192** (0.08)	0.245* (0.13)	0.138 (0.09)
Male	0.157*** (0.05)		
Male opponent	-0.027 (0.06)	-0.084 (0.06)	0.033 (0.07)
Risk tolerance	0.005 (0.01)	-0.011 (0.02)	0.020 (0.02)
Raven Score	-0.002 (0.03)	-0.023 (0.04)	0.017 (0.03)
Correct answers in part 1	0.049*** (0.01)	0.053*** (0.01)	0.043*** (0.02)
Male Teacher	-0.026 (0.07)	-0.062 (0.10)	0.022 (0.07)
Observations	1892	931	961

Standard errors bootstrapped and clustered at the school level.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Reported estimates are from OLS regressions. Dependent variable is normalized so that estimated coefficients are standard deviation effects

Table 12: Treatment Effects on Gender Stereotypes: Within-Task

	(1)	(2)	(3)
	All	Females	Males
Grit	-0.036 (0.03)	0.020 (0.04)	-0.089** (0.04)
Grit+FRM	-0.043 (0.03)	0.019 (0.06)	-0.095** (0.04)
FRM	-0.033 (0.05)	0.019 (0.06)	-0.092 (0.08)
Male	0.138*** (0.02)		
Male opponent	0.017 (0.01)	0.024 (0.02)	0.011 (0.03)
Risk tolerance	0.002 (0.01)	-0.001 (0.01)	0.002 (0.01)
Raven Score	-0.012 (0.01)	0.019 (0.02)	-0.037* (0.02)
Correct answers in part 1	-0.001 (0.00)	-0.005 (0.01)	-0.000 (0.01)
Male Teacher	-0.007 (0.02)	-0.057 (0.04)	0.032 (0.04)
Control Mean	0.29	0.17	0.39
N	1922	936	986

Standard errors bootstrapped and clustered at the classroom level.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Reported estimates are marginal effects from logistic regressions where the dependent variable is a dummy that takes the value 1 if the respondent stated that a boy would do better than a girl in the incentivized question and zero otherwise.

Table 13: Treatment Effects on Gender Stereotypes: Outside of the Task

	(1)	(2)	(3)
	All	Females	Males
Grit	-0.127* (0.07)	-0.062 (0.08)	-0.185** (0.08)
Grit+FRM	-0.571*** (0.07)	-0.454*** (0.08)	-0.676*** (0.09)
FRM	-0.164 (0.10)	-0.070 (0.14)	-0.242** (0.12)
Male	0.628*** (0.05)	.	.
Baseline gender bias	0.454*** (0.03)	0.449*** (0.03)	0.455*** (0.04)
Risk tolerance	0.010 (0.02)	0.022 (0.02)	-0.005 (0.03)
Raven Score	-0.129*** (0.02)	-0.158*** (0.03)	-0.104*** (0.03)
Male Teacher	0.147*** (0.05)	0.068 (0.05)	0.230*** (0.08)
Constant	-0.240*** (0.08)	-0.254*** (0.08)	0.403*** (0.11)
Observations	1560	763	797

Standard errors clustered at the classroom level.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Reported estimates are from OLS regressions where the dependent variable is normalized genderbias score

Figure 1: Response to Positive and Negative Feedback

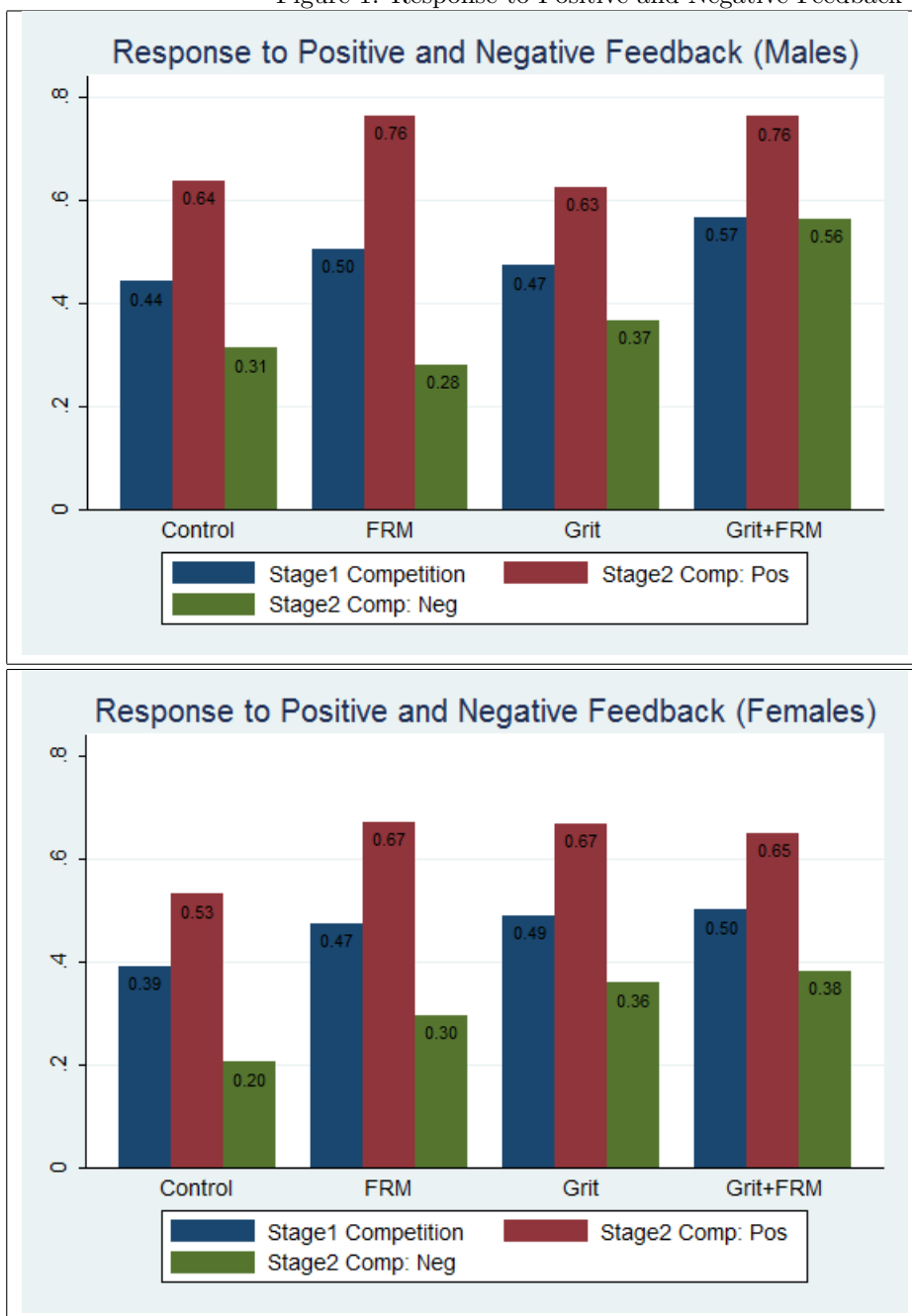
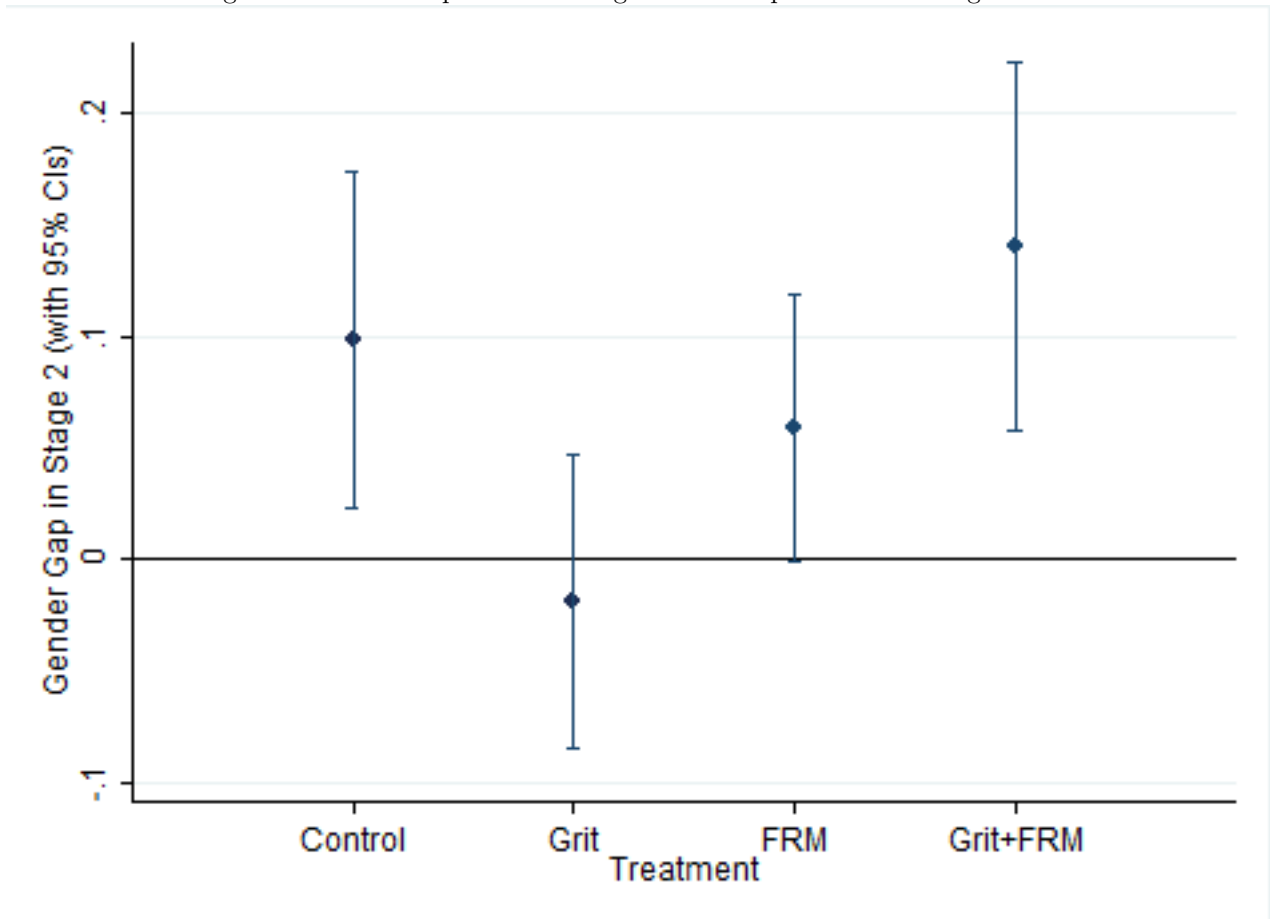


Figure 2: Gender Gap in the Willingness to Compete: Second Stage



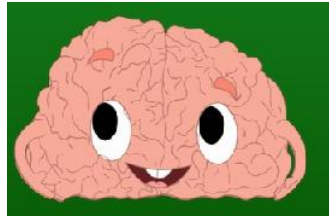
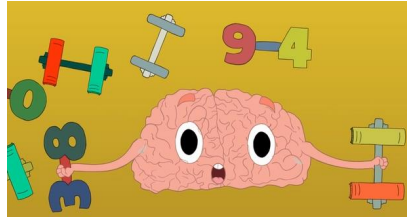
The Figure presents the marginal effects (and confidence bands) on “male” dummy obtained via logistic regressions separately for each treatment arm. Depicted coefficient estimates are obtained by controlling for initial performance, opponent’s gender, risk attitude, raven score, teacher’s gender and performance in the first stage. Confidence bands are obtained via bootstrapped and clustered standard errors.

Appendix

Education Materials

Poster below (a sample class activity) translates as: "I never quit".

Snapshots from Grit Treatment



Snapshots from FRM Treatment



Table 1: Treatment Effects on Competition Choice: Stage 1

	(1)	(2)	(3)
	All	Females	Males
Grit	0.082** (0.04)	0.115** (0.05)	0.056 (0.04)
Grit+FRM	0.170*** (0.06)	0.169* (0.09)	0.174*** (0.05)
FRM	0.139* (0.08)	0.162* (0.08)	0.114 (0.09)
Patience	0.051 (0.07)	0.078 (0.09)	0.025 (0.09)
Patience X FRM	0.058 (0.09)	0.040 (0.12)	0.081 (0.07)
Male	0.029 (0.02)		
Male opponent	0.014 (0.02)	0.001 (0.03)	0.032 (0.02)
Correct answers in part 1	0.017*** (0.00)	0.013*** (0.00)	0.020*** (0.00)
Risk tolerance	0.035*** (0.01)	0.045*** (0.01)	0.027** (0.01)
Raven Score	0.027** (0.01)	0.062*** (0.02)	-0.004 (0.02)
Male Teacher	-0.052 (0.04)	-0.008 (0.05)	-0.088** (0.04)
Percentage of Males	-0.008 (0.22)	-0.142 (0.21)	0.178 (0.26)
Control Mean	0.40	0.38	0.42
N	1920	933	987

Standard errors bootstrapped and clustered at the school level. * $p < 0.10$
** $p < 0.05$, *** $p < 0.01$. Reported estimates are marginal effects from
logistic regressions

Table 2: Treatment Effects on Competition Choice: Stage 2

	(1)	(2)	(3)
	All	Females	Males
Grit	0.071** (0.03)	0.118* (0.07)	0.019 (0.05)
Grit+FRM	0.180*** (0.04)	0.143* (0.08)	0.216*** (0.04)
FRM	0.075 (0.05)	0.093 (0.08)	0.042 (0.06)
Patience	-0.026 (0.05)	-0.029 (0.06)	-0.035 (0.09)
Patience X FRM	0.072 (0.07)	0.075 (0.07)	0.058 (0.09)
Male	0.065** (0.03)		
Male opponent	-0.008 (0.02)	-0.038 (0.03)	0.021 (0.04)
Correct answers in part 1	0.020*** (0.01)	0.024*** (0.01)	0.017*** (0.01)
Performed Worse in S1	-0.209*** (0.02)	-0.210*** (0.02)	-0.199*** (0.03)
Risk tolerance	0.018*** (0.01)	0.023** (0.01)	0.014 (0.01)
Raven Score	0.033** (0.01)	0.048** (0.02)	0.017 (0.02)
Male Teacher	-0.049 (0.03)	-0.023 (0.05)	-0.069* (0.04)
Control Mean	0.43	0.38	0.48
N	1926	934	992

Standard errors bootstrapped and clustered at the school level. * $p < 0.10$
** $p < 0.05$, *** $p < 0.01$. Reported estimates are marginal effects from
logistic regressions