Utrecht University
School of Economics

Do talented women shy away from competition?

Britta Hoyer
Thomas van Huizen
Linda Keijzer
Tahereh Rezai Khavas
Stephanie Rosenkranz
Bastian Westbrock

U.S.E. Discussion Paper Series nr: 16-06

Tjalling C. Koopmans Research Institute
Utrecht School of Economics
Utrecht University

Kriekenpitplein 21-22
3584 EC Utrecht
The Netherlands
telephone +31 30 253 9800
fax +31 30 253 7373
website www.koopmansinstitute.uu.nl

The Tjalling C. Koopmans Institute is the research institute and research school of Utrecht School of Economics.
It was founded in 2003, and named after Professor Tjalling C. Koopmans, Dutch-born Nobel Prize laureate in economics of 1975.

In the discussion papers series the Koopmans Institute publishes results of ongoing research for early dissemination of research results, and to enhance discussion with colleagues.

Please send any comments and suggestions on the Koopmans institute, or this series to J.M.vanDort@uu.nl

ontwerp voorblad: WRIK Utrecht

How to reach the authors

Please direct all correspondence to the second author.

Britta Hoyer
Warburger Str. 100
33098 Paderborn
Germany
Thomas van Huizen
Linda Keijzer
Tahereh Rezai Khavas
Stephanie Rosenkranz
Bastian Westbrock
Utrecht University
Utrecht School of Economics
Kriekenpitplein 21-22
3584 TC Utrecht
The Netherlands.
E-mail: t.m.vanhuizen@uu.nl

Utrecht School of Economics
Tjalling C. Koopmans Research Institute
Discussion Paper Series 16-06

Do talented women shy away from competition?

Britta Hoyer[a]
Thomas van Huizen[b]
Linda Keijzer[b]
Tahereh Rezai Khavas[b]
Stephanie Rosenkranz[b]
Bastian Westbrock[b]

[a]Center of International Economics
University of Paderborn

[b]Utrecht School of Economics
Utrecht University

September 2016

Abstract

We study the willingness to compete in a cognitive task among an entire cohort of fresh man business and economics students. Combining data from a lab-in-the-field experiment with university admissions data, we trace the gender gap in competitiveness at different levels of high school performance. Our results confirm that, on average, men choose to compete more often. The gender gap disappears, however, among students with above average high school performance. Female high school top performers are equally competitive as their male counterparts. In fact, the overall gender gap is entirely driven by the group of female high school under-performers who shied away from competition, even when they performed well in our task. Overall, our findings suggest that high school grades are more than just a signal of cognitive abilities, because they seem to influence the receivers self-perception of his or her performance in a competitive environment involved in later on in life.

Keywords: gender gap, competitiveness, performance feedback, high school grades, lab-in-the-field experiment

JEL classification: C93, J16, J24,I21

1 Introduction

It is now widely recognized that it takes more than just cognitive skills to succeed on the labor market (Heckman and Masterov, 2007; Lindqvist and Vestman, 2011). The general finding that non-cognitive traits matter suggests that gender differences in these traits may explain the persistent differences in labor market outcomes between men and women. More specifically, an explanation for the underrepresentation of women in top positions that has gained a prominent position in the literature is the gender difference in competitiveness: women tend to shy away from competitive payment schemes and are therefore less likely to climb all the way up the corporate ladder (Bertrand, 2011; Reuben et al., 2015). However, what explains these gender differences in competitiveness is not yet entirely understood.

The aim of this study is to investigate how the competitiveness of young men and women is influenced by their past high school performance. We do this in a lab-in-the-field experiment, where we administer an incentivized cognitive task based on Niederle and Vesterlund (2007) to an entire cohort of prospective students at a business economics faculty in the Netherlands. During the experiment, participants can earn lottery tickets by solving sets of math problems and by choosing between a piece-rate compensation scheme and a potentially more rewarding tournament compensation scheme. As this experiment is part of a mandatory full-day program, we are able to compare the choices within a representative population of freshman business and economics students, many of whom will fill vacancies in management positions after their study (Flynn and Quinn, 2010). By combining this experiment with university admissions data, we are able to examine the relationship between participants' high school grades and their tournament entry decisions. This comparison is facilitated by the fact that, in the Netherlands, the large majority of high schools is publicly financed and required to apply nationally standardized grading scales and examinations.

Our findings confirm Niederle and Vesterlund's results of a gender gap in the choice of the competitive scheme, even though the difference in our entire cohort is smaller than in the original U.S. sample.[1] Interestingly, we find that high school grades have no influence on the performance in the task, but a major impact on women's choices of the type of compensation scheme. In fact, the gender gap in willingness to compete disappears entirely within the subsample of prospective students in the upper part of the high school grade distribution.[2] It appears that the gender gap in the total sample is driven by only one subgroup: Dutch female participants with relatively low high school grades who choose the competitive scheme less often. In an attempt to explain this finding, we had a closer look at participants' task performances, before and after the choice of compensation scheme, only to notice no significant differences. What stood out is, however, that all subgroups but the one mentioned above respond to their prior task performance in the expected way: participants who solve more math questions correctly choose the competitive scheme more often. In contrast, female high school under-performers shy away from the competition regardless of their task performance.

The following simple theory helps to explain our findings: the choice of the competitive scheme depends on an individual's expectation to perform better than others. This expectation is formed based on available information on the task, such as an individual's prior performance in the task or the feedback provided during the experiment. However, as long as these signals are less than perfect, an individual will also draw on his or her experience within similar tasks. And even though both men and women may draw on their position in the high school grading curve, they may not do so equally.

In fact, there are several theoretical arguments why grades may matter more for girls. First, a

[1] A nonexclusive list of studies that succesfully replicated the original findings, yet sometimes with the same cultural attenuation as suggested by our findings, are Balafoutas and Sutter (2012); Healy and Pate (2011); Cason et al. (2010); Sutter and Rützler (2015); Niederle and Vesterlund (2011). Gneezy et al. (2003) and Gneezy and Rustichini (2004) extend the gender gap to performance under competition. Dreber et al. (2011, 2014) find that the gender gap is task dependent. Finally, Niederle et al. (2013) and Sutter et al. (2016) experimentally analyze the effect of policy interventions to close this gap.

[2] The same holds for our subsample of international participants, who can be expected to be among the better students in their home country. Pietro and Page (2008) indeed find that high school performance is positively correlated with studying abroad.

number of studies on gender-specific socialization at American high schools have noticed marked differences in the way boys and girls receive feedback on their classroom performance (see Roberts and Nolen-Hoeksema (1989) for a summary of this literature). Boys receive, on average, far more evaluative attention in the classroom than girls, mostly concerning classroom conduct or motivation. Girls, in contrast, receive more sparing feedback which is of a higher information content, as it is immediately related to their intellectual qualities. The implication for our study is that this learned way of dealing with feedback might have a long-lasting impact on how boys and girls assess their chances of succeeding in a competitive environment. Girls might become more responsive to external (past) performance evaluations, whereas boys are raised to ignore external feedback, in particular when it was negative.

Second, it is well documented that girls have higher grades in all or most subjects than boys (Alexander and Eckland, 1974; Halpern, 1997; Duckworth and Seligman, 2006). This may create higher grade expectations among teachers, parents, and peers. Girls who receive low grades therefore deviate from their social reference group, which may lower their confidence in their (relative) performance in future cognitive tests and tasks. Third, gender stereotypes may play a role. Because of the view that men surpass women in math tests, women experience typically higher levels of anxiety while making such tests, which lowers their performance (Steele, 1997; Spencer et al., 1999). Female students with above average (math) grades, however, may not suffer from anxiety, as they received frequent positive feedback on their skills.

Fourth, evidence indicates that boys exhibit lower levels of in-class citizenship, school motivation, and attachment (e.g. Downey and Vogt Yuan, 2005), suggesting that men simply care less about grades than women. They may therefore put more weight on their own evaluation of their performance in a task at hand. Hence, it is plausible that the weight individuals put on the information they garnered from their past high school performance is gender-specific. Whereas men's choices are relatively unaffected by their past grades, whether they performed poorly or not, women's choices are influenced by their past educational performance, in particular when they received a negative assessment.

Our findings contribute to a growing body of experimental studies with the aim to identify the reasons for women's overall lower inclination to compete (Eckel and Grossman, 2008; Eckel, 2008; Niederle and Yestrumskas, 2008; Kamas and Preston, 2012; Datta Gupta et al., 2013). They are most closely related to studies highlighting differences in self-confidence and feedback interpretation as the major cause. In particular, Mobius et al. (2014) investigate how participants in an online experiment process information on their relative position in an IQ test and find that women are significantly more conservative than men in updating their beliefs upon arrival of a noisy signal of their past test score. Similarly, Buser (2016) studies choices of a competitive compensation scheme in a series of cognitive tasks, but finds that women respond more than men to negative within-task feedback, whereas they respond less to positive feedback. Our findings on the average women are in line with both these studies, but we find that women who performed well on high school respond in no significant way different from their male counterparts.

Additionally, our study contributes to a body of literature with the same ambition as ours to attribute gender differences to different socialization processes (Gneezy et al., 2009; Booth and Nolen, 2012; Almås et al., 2015). For example, Booth and Nolen (2012) compare the competitive behavior of girls from single-sex and coeducational schools and find that girls from single-sex schools are significantly more competitive, so that their behavior actually resembles the one of boys from coeducational schools. The authors refer to another dimension of the high school socialization process as an explanation, namely the development of different gender identities in single- and mixed-sex groups. Our findings can be seen as complementary, because as part of the gender identity development at school, boys and girls may adopt different approaches of weighing their past performance evaluations (of others) in future decisions.

2 Experimental design and data

2.1 Experimental procedure

Our cognitive task data was collected in four sessions. Three of these sessions were organized during the so-called ”Matching Days” of the business economics department at a Dutch university in the spring and summer of 2015.[3] One session was organized in September 2015 as part of a required course. During the first sessions 97, 86 and 81 prospective students participated in our cognitive tasks, in the fourth session 109 registered students participated. The fourth session had mostly foreign participants, as the business economics department does not require them to participate in the Matching Days. All four sessions lasted one hour which was announced in advance, and were conducted as pen and paper experiments. The instructions can be found in the online appendix.

In each session, participants were first following a one and a half hour lecture with a break of 15 minutes before the actual experiment started. The gender composition of the participants within each session was not brought actively under the attention but the fact that the majority (approximately 66%) was male was salient. During the first three sessions, participants were divided into four smaller, more or less equally sized groups, which were seated in separate class rooms. During the fourth session, all participants were seated in a larger lecture hall. Our team of experimenters distributed and collected the hand-outs and answer sheets for each sub-task separately. Start and end times were publicly announced and behavior closely monitored. After completion of each sub-task, the answer sheets were immediately brought to another experimenter team who processed the data such that for each session the lottery could be executed and results could be communicated one and a half hour after the session ended.

[3]The university introduced the ”Matching Days” in 2012 as a compulsory requirement for all prospective students with a residence in the Netherlands to increase the fit between the students and their chosen study programs. This means that a prospective student has to visit the campus for an entire day and participate in a number of program related activities and tests. Based on his or her performance in these tests a non-binding advice is issued. Admission to the program is however declined, if a candidate did not participate in the Matching Day.

2.2 The cognitive tasks

Our experiments comprised a series of incentivized cognitive tasks. In each of the tasks, the participants could earn points. Each point represents a ticket for a lottery of an iPad2. The first part involved an individual task for which we used the same protocol as Niederle and Vesterlund (2007). In this task participants solved sets of math problems, first under a piece-rate compensation scheme (Task 1), then under a tournament compensation scheme (Task 2), and subsequently they were asked to choose one of the two compensation schemes (Task 3 choice) before solving another set of math problems (Task 3). At last, participants were asked to choose a compensation scheme for their first set of math problems solved under piece-rate compensation (Task 4 choice). While following the original Niederle and Vesterlund (2007) design closely we introduced three changes in the task:

(i) Participants had to add up only three two-digit numbers instead of five, and they had four minutes to do so instead of five. As in the original design, participants were allowed to use scrap paper but no calculators.

(ii) Each set of math problems consisted of a list of 30 sums handed out on a sheet of paper to the participants. Although we could not provide real-time feedback to the participants on their correct summations in the previous task, participants obtained indirect information about their relative performance. The maximal attainable score was salient to everyone so that participants could learn about their relative standing from the progress they made on their list, as well as from the observation that they finished the task later or earlier than their neighbors.

(iii) Finally, the tournament was performed among all participants in each of our four sessions of on average 90 participants, and not within groups of four as in the original design. Accordingly, the top 25% of the participants in each session earned 4 points for each correct answer in our set up, while all others earned nothing.

In the second part we measured participants' ability to reason about the mental states of others (an ability called "Theory of Mind"). Following Woolley et al. (2010), participants were shown

36 pairs of eyes, each for 10 seconds and participants were asked to indicate on a form sheet which of four possible mental states they read in the eyes. Each correct answer earned them one point. However, for this study we did not use the results of this test explicitly in our analysis.

The next task involved an assessment of risk attitudes and time preferences. We assessed risk attitudes in two ways. We used an incentivized method by administering the ”bomb risk elicitation task” as introduced by Crosetto and Filippin (2013). The points participants could earn corresponded to either 0 if the indicated number was larger than the randomly placed ”bomb”, or the points earned were equal to the indicated number. Additionally, we asked participants to self-assess in a non-incentivized way their risk attitudes on a 10 point Lickert scale (as in Dohmen et al., 2011).[4] We used the latter risk aversion measure in our main regression results (as this measure appears to be the stronger predictor).

Finally, we assessed participants willingness to pay for two kinds of information. For this we provided an endowment of 20 (two times 10) points and offered participants to use up to 10 points of this endowment to ”buy” general information on all tasks described above. The remaining 10 points could be used to ”buy” feedback information about how their own performance in the three tasks compared to that of other participants. At the end of the session the price for both kinds of information was randomly determined and when the participant specified an amount that was higher than this price, the respective information was sent via email on the next day. In our analyses we use the amount of tickets they were willing to give up for information about their (relative) performance to measure participants’ *preference for feedback*. The more points a participant was willing to give up to receive feedback, the higher his preference for feedback is.

Participants were informed about the sequence of tasks in the beginning of the experiment. Moreover, they were informed about the fact that instructions were identical for all participants, that the points they earned would be exchanged into lottery tickets, and that at the end of a session one lottery ticket would be randomly chosen to determine the winner. The winning prize was an

[4]We also include a non-incentivized 5 point scale item to measure time preferences.

iPad2 at the value of 475 Euros, which was shown around at the beginning of each session to make incentives salient.

2.3 Data on high school performance

To study the relationship between past high school performance and behavior in our cognitive tasks, we combined the experimental results with university admissions data. We used two main indicators to measure past high school performance: average final exam grades and whether the participant is an international student. The mandatory participation requirement of our Matching Days allows us to compare the choices within a large population of Dutch high school graduates. Moreover, as our program attracts a considerable number of foreign students, which are arguably a selection of the most successful high school graduates in their respective home countries, we additionally compare the choices of native and foreign students.

To capture high school performance of the Dutch students, we used data on the exam grades scored during the final year of high school. In the Dutch education system, the final exam grade for a subject is the unweighted average of a centralized written exam grade[5] and a school exam grade. Although the content and the level of the latter are up to the individual school, we still believe that the grades are comparable throughout the Netherlands. In the past there were indications that the level of the exams was not always equal among schools, but nowadays schools are monitored closely by the education inspectorate (see, for more detail, Inspectie van het Onderwijs, 2013). If for several years the school exam scores are substantially higher than the grades for the centralized exam, a school will be monitored more strictly (Inspectie van het Onderwijs, 2014).

A related issue is the comparability of average grades between students, as not all students take exams in the same subjects. Each student chooses one of four tracks, either nature and technology (NT), nature and health (NH), economics and society (ES) or culture and society

[5]This centralized written exam is the same for all students on all schools.

(CS). Some schools allow students to combine the two nature tracks (NT/NH) and society tracks (ES/CS). In addition to the specific track, students can in most cases choose one or several additional exam subjects. For instance, in our sample about half of the students with one of the nature tracks selected economics as an additional exam subject. The nature tracks are more science and math intensive and are therefore generally considered more challenging than the society tracks (see for example Buser et al., 2014). However, in our sample we do not find significant differences between the average exam grades of students with a nature track and students with a society track.[6] Nevertheless, to address this concern we performed robustness tests using the average grade of the three subjects that are mandatory in all tracks: Dutch, English and, Mathematics.[7]

Our study program attracts a considerable number of foreign students (68 out of 358 in 2015). For these international participants high school grades are not a reliable source of information, as there is a lot of variation between countries and even within countries. However, we believe that students who start their study abroad are highly likely to be among the best students in their home country. This is confirmed by Pietro and Page (2008). The authors find that high school performance is positively associated with the probability to study abroad. They argue that since the number of applicants tends to be higher than the number of places and scholarships available, home institutions often have to set up a selection process. The obvious criterion for this process is, of course, educational performance. The authors' results therefore confirm our intuition that excellent high school performance is positively associated with the probability to study abroad. Results concerning international students' performance abroad are mixed, but several studies have found that international students performed better than the home student population. Interestingly here, Morrison et al. (2005) and Dobson et al. (1998) find that female international students outperform their male counterparts, which we also observe in our data.

[6]The mean of the average exam grades is 6.72 and 6.64 for the society and the nature track students, respectively.
[7]The level of Mathematics differs between tracks though.

3 Results

3.1 The gender gap in tournament entry

We first present the results for our entire sample and show that overall, just as in Niederle and Vesterlund (2007), men and women differ in their preference for a piece rate versus a tournament compensation scheme. Table 1 provides the summary statistics for 358 Matching Day participants, for which we have complete information on all relevant variables.[8] The table highlights some clear gender gaps. Most notably, men and women differ in their compensation scheme choices for Tasks 3 and 4. Even though the magnitude of the gender gap is smaller than in the Niederle and Vesterlund (2007) sample of U.S. university students (73% versus 35%), 51% of our male participants chose the tournament scheme for Task 3, while only 33% of female participants did so. Similarly, 43% of men submitted their Task 1 solutions to a tournament in Task 4, while only 27% of women chose this option.[9] This gap implied an economic advantage for our male participants, as they earned on average more experimental tickets in both Tasks 3 and 4.

[Table 1 about here]

One of the obvious explanations is an accompanying gender difference in terms of how well our male and female participants performed in our math tests. Even though we find a gap of 1.71 in the average number of summations solved correctly for Task 1, this difference disappears in Task 2 and features only marginally in Task 3. As a consequence, there is also no significant gender gap in the opportunity to win the tournament of Task 3: 38% of our male and 33% of our

[8] While unusual in experiments to not have complete information for all participants, here this can be explained by our 'lab-in-the-field' setting. As we elicited the decisions and answers to a series of cognitive tasks with pen and paper and under time pressure, we could not check whether every participant performed all tasks.

[9] Other studies (e.g. Dreber et al., 2011; Booth and Nolen, 2012; Almås et al., 2015) find similar small gender gaps for participants from Sweden, the U.K., and Norway as we do. While this could be due to cultural differences between those countries and the U.S., two other potential factors in our case are that (i) the original Niederle & Vesterlund design might have triggered more competitive behavior in men, as the tournaments were taking place in small groups of four, and (ii) the fact that the male-female ratio was higher than one-to-one in our case, which might make our male participants shy away from competition, because the pool of competitors might have been perceived as stronger.

female participants would have won the tournament or actually did win it, in case they opted for it.[10] We therefore conclude that, based on performance alone, we would not suspect any gender difference in the compensation scheme choices for Task 3. On the other hand, the better male performance in Task 1 raised their Task 4 winning chances to a significant extent: 30% of the male participants scored among the top 25% in Task 1, while only 20% of the female participants did so.

3.2 The gender gap in general characteristics and information processing

If differences in math test performance are not the major cause for the gender gap in tournament entry choices, at least not for Task 3, an alternative explanation might be that our participants differed in their self-assessed, subjective winning chances. Indeed, several experimental studies show that under identical conditions women appear less self-confident and more risk averse (e.g. Niederle and Vesterlund, 2007; Eckel, 2008). Moreover, Mobius et al. (2014) find that women respond less to within-task feedback provided to them, making them appear more conservative information processors.

To check whether differences in general characteristics or information processing can explain the gender gap, we draw on the information available from our other Matching Days tasks. Our participants experienced the math test twice before they got to choose their preferred compensation scheme for Task 3. Hence, we measure gender differences in information processing by how well our male and female participants translate their perception on whether their Task 2 math score was (not) among the best in this task into a subsequent choice regarding the compensation scheme. Furthermore, to capture our participants' general characteristics, we draw on the elicited measures of (self-reported) risk and feedback preferences and, controlling for these

[10] Since in our set up a participant wins the Task 3 tournament with certainty if his or her performance is among the top 25% of all Task 2 performances, we know exactly how many problems are needed to be solved in Task 3 to be amongst the winners: 24 on the session of day one, 24 on the two sessions of day two, and 25 on the session on day three. To be among the winners for Task 4, a participant needed 22 correct summations in Task 1 in either of the four sessions.

preferences as well as a participant's Task 2 score, use their compensation scheme choices for Task 4 to measure self-confidence. As can be seen from Table 1, men and women indeed differ markedly and in the expected direction in terms of all these measures.

[Table 2 about here]

Table 2 presents the results of five OLS regressions with Task 3 choices as the dependent variable. Column (1) reports again the overall gender gap of approximately 18%. In column (2), we introduce a high Task 2 score as an explanatory factor. Because participants did not know exactly their own scores nor the scores of their competitors, we consider everything above the median as a high score in this column.[11] These results suggest that, unlike in the original study by Niederle and Vesterlund (2007), the average participant is more likely to enter the tournament when he or she scored high in Task 2. In column (3), we introduce our measures of feedback preference and risk aversion. The coefficient of our measure for risk aversion is significant and in the expected direction. Moreover, the gender gap in tournament entry drops but remains significant when we introduce our general characteristics measures. Interestingly, in contrast to the findings of Niederle and Vesterlund (2007), the gender gap becomes insignificant when we include a dummy indicating whether the participant opted for the tournament in Task 4 (column (4)). The final column of Table 2 introduces an interaction variable between *Female* and *Task 2 (Top 50%)* that captures gender-specific differences in dealing with the imperfect signal of a high performance in the previous task. Even though our female participants responded slightly less to an above median score than the male participants, the difference is not statistically significant, suggesting that there are no overall gender differences in information processing in our sample.

[11] Similar results as the ones shown in columns (2)-(5) are obtained when we consider a Task 2 score in the top quartile as our explanatory variable or use the raw Task 2 score as our measure of task-related performance. Also, we can reproduce the findings reported in Table 2, when we regress Task 4 tournament submissions on the listed explanatory variables except for, of course, *Tournament (Task 4)*.

3.3 The gender gap at different levels of high school performance

The results presented so far confirm that there is a gender gap in the choice of a competitive compensation scheme and that the tendency of women to be generally more risk averse and less confident are potential explanations for this gender gap. In the light of the literature on gender-specific high school socialization (e.g., Roberts and Nolen-Hoeksema, 1989; Downey and Vogt Yuan, 2005), the question arises whether this gap is the same for our female participants who received low grades during their high school career and those with outstanding marks.

Our findings for the different subgroups can be found in Table 3. A first observation is that, also within and in-between the different subgroups, there is no systematic gender gap in terms of math test performances. The picture from the entire sample is even turned around for some subgroups, as our female participants among the top 25% of Dutch high school students were actually doing slightly better in all three tasks than their male counterparts with the same high school performance.

[Table 3 about here]

Against this background, it is interesting to note the striking subgroup differences in terms of compensation scheme choices. In fact, the only group that confirms that there is a gender gap in willingness to compete is the group of Dutch high school graduates *below* the top 25% bracket, in particular the bottom 50% of students, where 59% of the male versus 25% of the female participants chose the tournament for Task 3. In contrast, the gender gap disappears entirely in the subgroup of the top 25% Dutch high school graduates and among our foreign participants. Hence, the results presented in Table 3 suggests that the overall gender gap in competitiveness, which we found in our total sample, is driven by women with relatively low grades.[12]

[12]The same picture emerges even after controlling for participants' Task 2 test scores or general characteristics, as done in the regression analysis presented in a separate appendix. There, we reproduce the analysis of Table 2 by subgroup and find that female high school under-performers are significantly less competitive than their male counterparts, even after controlling for test scores, risk aversion, preference for feedback, and the decision to submit to the tournament in Task 4. In contrast, no gender gap in the willingness to compete is found within the subgroup of participants with above median high school grades.

Because the total earnings (in terms of lottery tickets) is determined by both the performance in the tasks and the compensation scheme choices, it may not be surprising that there are important differences in earned lottery tickets. Table 3 shows that, while Dutch men on average earned significantly more than Dutch women in Task 3, the most successful female high school graduates earned more than their male counterparts. Within the group of international students, women also earned significantly more than men. In contrast, the female participants with below median high school performance earned fewer tickets in Task 3 than all other subgroups.

3.4 The effect of high school performance on tournament entry

Another way of looking at the results presented in Table 3 is to compare the tournament entry choices by gender at different levels of high school performance. In that way, the table sheds light on the impact of high school grades separately for men and women. The table points out a striking difference in payment scheme choices (Task 3) between female participants who performed well in high school and their under-performing counterparts: while 48% of both the group of female Dutch students with grades in the top quartile and the group of foreign students chose the tournament scheme, only 25% of the Dutch participants in the bottom 50% of high school graduates selected this payment scheme. The choices of the male participants, in contrast, appear to be unrelated to their high school achievement level.

Table 4 provides the regression results on the relation between high school grades and tournament entry. Columns (1) and (2) reproduce the findings of Table 2 for the subgroups of all Dutch participants (column (1)) and the participants for which the official high school grades are available (column (2)). Columns (3) and (5) introduce past high school performance as an additional explanatory factor and show that, for the average participant, belonging to the top 50%, respectively top 25%, at high school plays no role in the decision to enter the tournament. We interact past high school performance with the participant's gender in columns (4) and (6).

[Table 4 about here]

Consistent with Table 3, the effect of high school grades on tournament entry seems heterogeneous. The two columns provide evidence for the hypothesis that women's willingness to compete is higher, if their high school grades are higher. The combined effect of *High grade* and *Female x High grade* is insignificant in a Wald test when considering an above median grade as high, but significant (p=0.027) when we focus on the 25% highest grade bracket. Interestingly, columns (4) and (6) suggest that the relation is quite the other way around for men: the negative coefficients of *High grade* (significant at the $10\% level$) provide some weak evidence that our Dutch male participants are less competitive at higher levels of grade performance. Nevertheless, unlike the findings reported above on our female participants, the effect of past high school performance on the decisions taken by our male participants vanishes if we proxy past educational performance by the fact that a participant came from abroad, as done in some auxiliary regression analyses reported in the separate appendix.

3.5 Information processing at different levels of high school performance

Up until this point, the data suggests that our female participants draw on their past high school performance to judge about their winning chances in our math tests, while the choices of men seem to be rather unaffected by this task-independent piece of information. But does this mean that a position at the top end of the high school grading curve renders women equally focused on the task as men? And are female high school under-performers generally less confident, independent of their task performance?

To take a closer look at this, the two panels of Figure 1 plot the Task 3 tournament-entry choices of our participants against their Task 2 performances, conditional upon gender and high school grades. The most striking pattern is the difference between the female participants in the two subgroups: unlike the female high school graduates in the top tier, the under-performers do not seem to pick up to the positive signal of a higher Task 2 performance. As a consequence, less than 30% of them who answered (almost) all summations correctly actually chose the tournament option. The females in the top quartile, in contrast, responded to their Task 2 performance similar to our male participants.

[Figure 1 about here]

Likewise, Figure 2 plots the tournament-entry choices against Task 2 performance separately for the subgroups of Dutch and foreign participants. Most interestingly, the left panel suggests that our foreign female participants were even more selective in their tournament-entry choices than their male counterparts. That is, they chose the tournament more often in the upper quartiles 3 and 4 and less often in the losing quartiles 1 and 2.

[Figure2 about here]

Regression results for the different subgroups (reported in the separate appendix) confirm these patterns. In particular, the interaction effect between *Female* and *Task 2 top 50%* is insignificant for all subgroups but for the female high school under-performers, for which it is negative, significant and sizeable.[13] Thus, our findings point out an intriguing interaction between educational achievement and gender. While five of the six investigated subgroups respond appropriately to their prior task performance, it is the group of female high school under-performers that stands out in this regard and that drives the overall gender gap in our complete sample.

[13] We performed various other robustness checks on our results (available upon request), where we used other measures of task performance (e.g., top 25% in Task 2 performance) and educational attainment (distinction between students above and below the median high school grade, or tournament entry conditional on participants' average high school grades in Dutch, English and Mathematics). The results of all these checks are in line with Figures 1 and 2 and the corresponding table in the appendix can be reproduced.

4 Conclusion

A growing body of research shows that men and women differ in their choices to select into a competitive environment. This study provides evidence from a lab-in-the-field experiment on why they may do so. Our findings are inconsistent with a universal gender gap in willingness to compete. Instead, we find that the willingness to compete of our female participants is moderated by their past educational performance, such that the most successful among them appear to be as competitive as the male participants in our experiment.

Even though we are not in the position to draw definite conclusions, a story consistent with our findings is that the negative feedback, i.e., a low grade, received during their high school career spilled over to women's decisions in our experiment, whereas this is not the case for men. Overall, the latter quite correctly interpreted an imperfect signal about their own task performance during the experiment and adjusted their choices accordingly: those who performed better in our math tests prior to the tournament entry decisions are more likely to select the competitive scheme. We also find the same responsiveness in women with an above average high school performance and those who were selected to study abroad. Yet, female high school under-performers did not respond to their own prior math test performance. They appear to use their low position on the grading curve as a reference in our experimental contest against a similar pool of contestants and, accordingly, underestimate their winning chances, even if they would have won the tournament by a substantive margin.

Our findings add a nuanced view to established theories in social psychology (Garcia et al., 2013) that women, in general, respond more conservatively to positive feedback and are more sensitive to negative feedback. They contradict Niederle and Vesterlund's conjecture that "much may be gained if we can create environments in which high-ability women are willing to compete" (Niederle and Vesterlund, 2007, p. 1100). The talented female participants in our experiment were by no means less competitive than their male counterparts. Attention should rather go to those females who underperformed in terms of their educational achievements earlier in life.

References

Alexander, K. L. and Eckland, B. K. (1974). Sex differences in the educational attainment process. *American Sociological Review*, pages 668–682.

Almås, I., Cappelen, A. W., Salvanes, K. G., Sørensen, E. Ø., and Tungodden, B. (2015). Willingness to compete: Family matters. *Management Science*.

Balafoutas, L. and Sutter, M. (2012). Affirmative action policies promote women and do not harm efficiency in the laboratory. *Science*, 335(6068):579–582.

Bertrand, M. (2011). New perspectives on gender. *Handbook of labor economics*, 4:1543–1590.

Booth, A. and Nolen, P. (2012). Choosing to compete: How different are girls and boys? *Journal of Economic Behavior & Organization*, 81(2):542–555.

Buser, T. (2016). How does the Gender Difference in Willingness to Compete evolve with Experience? *Mimeo*.

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

Cason, T. N., Masters, W. A., and Sheremeta, R. M. (2010). Entry into winner-take-all and proportional-prize contests: An experimental study. *Journal of Public Economics*, 94(9):604–611.

Crosetto, P. and Filippin, A. (2013). The ”bomb” risk elicitation task. *Journal of Risk and Uncertainty*, 47(1):31–65.

Datta Gupta, N., Poulsen, A., and Villeval, M. C. (2013). Gender matching and competitiveness: Experimental evidence. *Economic Inquiry*, 51(1):816–835.

Dobson, I., Sharma, R., and Calderon, A. (1998). The comparative performance of overseas and australian undergraduates. In *12th Australian International Education Conference, Canberra*, pages 3–18.

Dohmen, T., Falk, A., Huffman, D., Sunde, U., Schupp, J., and Wagner, G. G. (2011). Individual risk attitudes: Measurement, determinants, and behavioral consequences. *Journal of the European Economic Association*, 9(3):522–550.

Downey, D. B. and Vogt Yuan, A. S. (2005). Sex differences in school performance during high school: Puzzling patterns and possible explanations. *The Sociological Quarterly*, 46(2):299–321.

Dreber, A., von Essen, E., and Ranehill, E. (2011). Outrunning the gender gap - boys and girls compete equally. *Experimental Economics*, 14(4):567–582.

Dreber, A., von Essen, E., and Ranehill, E. (2014). Gender and competition in adolescence: task matters. *Experimental Economics*, 17(1):154–172.

Duckworth, A. L. and Seligman, M. E. (2006). Self-discipline gives girls the edge: Gender in self-discipline, grades, and achievement test scores. *Journal of educational psychology*, 98(1):198.

Eckel, C. (2008). Experiments on gender differences. In *The New Palgrave Dictionary of Economics, Second Edition, Edited by Steven N. Durlauf and Lawrence E. Blum*. Palgrave Macmillan (Basingstoke and New York).

Eckel, C. C. and Grossman, P. J. (2008). Forecasting risk attitudes: An experimental study using actual and forecast gamble choices. *Journal of Economic Behavior & Organization*, 68(1):1–17.

Flynn, P. M. and Quinn, M. A. (2010). Economics: Good choice of major for future ceos. *The American Economist*, 55(1):58–72.

Garcia, S. M., Tor, A., and Schiff, T. M. (2013). The psychology of competition a social comparison perspective. *Perspectives on Psychological Science*, 8(6):634–650.

Gneezy, U., Leonard, K. L., and List, J. A. (2009). Gender differences in competition: Evidence from a matrilineal and a patriarchal society. *Econometrica*, 77(5):1637–1664.

Gneezy, U., Niederle, M., Rustichini, A., and others (2003). Performance in competitive environments: Gender differences. *QUARTERLY JOURNAL OF ECONOMICS-CAMBRIDGE MASSACHUSETTS-*, 118(3):1049–1074.

Gneezy, U. and Rustichini, A. (2004). Gender and competition at a young age. *The American Economic Review*, 94(2):377–381.

Halpern, D. F. (1997). Sex differences in intelligence: Implications for education. *American Psychologist*, 52(10):1091.

Healy, A. and Pate, J. (2011). Can Teams Help to Close the Gender Competition Gap? *The Economic Journal*, 121(555):1192–1204.

Heckman, J. J. and Masterov, D. V. (2007). The productivity argument for investing in young children. *Applied Economic Perspectives and Policy*, 29(3):446–493.

Inspectie van het Onderwijs (2013). Onderwijsverslag 2012/2013 - centraal examen en schoolexamen.

Inspectie van het Onderwijs (2014). Beoordeling indicator verschil cijfer schoolexamen en centraal examen.

Kamas, L. and Preston, A. (2012). The importance of being confident; gender, career choice, and willingness to compete. *Journal of Economic Behavior & Organization*, 83(1):82–97.

Lindqvist, E. and Vestman, R. (2011). The labor market returns to cognitive and noncognitive ability: Evidence from the swedish enlistment. *American Economic Journal: Applied Economics*, 3(1):101–128.

Mobius, M. M., Niederle, M., Niehaus, P., and Rosenblat, T. S. (2014). Managing self-confidence: Theory and experimental evidence. Technical report, National Bureau of Economic Research.

Morrison, J., Merrick, B., Higgs, S., and Le Métais, J. (2005). Researching the performance of international students in the uk. *Studies in Higher Education*, 30(3):327–337.

Niederle, M., Segal, C., and Vesterlund, L. (2013). How costly is diversity? Affirmative action in light of gender differences in competitiveness. *Management Science*, 59(1):1–16.

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

Niederle, M. and Vesterlund, L. (2011). Gender and competition. *Annu. Rev. Econ.*, 3(1):601–630.

Niederle, M. and Yestrumskas, A. H. (2008). Gender differences in seeking challenges: The role of institutions. Technical report, National Bureau of Economic Research.

Pietro, G. D. and Page, L. (2008). Who studies abroad? evidence from france and italy. *European Journal of Education*, 43(3):389–398.

Reuben, E., Sapienza, P., and Zingales, L. (2015). Taste for competition and the gender gap among young business professionals. Technical report, National Bureau of Economic Research.

Roberts, T.-A. and Nolen-Hoeksema, S. (1989). Sex differences in reactions to evaluative feedback. *Sex Roles*, 21(11-12):725–747.

Spencer, S. J., Steele, C. M., and Quinn, D. M. (1999). Stereotype threat and women's math performance. *Journal of experimental social psychology*, 35(1):4–28.

Steele, C. M. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. *American psychologist*, 52(6):613.

Sutter, M., Glätzle-Rützler, D., Balafoutas, L., and Czermak, S. (2016). Cancelling out early age gender differences in competition: an analysis of policy interventions. *Experimental Economics*, 19(2):412–432.

Sutter, M. and Rützler, D. (2015). Gender differences in competition emerge early in life and persist. *Management Science*, 61(10):2339 – 2354.

Woolley, A. W., Chabris, C. F., Pentland, A., Hashmi, N., and Malone, T. W. (2010). Evidence for a collective intelligence factor in the performance of human groups. *Science*, 330(6004):686–688.

Table 1: Results for entire sample

	Overall means	Male means	Female means
Dutch nationality	80.52%	81.86%	77.97%
Avg. High school Math Grade (only Dutch)	6.82	6.71	7.11***
Avg. High school Grade (only Dutch)	6.67	6.59	6.86***
Task 1 Performance	17.80	18.38	16.69***
Task 2 Performance	19.06	19.26	18.69
Task 3 Tournament choice	44.77%	50.88%	33.05%***
Task 3 Performance	21.34	21.73	20.59*
Task 3 Tournament winners	36.59%	38.30%	33.33%
Task 3 Tickets earned	30.87	32.09	28.52
Task 4 Tournament submission	37.50%	42.92%	27.12%***
Task 4 Tournament winners	25.00%	30.21%	20.33%**
Task 4 Tickets earned	27.54	30.30	22.22**
Bomb task risk aversion	59.01	57.61	61.69*
Self-reported risk aversion	3.95	3.69	4.44***
Preference for feedback	3.35	3.59	2.88*
Observations	358	235	123

NOTES: An asterisk indicates a significant difference between the male and female mean in a two-sided t-test: *** p<0.01, ** p<0.05, * p<0.1.

Table 2: OLS results for Tournament-Entry Decision (Task 3): Full sample

	(1)	(2)	(3)	(4)	(5)
Female	-0.177***	-0.148***	-0.117**	-0.0806	-0.0791
	(0.0547)	(0.0528)	(0.0539)	(0.0529)	(0.0755)
Task 2 (Top 50%)		0.281***	0.253***	0.233***	0.234***
		(0.0503)	(0.0505)	(0.0495)	(0.0611)
Preference for feedback			0.0157**	0.0111	0.0111
			(0.00725)	(0.00711)	(0.00713)
Risk aversion			-0.0391***	-0.0336**	-0.0336**
			(0.0141)	(0.0137)	(0.0139)
Tournament (Task 4)				0.253***	0.253***
				(0.0518)	(0.0519)
Female x Task 2 (Top 50%)					-0.00288
					(0.104)
Constant	0.511***	0.349***	0.453***	0.351***	0.351***
	(0.0321)	(0.0422)	(0.0722)	(0.0734)	(0.0749)
Observations	358	358	348	344	344
R-squared	0.029	0.107	0.138	0.195	0.195

NOTES: Standard errors are reported in parentheses. Sample sizes are smaller in columns (3)-(5) due to missing values for elicited preference measures.*** p<0.01, ** p<0.05, * p<0.1

Table 3: Results for different subsamples

	Dutch prosp. students			Dutch students									Foreign students		
				Bottom 50% grades			50-25% grades			Top 25% grades					
	Avg.	Male	Female	Avg.	Male	Female	Avg.	Male	Female	Avg.	Male	Female	Avg.	Male	Female
Task 1 Performance	18.18	18.81	16.9***	19.25	19.62	17.85	17.51	17.80	16.75	17.44	17.13	18.23	16.19	16.41	15.85
Task 2 Performance	19.20	19.6	18.42	20.12	20.42	18.95	18.51	18.96	17.33	19.43	18.40	20.9*	18.45	17.71	19.44
Task 3 Tourn. choice	43%	50%	28%***	52%	59%	25%***	35%	42%	17%	43%	40%	48%	52%	54%	48%
Task 3 Performance	21.36	21.92	20.23**	22.35	22.54	21.60	20.42	21.13	18.58	21.16	20.90	21.85	21.23	20.85	21.85
Task 3 Tickets earned	31.41	34.30	25.55*	33.89	37.25	21.30	31.30	34.65	22.67	27.84	26.70	29.48	28.56	21.63	39.07*
Task 4 Tourn. choice	35%	40%	26%***	37%	37%	35%	33%	39%	17%	43%	48%	35%	49%	56%	37%
Task 4 Tickets earned	29.06	31.24	24.60	34.20	34.04	34.75	27.26	29.48	21.50	23.73	21.76	26.60	21.13	25.88	13.93
Bomb task risk av.	59.10	57.91	61.49	59.35	58.13	63.90	60.12	58.55	63.42	61.55	59.4	60.44	58.63	56.22	62.42
Self-reported risk av.	3.98	3.71	4.52***	3.60	3.58	3.70	4.38	3.90	5.6***	4.29	3.68	5.11***	3.82	3.59	4.19
Feedback preference	3.05	3.29	2.59	3.32	3.47	2.75	2.52	2.39	2.84	2.85	3.07	2.56	4.57	4.98	3.92
Observations	290	194	96	95	75	20	43	31	12	51	30	21	68	41	27

NOTES: Asterisks indicate significant difference from male mean in the same subsample in a two-sided t-test: *** p<0.01, ** p<0.05, * p<0.1.

Table 4: OLS results for Tournament-Entry Decision (Task 3): Native students

	(1)	(2)	(3)	(4)	(5)	(6)
Female	-0.0891	-0.0256	-0.120	-0.177	-0.125	-0.137
	(0.0862)	(0.116)	(0.0768)	(0.137)	(0.0773)	(0.121)
Task 2 (Top 50%)	0.236***	0.262***	0.210***	0.242***	0.212***	0.247***
	(0.0680)	(0.0796)	(0.0683)	(0.0798)	(0.0685)	(0.0790)
Preference for feedback	0.0183**	0.0213**	0.0196**	0.0207**	0.0199**	0.0219**
	(0.00811)	(0.00968)	(0.00963)	(0.00962)	(0.00965)	(0.00953)
Risk aversion	-0.0245	-0.0208	-0.0139	-0.0256	-0.0168	-0.0257
	(0.0161)	(0.0198)	(0.0198)	(0.0204)	(0.0197)	(0.0197)
Tournament (Task 4)	0.242***	0.303***	0.311***	0.318***	0.309***	0.302***
	(0.0587)	(0.0699)	(0.0699)	(0.0697)	(0.0702)	(0.0690)
Female x Task 2 (Top 50%)	-0.0673	-0.180		-0.191		-0.241
	(0.117)	(0.151)		(0.153)		(0.153)
High grade (Top 50%)			-0.0729	-0.140*		
			(0.0680)	(0.0790)		
Female x High grade (Top 50%)				0.306**		
				(0.154)		
High grade (Top 25%)					-0.0259	-0.154*
					(0.0758)	(0.0921)
Female x High grade (Top 25%)						0.444***
						(0.160)
Constant	0.297***	0.244**	0.283***	0.330***	0.265***	0.304***
	(0.0838)	(0.100)	(0.101)	(0.108)	(0.100)	(0.102)
Observations	277	182	182	182	182	182
R-squared	0.202	0.228	0.227	0.248	0.222	0.261

NOTES: Standard errors are reported in parentheses. Sample sizes are smaller than in Table 3 due to some missing values for preference measures.*** p<0.01, ** p<0.05, * p<0.1

Figure 1: Proportion of high and low performing high school graduates choosing tournament for Task 3 conditional on Task 2 performance quartiles

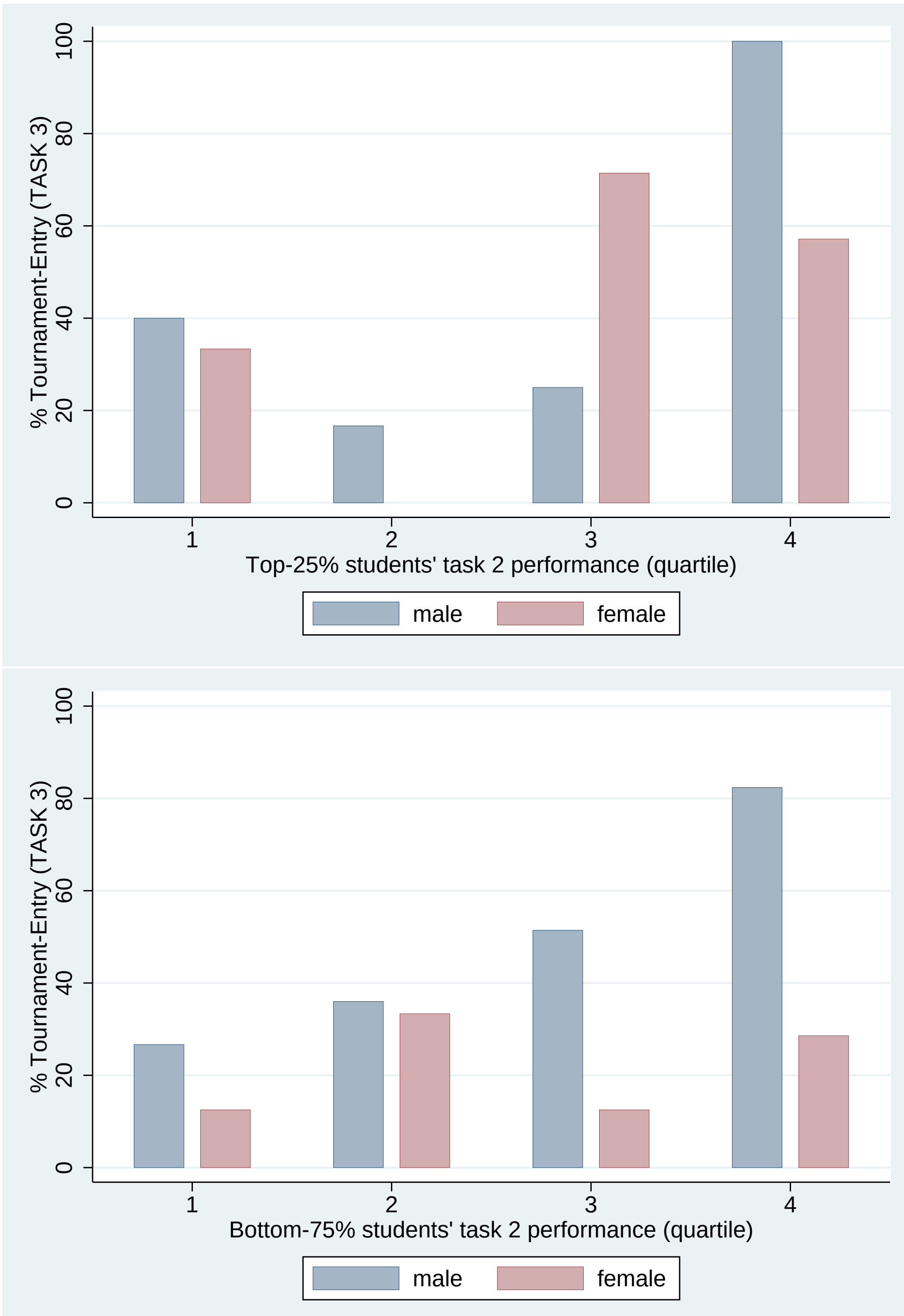

Figure 2: Proportion of native and foreign students choosing tournament for Task 3 conditional on performance quartiles