Gender Differences Disappear with Exposure to Competition

Christopher Cotton

Frank McIntyre

Joseph Price*

September 2009

Abstract

Past research finds that males outperform females in competitive situations. Using data from multiple-round math tournaments, we verify this finding during the initial round of competition. The performance gap between males and females, however, disappears after the first round. In later rounds, only math ability (not gender) serves as a significant predictor of performance. Several possible explanations are discussed. The results suggest that we should be cautious about using data from one-round experiments to generalize about behavior. (*JEL* J16, C93)

Keywords: competitiveness; gender differences; field experiment

*Cotton: Department of Economics, University of Miami, Coral Gables, FL 33146; cotton@miami.edu. McIntyre and Price: Department of Economics, Brigham Young University, Provo, UT 84602; mcintyre@byu.edu, joe_price@byu.edu.

1 Introduction

In their important paper, Gneezy et al. (2003) show that males and females respond differently to competition. The authors conduct experiments in which college students are paid to solve mazes, either on their own or in competition with others. They show that competition causes males to increase their performance relative to females. Gneezy and Rustichini (2004) study footraces between fourth graders and find a similar result: boys respond favorably to competition, while girls do not.[1] These results imply that males perform better in competitive settings than otherwise similar females,[2] and the results may help explain achievement differences between males and females in professional and academic settings.[3]

These past papers consider one-round competitions. It is reasonable, however, that the gender differences do not persist over multiple rounds of competition. Males, for example, may have more competitive (e.g., sports) experience and find the idea of competing less intimidating or more exciting than females. If this is the case, females may improve their performance over time as they become more comfortable with the competitive environment, or male performance may fall as their enthusiasm dissipates. We attempt to better understand gender differences in competition by considering how robust the earlier gender-gap results are to multiple rounds of participation. As competitors gain experience, do the gender differences persist or do they disappear?

We hold a series of multiple-round math competitions in which participants race against a randomly assigned opponent to complete math problems. Controlling for math ability (based on prior test scores), males perform over 0.4 standard deviations better than females

[1] Inzlicht and Ben-Zeev (2000) find evidence that one's performance on difficult verbal and math tests may depend on the gender composition of the group of people sitting in close proximity, even when they are not directly competing. Antonovics et al. (2009) showed that males were more likely to answer trivia questions correctly when a larger fraction of their competitors were female. In the present analysis, we focus on whether males perform better than females, on average, in competitive environments. We are less concerned with whether one's performance depends on competitor gender. Although we acknowledge that players may put in more or less effort when competing against males or females, we find that any intra-gender performance gap is much less significant than the inter-gender performance gap.

[2] Gneezy et al. (2003) show that males solved more mazes when they were paid based on relative performance than when they were paid piece-rate. The effect of competition on females was less pronounced, although there was a more modest positive effect when females competed against other females. In both Gneezy et al. (2003) and Gneezy and Rustichini (2004), there were no statistically significant differences between the performances of females who competed against males and females who competed against females, or between the performances of males who competed against males and males who competed against females. Based on the significance of the results, we can conclude from these papers that males respond better to competition than similar females (although we cannot say how the results were affected by opponent gender).

[3] For example, only 2.5% of the highest-paid U.S. executives are women (Bertrand and Hallock (2001); for an overview of the gender pay gap, see Blau and Kahn (2000)). Kleinjans (2009) presents evidence that differences in male-female taste for competition may help explain a portion of the sorting into different professions. Price (2008) shows that an increase in competition for funding in graduate school causes males to increase their performance relative to females.

during the first round of competition. This result is consistent with the findings from the one-round experiments in the literature.

During the first round of competition, one's gender and ability, as well as opponent ability, are all significant predictors of performance. In round two and onward, *only* math ability serves as a significant predictor of performance; gender and opponent ability no longer matter. After the first round, the performance gap disappears and females perform just as well as males.

The short-lived nature of the gender performance gap is consistent with some types of gender differences, but not others. Because the gender gap disappears, the data do not support an explanation in which males and females differ in significant, permanent ways (e.g., males are always better at dealing with the stresses of competition, care more about winning, or get more enjoyment from competing).[4] The fact that the gender differences vanish after only one round (rather than gradually) does not lend support to models in which participants learn about their relative ability or build self-confidence over time.[5] Furthermore, we find evidence that the initial gender gap is caused by male overperformance rather than female underperformance. This means the data do not support an explanation in which females are hesitant to compete. Rather, the evidence is consistent with a model in which males are initially more enthusiastic about competing than females, but this overexcitement quickly disappears with exposure to competition.[6]

In later rounds of competition, both males and females steadily improve their perfor-

[4] Even if males are no better than females at performing outside of competition, males might still have an ability advantage over females if they can better deal with the stresses of competition. Only by considering multiple rounds of competition are we able rule out this possibility.

[5] Gneezy et al. (2003) suggest that females underestimate female ability relative to male ability. However, this conclusion is based on the observation that females respond more to competition when they are competing against other females than when they are competing against males. Not only is this observation not significant (p-value = 0.1025), it is not replicated in Gneezy and Rustichini (2004), and does not provide sufficient evidence to rule out other explanations of the performance gap. For example, their data are also consistent with a model in which females tend to be more nervous when competing against males than when competing against other females.

[6] This is consistent with Niederle and Vesterlund (2007) who show that, given a choice, males are more likely to compete than females. Similarly, Sapienza et al. (2009) claim that at Northwestern University, 36% of female versus 57% of male MBA students choose competitive finance careers. Additionally, Booth and Nolen (2009) and Gneezy et al. (forthcoming) show that preferences towards competition may be due to past experience. Specifically, Booth and Nolen (2009) show that females who attend all girls schools are more likely to choose competition (even when competition was against males) than were females in coed schools. Gneezy et al. (forthcoming) show that in a matrilineal society, women prefer competition more than males. Dreber et al. (2009) find no gender differences in female activities including skipping rope and dancing. Günther et al. (2008) find no ascertainable difference between male and female performance when participants compete solving word games, a task they claim is gender neutral. They argue that the gender differences go away because there are no gender stereotypes associated with completing the task. Our results suggest an alternative explanation in which the gender differences would go away because males are less likely (or females are more likely) to be overeager to compete in gender-neutral or female activities.

mance. This is consistent with a model in which participants learn to compete or become more comfortable with competition over time. This gradual improvement in the later rounds is present for both males and females, and cannot explain the first round gender differences. We find no evidence that either males or females perform better in the long run.

The results advance our understanding of gender differences in competition. The findings also have implications for experimental research more generally. We show that performance during the first round of our experiment is very different from performance after the first round. This suggests that there may be value to repeating experiments more than once, and that we need to be careful about using results from one round to generalize about behavior.

Section 2 describes our field experiment in detail. In Section 3, we conduct the analysis and discuss possible explanations for the gender differences. Section 4 concludes with a discussion of the results, policy implications, and directions for future research.

2 Math Competitions

Working with school officials and teachers, we held a series of in-classroom math competitions with a total of 185 fourth grade students. In each of eight classrooms, students participated in up to five rounds of competition.[7] In each round, each student was randomly paired with another student, and then raced against this opponent to complete a series of math questions drawn from past fourth grade state assessments. The competitions occurred in the spring, a few weeks before students took the state assessment. This assured the questions were grade appropriate and, because the students had been together all year, meant that students were familiar with one another and were likely to have some idea of the math ability of the other students.

In each round, students were given a sheet with 10 multiple choice questions. They had five minutes to solve as many questions as possible. If a student finished before the five minutes were up, he or she could "buzz in" early. The student with the most correct answers in each pair won the round. If both students had the same number of correct answers, the student who buzzed in first won the round. Winning a round entitled the student to two raffle tickets. In the case of a tie when neither student buzzed in early, both students received one ticket. At the end of the tournament, we randomly drew three raffle tickets for each classroom and the winners received candy bars.[8]

[7]Four classes participated in five rounds of competitions. Due to time constraints, two classes had only three rounds and two classes had only four rounds.

[8]Our tournaments took place a few weeks before the state assessment tests and were used as a way of preparing the students for those tests. If a student buzzed in early, he or she was required to stop working; the opponent could continue working until the end of the five minutes. There were no penalties for incorrect

The school district provided prior-year (third grade) standardized math test results for all students except those who recently moved into the district or for some other reason did not take the math assessment the year before. In total, we have test performance data for ninety percent of the competitors. The state uses its own criterion-referenced test for its end-of-year assessment. The un-normalized student state assessment scores in our sample have a mean of 172 (standard deviation of 12) with no significant differences in the scores between boys and girls.

Although parents could opt out from their child participating, the competitions were presented as school activities, and no one opted out. We are, therefore, unconcerned about selection bias.

3 Identifying Gender Differences

3.1 Empirical Strategy

Let $y_{i,r}$ be student i's score in round r, where all scores are normalized by round and class to be mean 0 and standard deviation 1.[9] Let $\bar{y}_{i,-r}$ be i's average score over the other four rounds of competition, a noisy measure of ability. The average score of i's round r opponent (in the opponent's other rounds) is $\bar{y}_{-i,-r}$. We run regressions of the following form:

$$y_{i,r} = \alpha + \beta_1 \bar{y}_{i,-r} + \beta_2 \bar{y}_{-i,-r} + g(i, -i) + \epsilon_{i,r}.$$

The function $g(i, -i)$ is used to capture gender effects based on the student's gender and that of his or her opponent. As written, the OLS regression may be inconsistent due to measurement error in both ability scores. We address this by instrumenting student ability and opponent ability measures with the student and opponent's normalized scores on the prior year state assessment test.[10]

Table 1 reports the results from the IV analysis. For those interested, Table 2 provides results on the male-female gap from several alternate specifications. There we consider OLS specifications using state assessment scores for our two ability measures. We also run regressions excluding some or all of the ability controls entirely or allowing the ability

answers. After each round of competition, the students graded each other's test and each pair agreed upon the outcome of the round. We observed no collusion or cheating among the students.

[9]We normalize scores for the regressions to make results easily comparable to the education research on tests scores and to remove from the analysis classroom idiosyncrasies in ability or testing environment. As we show in Table 2, regressions using raw scores or when normalized just by round produce the same substantive results.

[10]The instrumental variables specification is not needed to correct any simultaneity bias, it is simply a correction for the measurement error inherent in measuring math ability with finite resources.

coefficients to differ by gender. Lastly, we consider different normalizations of the competition data. In all of these alternative specifications the estimated male-female gap is negative and significant in the first round, and small and never significantly negative in the later rounds.

In addition to these regressions, we consider data regarding round-by-round scores by both gender and ability. This data allows us to distinguish between male over-performance and female under-performance in the initial round of competition. We also use this data and the regression results to better understand changes in the performance of both males and females over the later rounds of competition.

3.2 First Round

We first identify significant gender differences in performance during the first round of competition. This gender gap is consistent with the literature using one-round competitions.

In Table 1 results, regression 1 controls for a participant's own gender. Regression 2 controls for both participant and opponent gender, considering four distinct cases: male subject against male opponent (MvM), male subject against female opponent (MvF), female subject against male opponent (FvM), and female subject against female opponent (FvF).

There is a significant and very large negative effect of being female on performance. In regression 1, the gap between female and male competitors is provided by the coefficient on the Female dummy variable, which shows females doing 0.44 standard deviations worse than males of similar ability. The first round performance gap between males and females is highly significant (p-value = 0.007).

We use regression 2 to test whether opponent gender affects performance. Opponent gender does not appear to have a significant impact on performance. That is, there is no significant difference between the FvM and FvF coefficients, or between the MvF coefficient and the MvM baseline. Therefore, although females perform significantly worse than males, their performance does not appear to be influenced by the gender of their opponent.

In both regressions 1 and 2, the coefficient on opponent ability is significant. That is, one's performance is decreasing in opponent ability. We discuss this result further in Section 3.4.

3.3 Multiple Rounds

After the first round of competition, the gender difference disappears. In the second round of competition, and in all subsequent rounds, only math ability significantly affects performance.

Regressions 3 and 4 are identical to the first two regressions, except they now consider the second round of competition. Regressions 5 and 6 pool the data from rounds two through five.[11] In regressions 3 and 5 we see that the female effect (which now has a positive coefficient) is insignificant (p-values = 0.92 and 0.21). We can also formally reject that the female effect is the same between rounds 1 and 2 (p-value = 0.04). Regressions 4 and 6 control for opponent gender and continue to find no gender effects. We can statistically reject the hypothesis that the female effect is the same in regressions 1 and 5 (p-value = 0.02). We find no evidence that females continue to perform worse than males after the first round of competition.

A few other patterns in the multiple round data are worth mentioning. From Table 1, we see that the coefficient on ability goes from 0.46 in round one, to 0.68 in round two, to 1.30 in the pooled data for rounds two through five. The difference between the round 1 and 2 ability coefficients is not significant, but the difference between round 1 and the pooled rounds 2 through 5 is significant (p-value = 0.03). Furthermore, the coefficient on opponent ability is negative and significant in the first round; after the first round it is insignificant, and it becomes smaller over time; although the results are not precise enough to reject the hypothesis that the coefficient is the same across all specifications.

Figure 1 shows the average number of correct answers (out of a possible 10) by round by gender. The data show that females on average perform just as well in the second round as they do in the first round. Males, however, perform significantly better in the first round than in the second. Formally, we cannot reject a hypothesis that female scores in the first two rounds come from the same distribution (p-value = 0.41), and we can reject a hypothesis that male scores in round one come from the same distribution as round two scores (p-value = 0.004). After the second round, the performance of both males and females steadily improves over the remaining rounds.

Additionally, we consider whether the effect of experience depends on relative ability. To do so, we divide our subjects into three equal-sized categories based on prior-year standardized test scores: low-ability, middle-ability, and high-ability. Figure 2 shows the average number of correct answers by ability level for each round of competition. Low-ability participants, for example, correctly answer an average of 3.6 (out of a possible 10) questions during the first round of competition and improve to 3.8 correct answers by the final round, an eight percent improvement. High-ability participants do slightly better in the first round of competition answering an average 4.3 questions correctly. By the final round of competition, high-ability participants answer an average of 7.3 questions correctly, a 67 percent improve-

[11] In unreported results we ran individual regressions by round. For rounds 2 through 5 the gender gap was never significantly different from zero.

ment. We test if the changes between rounds are the same across ability groups, and reject this hypothesis (p-value = 0.004). (Similar results are found if we look at round 2 performance rather than round 1 performance; thus, the initial gender differences are not driving these findings.) These results will help when we work to distinguish between alternate types of gender differences in Section 3.4.[12]

In summary, the analysis finds the following results:

1. Males outperform females in the first round of competition, but not thereafter.
2. Males perform better in the first round than in the second round; females perform equally well in the first and second round.
3. Both males and females steadily improve over rounds two through five.
4. High ability participants improve more than low ability participants in rounds two through five.
5. The coefficient on ability increases and becomes more significant, and the coefficient on opponent ability decreases and becomes less significant over multiple rounds.

3.4 Explaining Gender Differences

The primary goal of this analysis is to better understand gender differences in competition. Prior analyses show that during a single-round of competition, males outperform females. Possible explanations for this performance gap included (i) males being inherently better competitors than females, (ii) males being overconfident or females being underconfident in their respective ability, (iii) females being more nervous or intimidated by competition than males, and (iv) males being more eager to compete than females. In the appendix, we provide a game theoretic model of competition and show that each of these types of gender differences can result in males outperforming females in a single round of competition, but the four possibilities have different implications about performance in the long run. All of these possibilities cannot explain the short-lived nature of the gender differences in our data.

The disappearance of gender differences suggests that the male-female performance gap is not caused by inherent differences in ability or preferences. After the first round of competition, we find no evidence that males are better competitors than females. Thus, we find no support for the possibility that males are inherently better competitors than females.

[12] As discussed earlier, some schools only competed in 3 or 4 rounds. Excluding these classes yields similar figures and the same patterns emerge.

If the gender performance gap resulted from differences in self confidence, then we expect the performance gap to gradually disappear as participants gain experience. However, we observe that the gender gap does not gradually disappear, but rather vanishes after only one round of competition. The sudden disappearance of the gender differences is not consistent with a model of learning. It's unreasonable to believe that females, for example, remained underconfident in their abilities through all of their past experiences, only to completely overcome this underconfidence during one round of math competition. Thus, we find no support for an explanation in which gender differences result from differences in self confidence.

Explanations in which females are more nervous or males are more eager to compete are both consistent with the short-lived nature of the gender differences. It is reasonable that female nervousness or male excitement settles down after one round. We show, however, that female performance is unchanged between the first and second round, suggesting that female nervousness is not causing the performance gap in the first period. On the other hand, male performance decreases between the first and second round of competition lending support to the explanation in which males are more excited to compete during the first round, and that this eagerness dissipates before the second round of competition.

This finding has interesting implications. It suggests that the initial gender gap is due to male overperformance rather than female underperformance. The high male performance may result because males are more excited about competing than females. Males may overestimate their enjoyment from competition; they are "quick out of the gate" at the start of the competition, but then their enthusiasm about competing quickly settles down to the same level as females. This model is consistent with other papers showing that males are overeager to compete (e.g., Niederle and Vesterlund 2007).

3.5 Effect of Experience in Long Run

A secondary goal of the analysis is to understand how competitive performance changes with experience. We observe a steady increase in the performance of both males and females after rounds two.[13] Possible explanations for this include participants becoming better competitors, more confident, or more comfortable (i.e., less nervous) over the course of the competition. We argue that the data is most consistent with a model in which competitors

[13]Tables 1 and 2 show that the female coefficient small insignificant in round 2 and in the pooled rounds 2-5. This is true for the IV specification in Table 1, as well as in 11 of the 12 round 2-5 robustness checks in Table 2. The only exception is when we exclude opponent ability in the IV analysis, in which case the female coefficient for pooled rounds 2-5 is positive and significant, yet relatively small. However, given that the competitors in our experiment have spent almost an entire school year in the same classroom as their opponents, we do not put much faith in this particular regression relative to each of our other regressions in which the coefficient was insignificant.

become more comfortable over time, although we cannot rule out the other possibilities.

In Table 1 the coefficient on ability increases and becomes more significant in rounds two through five. Furthermore, the coefficient on opponent ability becomes less significant over multiple rounds. These pieces of evidence are consistent with participants becoming more focused. One reason for this may be if participants are becoming less nervous. Furthermore, our finding that high-ability participants improve more than low-ability participants over the course of the competition is also consistent with an explanation in which participants become more comfortable with competition over time. In particular, the evidence is consistent with the psychology literature, which shows that high-ability people are more likely to be effected by nerves and to choke under pressure than low-ability people (e.g., Beilock and Carr 2005, show this in math competitions).

4 Conclusion

Using data from multiple-round math competitions, we identify a performance gap between males and females during the first round of competition; a result that is consistent with the past literature. We then show that these differences disappear after the first round of competition. Through the later rounds of competition, we find no evidence that males outperform females of similar ability.

The analysis considers various explanations for the initial–but transitory–gender differences in competition. We find no evidence that males and females differ in their ability to compete, or that the gender gap is caused by female underconfidence of male overconfidence. Rather, the evidence is consistent with an explanation in which males initially tend to be overeager to compete.

The results suggest that programs designed to encourage female participation in competition may increase the number of females entering competitive fields of employment. While not specifically about entry into competitive fields, Stevenson (forthcoming) finds that the increase in sports participation that accompanied the passage of Title IX led to a 15 percent increase in the fraction of women entering traditionally male occupations.[14] According to our results, however, the initial gender gap and the sorting into different fields may be due to males overestimating their enjoyment from competition (not from differences in competitive ability, self confidence, or comfort in competitive settings). This means that Title IX may encourage more females to enter competitive fields by increasing female enthusiasm towards competition.

[14]Title IX of the Education Amendments of 1972 requires that U.S. schools provide equal funding and attention for both female and male sports teams.

At a minimum, our results suggest that we are far from a complete understanding of gender differences in competition. Future work is needed to determine to what extent the one-round gender differences may affect sorting of males and females into different professions, or long run achievement in competitive settings. If, for example, males are always overly eager to compete, then the gender differences may result in large long-run differences in performance. Future work is also needed to determine the extent to which the (round-one) gender differences identified here and elsewhere may be eliminated through repeated exposure to competitive experience. For example, can encouraging sports participation amongst females help eliminate the initial gender differences? It would also be interesting to consider whether our findings carry over to other activities. The gender differences may go away after one round because the math competition was less enjoyable than males expected; maybe other activities turn out to be more enjoyable than expected.

Our results also have implications for experimental research more generally, as they suggest that there may be value to repeating experiments more than once. Our findings in rounds two and onward in our experiment are vastly different from our first round results. This suggests there is risk in using first round results to generalize about behavior, something that is often done in the experimental literature.

5 Appendix

5.1 Game Theoretic Model of Competition

Consider the following model of competition. Two players, $i \in \{1,2\}$, compete for a prize. The function A describes a player's overall ability, where i's ability depends on both his or her math ability m_i and competitive ability a_i (e.g., ability to focus under pressure). For player i, $A(m_i, a_i) > 0$, $\partial A/\partial m_i > 0$, and $\partial A/\partial a_i > 0$ for all possible m_i and a_i.

Both players simultaneously choose their level of effort $e_i > 0$, with the likelihood of winning the prize dependent on both players' effort and ability. Let $\theta_i(e; m, a)$ denote the probability i wins the prize given $e = \{e_1, e_2\}$, $a = \{a_1, a_2\}$ and $m = \{m_1, m_2\}$, where

$$\theta_i(e; a) \equiv \frac{A(m_i, a_i)e_i}{A(m_i, a_i)e_i + A(m_{-i}, a_{-i})e_{-i}}.$$

We refer to $p_i \equiv A(m_i, a_i)e_i$ as i's performance, and the subscript $-i$ refers to i's opponent.[15] Players care about winning the prize and find effort costly. If player i wins the prize, he or she

[15] The formulation of the model incorporates uncertainty about performance in the competition through $\theta_i(e; m, a)$ being between 0 and 1. Even if a player knows that she is higher ability than her opponent, there is still a positive probability that her opponent wins in equilibrium.

earns $U_i = v - c_i e_i$, where $v > 0$ is the value of winning and $c_i > 0$ is the cost of effort. If i does not win the prize, $U_i = -c_i e_i$.

We solve for the symmetric subgame perfect Nash equilibrium under complete information.

Player i chooses effort e_i to maximize his expected payoff

$$EU_i = \frac{A(m_i, a_i)e_i}{A(m_i, a_i)e_i + A(m_{-i}, a_{-i})e^*_{-i}(m, a, v, c)} v - c_i e_i,$$

where $e^*_{-i}(v, a, c)$ denotes the opponent's equilibrium effort choice given m, a, $c = \{c_1, c_2\}$, and v. In equilibrium, i chooses effort

$$e^*_i(v, a, c) = \frac{A(m_i, a_i)A(m_{-i}, a_{-i})c_{-i}v}{(A(m_i, a_i)c_i + A(m_{-i}, a_{-i})c_{-i})^2}, \tag{1}$$

and his performance equals

$$p^*_i(v, a, c) = \frac{A(m_i, a_i)^2 A(m_{-i}, a_{-i})c_{-i}v}{(A(m_i, a_i)c_i + A(m_{-i}, a_{-i})c_{-i})^2}. \tag{2}$$

Notice that p^*_i is strictly increasing in both m_i and a_i, and strictly decreasing in c_i. Furthermore, $p^*_i > p^*_{-i}$ if and only if $A(m_i, a_i) > A(m_{-i}, a_{-i})$; if both i and $-i$ have the same math ability, then $p^*_i > p^*_{-i}$ if and only if $a_i > a_{-i}$.

We can use this framework to consider the possible types of gender differences from Section 3.4. We discuss cases (i), (iii), and (iv), which each involve *actual* differences in the model's parameters, before discussing case (ii), which involves *perceived* differences in parameters. First, suppose that males are inherently better competitors than females. The is possibility (i) from the text. In this case, males tend to have higher a_i than females. Then, assuming both males and females have similar m and c, males will tend to perform better than females. In this case, the differences in a are inherent and should persist over multiple rounds of competition. Females may learn to compete, but we expect this learning to require multiple rounds of experience.[16]

Alternatively, females may tend to be more nervous than males. This is possibility (iii) from the text. In this case, nerves may result in females starting the competition with a lower a_i than males. However, once females become used to the competition, they will no longer be at a disadvantage. That is, if females are more nervous than males, we expect that females will initially perform worse than males, but that this performance difference may disappear after exposure to at least one round of competition. This possibility, however, predicts that the gender differences are eliminated due to an increase in female performance.

If males tend to be more excited about competing, they initially have a lower c_i than females. This is possibility (iv) from the text. The predictions are similar to the case of nerves, as the excitement may dissipate after the first round. However, this model predicts that the gender

[16]If males and females have the same ability, but instead differ in terms of their value of winning or cost of effort, the model will continue to produce similar results.

differences are eliminated due to a decrease in male performance, rather than a change in female performance.

A final possibility is that females and males differ in their self confidence. This is possibility (ii) from the text. Suppose that player i underestimates her own $A(m_i, a_i)$, which may be because she underestimates either m_i or a_i. Player $-i$ has correct beliefs about his ability. To keep the example as straightforward as possible, we assume that neither player knows that player i may underestimate her ability. Let $\tilde{A} < A(m_i, a_i)$ denote i's beliefs about her ability. In this case, i puts in effort $e_i^* = \tilde{A}A(m_{-i}, a_{-i})v/((\tilde{A} + A(m_{-i}, a_{-i}))^2 c)$ achieving performance $p_i^* = \tilde{A}A(m_i, a_i)A(m_{-i}, a_{-i})v/((\tilde{A} + A(m_{-i}, a_{-i}))^2 c)$. Player $-i$ puts in effort and achieves performance according to equations 1 and 2. The player who underestimates her own ability puts in lower effort and achieves lower performance than she otherwise would. If males and females have the same actual ability, but females underestimate their ability, then females will tend to perform worse than males. They may learn about their true ability over time, but again, we expect this learning requires multiple rounds of experience.

References

Antonovics, Kate, Peter Arcidiacono, and Randall Walsh, "The Effects of Gender Interactions in the Lab and Field," *Review of Economics and Statistics*, 2009, *91* (1), 152–162.

Beilock, Sian L. and Thomos H. Carr, "When High-Powered People Fail: Working Memory and 'Choking Under Pressure' in Math," *Psychological Science*, 2005, *16*, 101–105.

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

Blau, Francine and Larry Kahn, "Gender Differences in Pay," *Journal of Economic Perspectives*, 2000, *14* (4), 75–99.

Booth, Alison L. and Patrick J. Nolen, "Choosing to Compete: How Different Are Girls and Boys?," February 2009. IZA Discussion Paper Series, No. 4027.

Dreber, Anna, Emma von Essen, and Eva Ranehill, "Outrunning the Gender Gap: Boys and Girls Compete Equally," 2009. SSE/EFI Working Paper Series in Economics and Finance, No. 709.

Gneezy, Uri and Aldo Rustichini, "Gender and Competition at a Young Age," *American Economic Review*, 2004, *94* (2), 377–381.

_ **, Kenneth Leonard, and John A. List**, "Gender Differences in Competition: Evidence from a Matrilineal and Patriarchal Societies," *Econometrica*, forthcoming.

_ , **Muriel Niederle, and Aldo Rustichini**, "Performance in Competitive Environments: Gender Differences," *Quarterly Journal of Economics*, 2003, *118* (3), 1049–1074.

Günther, Christina, Neslihan Arslan Ekinci, Christiane Schwieren, and Martin Strobel, "Women can't jump? An experiment on competitive attitudes and stereotype threat," 2008. Universitat Heidelberg working paper.

Inzlicht, Michael and Talia Ben-Zeev, "A Threatening Intellectual Environment: Why Women Are Susceptible to Experiencing Problem-Solving Deficits in the Presence of Males," *Psychological Science*, 2000, *11* (5), 365–371.

Kleinjans, Kristin J., "Do gender differences in prefereces from competition matter for occupational expectations?," *Journal of Economic Psychology*, 2009, *30* (5), 701–710.

Niederle, Muriel and Lise Vesterlund, "Do Women Shy Away from Competition? Do Men Compete too Much?," *Quarterly Journal of Economics*, 2007, *122* (3), 1067–1101.

Price, Joseph, "Gender Differences in Response to Competition," *Industrial and Labor Relations Review*, 2008, *61* (3), 320–333.

Sapienza, Paola, Luigi Zingales, and Dario Maestripieri, "Gender differences in financial risk aversion and career choices are affected by testosterone," *Proceedings of the National Academy of Sciences*, 2009. http://www.pnas.org/cgi/doi/10.1073/pnas.0907352106.

Stevenson, Betsey, "Beyond the Classroom: Using Title IX to Measure the Return to High School Sports," *Review of Economics and Statistics*, forthcoming.

Table 1: Instrumental Variable Regressions of Student Score on Ability and Gender

	Round 1		Round 2		Rounds 2-5	
	(1)	(2)	(3)	(4)	(5)	(6)
Ability	0.46*	0.47*	0.68***	0.65***	1.30***	1.30***
	[0.27]	[0.27]	[0.20]	[0.20]	[0.13]	[0.13]
Opponent Ability	-0.48*	-0.47*	-0.19	-0.19	-0.07	-0.07
	[0.25]	[0.24]	[0.19]	[0.19]	[0.14]	[0.13]
Female	-0.44***		0.02		0.08	
	[0.17]		[0.15]		[0.06]	
Female: Male		-0.49*		0.04		0.02
Opponent		[0.26]		[0.17]		[0.10]
Female: Female		-0.44*		0.17		0.1
Opponent		[0.26]		[0.20]		[0.10]
Male: Female		-0.04		0.17		-0.05
Opponent		[0.25]		[0.21]		[0.13]
Constant	0.15	0.17	0.01	-0.07	-0.03	-0.01
	[0.12]	[0.21]	[0.10]	[0.12]	[0.05]	[0.07]
Observations	152	152	150	150	516	516

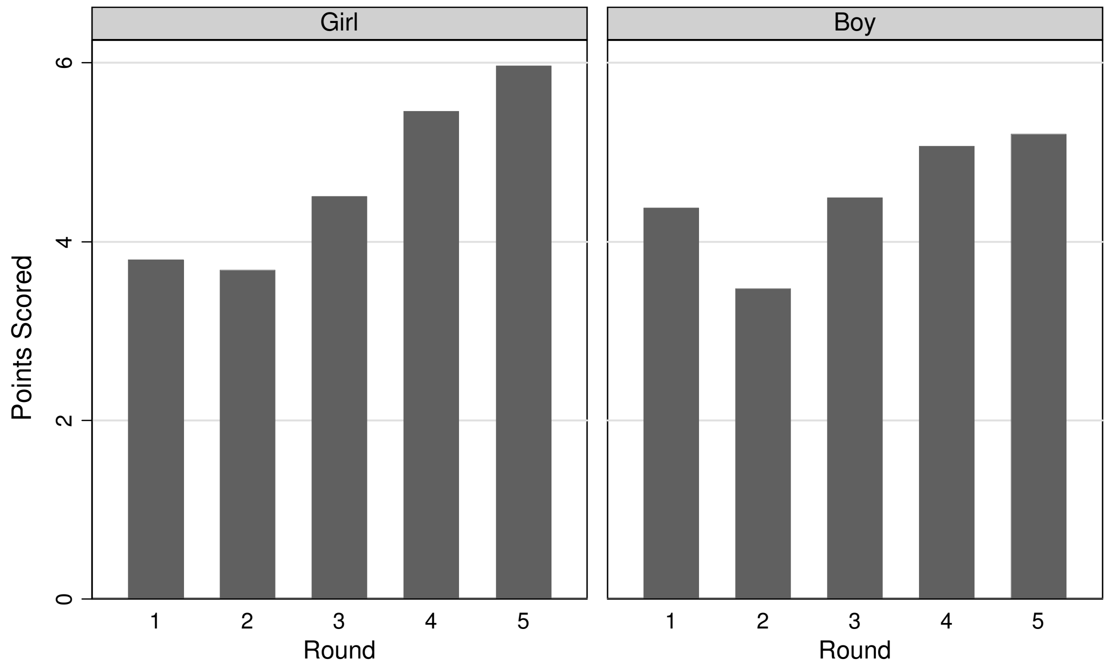
Notes: Dependent variable is the student's normalized score. Robust standard errors in brackets are clustered by student in regressions (5) and (6). Unreported non-robust standard errors are similar. * significant at 10%, ** significant at 5%, *** significant at 1%. Ability and Opponent Ability are the student's average score across the other rounds of the tournament. Both are instrumented by student and opponent state assessment scores. Observation numbers change between first and second rounds due to students switching opponents, some of whom are missing data on the state assessment test score instrument. First stage F-statistics for excluded instruments are 24.5 and 25.1 in regression (1) and the same or larger in other rounds.

Table 2: Robustness of Table 1 Results -- Female Coefficient

	OLS			IV		
	Round 1	Round 2	Rounds 2-5	Round 1	Round 2	Rounds 2-5
Baseline Regression	-0.35**	-0.09	-0.05	-0.44***	0.02	0.08
	[0.15]	[0.16]	[0.11]	[0.17]	[0.15]	[0.06]
Exclude Opponent Ability	-0.36**	-0.02	-0.01	-0.36**	0.06	0.11*
	[0.15]	[0.15]	[0.10]	[0.14]	[0.13]	[0.06]
Exclude Both Ability Controls	-0.30**	0.09	0.08	-	-	-
	[0.14]	[0.15]	[0.11]			
Allow Ability Coefficients to Differ by Gender	-0.36**	-0.09	-0.05	-0.44**	0.01	0.09
	[0.16]	[0.16]	[0.10]	[0.17]	[0.15]	[0.07]
Normalize Scores by Round Only	-0.28*	-0.07	0.02	-0.35**	-0.03	0.08
	[0.16]	[0.16]	[0.12]	[0.16]	[0.14]	[0.06]
Use Raw Scores for Dep. Var.	-0.64*	-0.18	0.02	-0.81**	0.05	0.34
	[0.37]	[0.42]	[0.32]	[0.38]	[0.39]	[0.27]

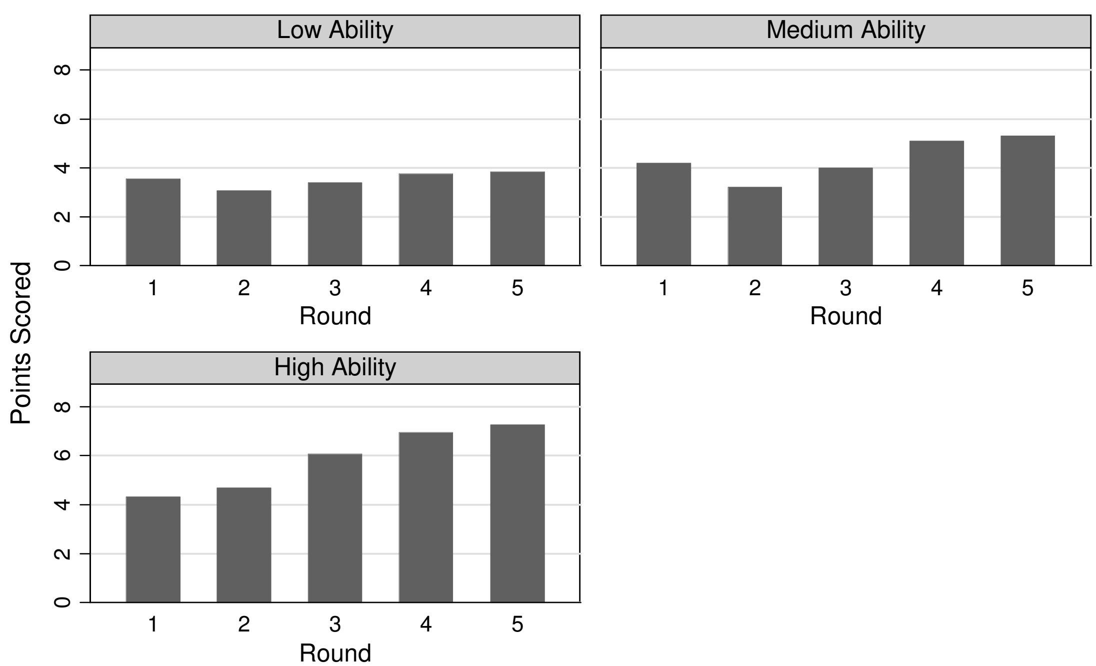
Notes: Reported values are the coefficient and standard error on the "Female" dummy variable in regressions of the form reported in Table 1. Robust standard errors in brackets are clustered by student for Rounds 2-5. * significant at 10%, ** significant at 5%, *** significant at 1%. OLS regressions control for ability and opponent ability using prior year state assessment math test. IV regressions instrument student performance in other rounds with prior year state assessment math test. Since ability covariates are mean zero for both girls and boys, the dummy variable in Row 4 still reports the mean Male-Female gap. For the same regression, one cannot reject the hypothesis that there are no differences in the ability coefficients across gender, though results are imprecise.

Figure 1: Raw Scores by Round and Gender



Graphs by Gender

Figure 2: Raw Scores by Round and Prior Ability



Graphs by Performance Terciles on State Assessment Test