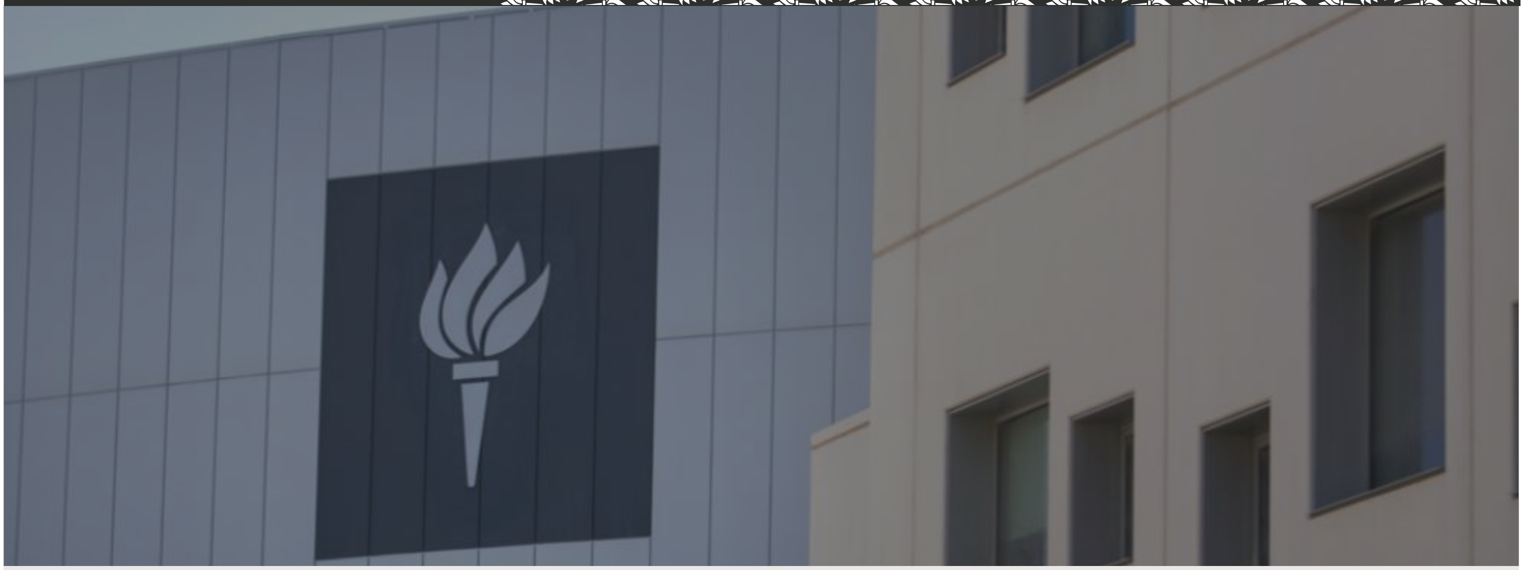
جامعة نيويورك أبوظبي
NYU | ABU DHABI



Competitiveness, selection bias and gender differences among economics majors

Aurelie Dariel, Nikos Nikiforakis, Jan Stoop

Working Paper # 0074
January 2022

Division of Social Science Working Paper Series
New York University Abu Dhabi, Saadiyat Island P.O Box 129188, Abu Dhabi, UAE
http://nyuad.nyu.edu/en/academics/academic-divisions/social-science.html

Competitiveness, selection bias and gender differences among economics majors[*]

Aurélie Dariel[†] Nikos Nikiforakis[‡] Jan Stoop[⁑]

Abstract

Evidence from behavioral experiments with volunteer samples suggests that there exists a substantial gap in the willingness of men and women to compete. We ask whether a similar gap can be found in a population of economics majors – a population of interest as questions loom regarding the reasons for the underrepresentation of women in economics. We find a substantial gender gap in competitiveness – as well as in risk attitudes – among economics majors. We also find that self-selection into the lab causes us to overestimate this gap among volunteers by a factor of 2 to 3 depending on the econometric model.

Key words: gender gap, diversity in economics, selection bias, lab experiment, competitiveness.

JEL Classification: C90, D03

[*] We would like to thank Moritz Janas, Ernesto Reuben and Lise Vesterlund for comments, as well as participants at the 2019 Asia Pacific Meeting of the Economic Science Association, and the 2019 International Symposium for Experimental Economics at NYU Abu Dhabi. Special thanks go out to Bas Karremans, Omar Ricket and Tim Benning for helping conduct the experiment. For financial support, we are thankful to NWO, the Netherlands (Veni grant 016.155.026 and Vidi grant VI.Vidi.195.061), and Tamkeen under the NYUAD Research Institute Award CG005. Experimental instructions and all communications with students can be found in the Online Supplementary Material. The protocol for the study was approved by the IRB at NYU Abu Dhabi (006-2018) and at Erasmus University Rotterdam (17.097).

[†] Division of Social Science, New York University Abu Dhabi, PO Box 129188, United Arab Emirates. E-mail: apd5@nyu.edu. Phone: +971 26285436.

[‡] Division of Social Science, and Center for Behavioral Institutional Design, New York University Abu Dhabi, PO Box 129188, United Arab Emirates. E-mail: nikos.nikiforakis@nyu.edu. Phone: +971 26285436.

[⁑] Erasmus School of Economics, Erasmus University Rotterdam, PO Box 1738, 3000 DR Rotterdam, the Netherlands. E-mail: stoop@ese.eur.nl. Phone: +31 104082708.

1. Introduction

The persistent differences in labor market outcomes between men and women have been a topic of continuous interest for researchers, policy makers, and managers (Bertrand 2020; Blau and Khan 2017; Goldin, 2014). Traditional explanations for the gap include the gender disparities in human capital (Becker 1991), the division of labor in couples (Becker 1985), and discrimination by employers (Neumark, Bank and Van Nort 1996). More recently, advances in experimental research allowed researchers to identify other discrepancies between men and women that are difficult to observe in natural settings. A prime example is the evidence from behavioral experiments showing a substantial gap in the willingness of men and women to compete (Niederle and Vesterlund 2007; henceforth NV2007).[1] Given the importance of competitive incentives in modern organizations, the relative reluctance of women to engage in competition has been proposed as a plausible explanation for women's lower wages and lower representation in high-level jobs (Blau and Khan 2017; Goldin 2014). Despite the publication of several dozens of experimental studies on the topic of competitiveness, however, some important questions linger.

In their authoritative review of the literature on the gender wage gap, Blau and Kahn (2017, p. 837) write: "[T]here are reasonable concerns about generalizing the results of such experiments outside the lab". One of the main concerns when generalizing findings from lab experiments such as those on gender arises from the use of volunteer subjects, i.e., individuals that self-select into an experiment (Falk and Heckman 2009; Levitt and List 2007). If people's attitudes toward competition affect selection into experiments, inferences about the gender gap in competitiveness drawn from volunteer samples could be biased, i.e., not reflect the gender gap in the population from which the volunteers were recruited. Specifically, inferences regarding the gender gap *will be* biased if attitudes toward competition affect differently the likelihood of men and women to participate in lab experiments. To the extent that experimental measures of competitiveness are used to champion specific policies for reducing the gender gap such as gender quotas or mentoring (Balafoutas and Sutter 2012; Brandts, Groenert and Rott 2015; Ip, Leibbrandt and Vecci 2019; Leibbrandt, Wang and Foo 2018; Niederle, Segal and Vesterlund 2013), the issue of selection bias is not merely of academic interest, but of practical relevance.

[1] Following the pioneering study by NV2007, several dozen experiments have documented the tendency of women to shy away from competition, and of men to compete "too much." For extensive reviews of the literature, see Dariel et al. (2017) and Niederle (2016).

The present study has two goals. The first one is to evaluate whether there exists a gender gap in competitiveness in a population that we believe to be of interest: economics majors. Economics as a field has recently come under scrutiny for the low representation of women among both students and faculty members. Specifically, whereas women make up about half of the entire undergraduate student population, they account for approximately 30% of economics majors the US (Bayer and Rouse 2016; Buckles 2019; Lundberg and Stearns 2019). Bayer and Rouse (2016) show that the problem is substantially worse in economics than in STEM, where approximately 60% of majors are female. With regards to faculty members, the fraction of women in economics appears to steadily decline as one moves from graduate school, to the ranks of assistant professor, associate professor and, finally, full professor – a trend referred to as "the leaky pipeline" – both in US and European departments (Auriol, Friebel and Wilhelm 2019; Buckles 2019; Lundberg and Stearns 2019). A recent report by the *Royal Economic Society* documented a very similar situation in the UK (Bateman et al. 2021).[2] Given that a considerable fraction of the observed variance in gender representation remains unexplained (Bayer and Rouse 2016), it seems interesting to explore whether there exists a gender gap in competitiveness among economics majors. To our knowledge, our study is the first to address this question.

It is not obvious that a gender gap in competitiveness should exist among economics majors. If the most competitive individuals select into math intensive fields such as economics as evidence suggests (Buser, Niederle and Oosterbeek 2014; Buser, Peter and Wolter 2017), then we would expect that the gap in competitiveness between men and women majoring in economics should be small, if any. However, if a gender gap in competitiveness does exist, it could help explain some of the "leak" in the pipeline, and suggest possible "fixes." To evaluate whether a gap exists in this population, we take advantage of a compulsory class with mandatory attendance for all students majoring in economics at one of the largest universities in Europe – Erasmus University Rotterdam. As we show in the next section, the patterns of female (under)representation in this university appear to be very similar to those in the UK and the US.

The second goal of our study is to explore whether self-selection into the lab biases the estimated gender differences in competitiveness. If a bias exists, we would like to know whether self-selection causes us to overestimate the gap in competitiveness or to *under*estimate it. The latter

[2] Megalokonomou, Vidal-Fernandez and Yengin (2021) provide evidence that, in Europe, there is a similar conversion rate of graduates to faculty in economics and STEM.

could happen if women in the population are less willing to compete than men, but only men and women with a certain tolerance for competition volunteer for experiments.[3] To test for the existence of selection bias, we use a method introduced by Cleave, Nikiforakis and Slonim (2013). At time t_0, we elicit individuals' willingness to compete during normal class hours using a design similar to that of NV2007. As the class we use for our study is compulsory with mandatory attendance for all first-year economics majors, selection to our study should be minimal. Even if some students fail to show up in class on the day of our study, self-selection should not bias our estimates of the gender gap in competitiveness as the experiment was not pre-announced. For brevity, therefore, we will refer to the 1,145 students in the classroom experiment at time t_0 as the "population." At time t_1, we identify the subset of "lab volunteers" in the population by observing who accepts an invitation to participate in lab experiments. To evaluate the effect of selection bias, we compare the gender gap among lab volunteers measured at time t_0 and the gender gap in the population.[4]

The data reveal that there is a substantial gender gap in competitiveness among economics majors. Specifically, men majoring in economics are 15 percentage points more likely to compete than women. We also find that men majoring in economics are more risk tolerant than women, and more likely to select into competition when their performance is poor. In other words, we obtain qualitatively similar results in this population as those found in previous behavioral experiments. We also find evidence of substantial selection bias in our lab estimates. To be precise, self-selection causes us to *over*estimate the gap in competitiveness by a factor between 2 and 3,

[3] In Appendix A, we present evidence from past studies that is in line with the hypothesis that self-selection causes us to *over*estimate the gender gap in competitiveness. Specifically, we use on the survey of Dariel et al. (2017) to classify all studies that were either published or accepted for publication by December 2017 into those that rely on self-selected samples and those that do not. While we find a substantially larger gender gap in studies with self-selected samples, there is a confound that prevents us from safely attributing the difference to self-selection.

[4] As we clarify in the next section, in actuality, there is a sequence of events between time t_0 and time t_1 which relate to the usual steps of recruiting participants to experiment. Importantly, note that, we compare the gap among volunteers at time t_0 and not at time t_1 as learning, time and environmental effects would confound our estimates in that case. Specifically, let $\Delta_P^{t_0}$ denote the gender gap in competitiveness in the population measured at t_0, i.e., in the classroom experiment. Let $\Delta_V^{t_0}$ denote the gender gap in competiveness among lab volunteers also measured at t_0. We will say self-selection biases our estimate of the gender gap if $\Delta_P^{t_0} \neq \Delta_V^{t_0}$. If $\Delta_P^{t_0} < \Delta_V^{t_0}$, we will say selection bias causes us to *over*estimate the gender gap. If $\Delta_P^{t_0} > \Delta_V^{t_0}$, we will say selection bias causes us to *under*estimate the gender gap.

depending on the controls in the econometric model. The gender gap among our lab volunteers is similar in magnitude to that in previous behavioral experiments.

The study that is most closely related to ours is Cleave, Nikiforakis and Slonim (2013). Cleave et al. investigate the extent to which social and risk preferences drive selection into lab experiments. We use their method to address our second question, namely, whether self-selection causes us to overestimate the gender gap in competitiveness in the population. Apart from our focus on attitudes toward competition, an important difference is that Cleave et al. (2013) do not target students from a particular academic major. In particular, they target a course that is taken by economics majors, but also students majoring in Management and Marketing, Accounting, Finance, Arts, Science, Engineering, and Environments. As their data does not permit them to identify who majors in economics, they cannot explore whether there is a gender gap in risk attitudes between economics majors. Given that risk attitudes are commonly measured when studying individuals' willingness to compete, and the fact that there is evidence of gender differences in risk taking, we include a separate section in which we address this question.

The rest of this paper is organized as follows. Section 2 describes in detail the experiment. Section 3 presents the experimental results, and section 4 concludes with a discussion.

2. The experiment

2.1 The population

Our population consists of students enrolled in the *Academic Skills* course at the Erasmus School of Economics. Students in this course are taught presentation and discussion skills as well as how to study effectively. The course is compulsory for all first-year students majoring in economics, but is not available for students from different majors.[5]

A total of 1,395 students were enrolled in the course at the time of our study. This particular course has also another advantage which is that it has a large number of mandatory lab tutorials, which simplifies the administration of a large-scale computerized experiment. Specifically, students in this course are divided into 96 tutorial groups with as many tutors. Each group consists

[5] Academic Skills is compulsory for all students majoring in economics, econometrics, and law and economics. Since most universities outside the Netherlands do not offer separate majors in econometrics or law and economics, and given that all students in these majors could enrol to graduate programs in economics (if they so desire), we refer to them collectively as economics majors.

of 15 students at most, implying that the tutor can easily monitor participants and ensure they do not communicate during the experiment.

The experiment took place during the first tutorial, in the first week of the academic year 2018-2019. Participants, therefore, had no previous exposure to laboratory or other classroom experiments. To rule out that self-selection could bias our estimates, the experiment was not announced to students in advance. One week prior to the experiment, we gave tutors – who were unaware of our research question – precise instructions about how to run the experiment in class. Once in class, tutors guided students to a link on the course's webpage that directed them to the online experiment (see Section 1 in the *Online Supplementary Material, OSM*). Tutorials last 90 minutes and we purposefully limited the duration of the experiment (about 15 minutes) to ensure that students would not leave the class. If students did not wish to participate, tutors asked them to remain seated and be silent.

Of the 1,216 students that showed up for the tutorials, 1,145 (94.2%) agreed to participate in the experiment. These 1,145 individuals are our "population." The fraction of women in our population is the same as that in economics program in the US (Bayer and Rouse 2016; Buckles 2019; Lundberg and Stearns 2019): 29.7% (811 males and 334 females).[6] With regards to faculty, as in the US, there appears to be clear evidence of a "leaky pipeline": the fraction of female economists in 2021 in the Netherlands, is 42.9% among PhD students, 37.4% among Assistant Professors, 21.2% among Associate Professors and 15.4% are Professors.[7]

2.2 Measuring the willingness to compete in the population

The experiment consists of four tasks and a short post-experiment questionnaire. The first three tasks are built on the paradigm introduced by Niederle and Vesterlund (2007) (NV2007). In each of them, participants are asked to add up a series of three two-digit numbers for 90 seconds. The tasks differ in how performance translates into earnings. In Task 1 (Individual Performance), participants are paid for each correct summation using a *piece rate* of 1 Euro. In Task 2 (Compared Performance), participants compete in a *tournament* against three other anonymous students that are randomly selected from the population (not from their tutorial). The participant that correctly

[6] Erasmus University Rotterdam has data on the gender of 1,366 out of 1,395 students enrolled in the course. The fraction of women among these students is 27.2%.

[7] Source: https://www.rathenau.nl/en/science-figures/personnel/women-science/women-academia.

solves the highest number of additions receives 4 Euros per correct answer, while the other three people in the group receive nothing. Ties for the first place are randomly broken.

In the third task (Choice), participants must choose whether they wish their earnings to be determined by their individual performance (as in Task 1) or by their compared performance (as in Task 2). After making their choice, participants have to add up a series of three two-digit numbers for a final time. The Choice task measures individuals' attitudes towards competition and is, therefore, our main variable of interest. Like in NV2007, if participants choose compared performance, their score is compared to that of participants in Task 2. In Task 4 (Lottery), students are given six lotteries and have to choose the one they prefer. The task is similar to that in Eckel and Grossman (2008), but lottery payoffs are chosen such that they reflect the incentives in the Choice task. The latter is important in light of a recent debate about the extent to which differences in the Choice task simply reflect gender differences in risk attitudes (Gillen, Snowberg and Yariv 2019; van Veldhuizen, 2018). The experimental instructions are available in the *OSM*.

The experiment concluded by asking participants to fill out a short questionnaire with questions about their gender, how much time they have for leisure activities, and the income group they belong to (from a scale from 1 to 10) (see Section 2 in the *OSM*). The survey also included a question aimed to capture participants' confidence in their abilities in Task 2. In particular, they were asked to state if they agreed or not with the following statement: "I believe there is a high chance I will be the one who solved the highest number of correct sums in my group in Task 2". Given the limited time available for the classroom experiment, responses to this question were not incentivized. However, as we will see in sections 3.1 and 3.2.2, the answers obtained exhibit the same patterns as those in prior experiments with incentivized belief elicitation.

Participants were informed that one in every four participants would be selected for payment and that, if selected, one of the four tasks would be selected for payment. While participants learned their individual score at the end of each of tasks 1-3, they were not informed about whether they won the tournament/lottery, their earnings or which task had been selected for payment. Participants were told that they will receive an email with information on who was selected for payment (see Section 3 in the *OSM*).[8]

[8] Students were informed that their earnings would be determined in a public event that would take place in the second week of the academic year. Subsequently, an announcement was posted on the course's webpage with the student ID

2.3 Identifying the lab volunteers in the population

We used a four-step process to identify lab volunteers in our population. We discuss each step in detail.

(1) Expressing an interest in experiments. We asked all participants in the classroom experiment if they would be interested to receive an e-mail with information about participating in future laboratory experiments at ESE.[9] To test whether the classroom experiment affected participants' willingness to volunteer in future lab experiments, half of the participants were randomly selected to be asked this question at the very first screen of the experiment, whereas the remaining half was asked the same question on the very last screen. (A detailed analysis of the effect of this variation is presented in Appendix B.) As can be seen in Figure 1, 76.9% of the individuals in our population expressed an interest in receiving the email. The order of the question does not have a significant effect on who expresses an interest (Fisher Exact test, N = 1,145, p = 0.14) or volunteers for experiments (Fischer Exact test, N = 1,145, p = 0.38). Importantly, in Appendix B, we provide a detailed analysis showing that the *exposure* to a classroom experiment does *not* affect who selects into the lab.

(2) Registering in the database of future volunteers. Two weeks after the last tutorial, all participants that expressed an interest in future experiments received an email by the lab manager (not us) to register to the ORSEE volunteer database (Greiner, 2015) (see Section 4 in the *OSM*). The email reminded them that they had expressed an interest in participating in lab experiments and contained a link for registering to the database. Reminder mails were sent twice and students who had provided us with multiple email addresses, received emails in all addresses. Note, that by the time they were invited to register in the database, students had been informed of their individual earnings from the classroom experiment. However, they did not receive any information about which task had been selected for payment or how their earnings were determined. Figure 1 shows that, similar to Cleave, Nikiforakis and Slonim (2013), only 19.7 percent of the entire population registered in the database of volunteers.

numbers of those that were selected for payment. Students could pick up their earnings two weeks after the experiment took place, at which point, they learned their exact earnings but not of which task was selected for payment.

[9] As this was the first official recruitment drive of the academic year, it was extremely unlikely our first-year students would have signed up to the database prior to receiving this email.

(3) Signing up for an experiment. Twelve weeks after the start of the classroom experiment, an email was sent to all those in our population that registered in the database of future volunteers (see Section 5 in the *OSM*). The email included links that enabled recipients to enroll for one of 20 laboratory sessions. The sessions were spread out over a week, the first being scheduled to run 10 days after the invitation email. None of the 20 sessions filled up to the maximum. We are therefore confident that lab constraints did not preclude individuals from participating in an experiment. Despite this, as can be seen in Figure 1, less than half of those who registered in the database signed up for an experiment (i.e., 8.6% of our population).

(4) Participating in an experiment. A large majority of those signing up for an experiment, participated in one. Specifically, 7.9% of the starting population participated in a lab experiment (90 out of 1,145 individuals).[10] This, again, is similar to the fraction found in Cleave, Nikiforakis and Slonim (2013) showing that the high level of attrition is not unique to our study. Following Cleave, Nikiforakis and Slonim (2013), we classify all those that participated in a lab experiment as "lab volunteers". This seems natural for our purposes as it is these subjects that we would normally use to estimate the gender gap in competitiveness in a lab experiment.[11] This implies that our sample of lab volunteers (N=90) consists of a similar number of individuals as NV2007 (N=80).

We use the lab experiments only for the purpose of identifying who is a volunteer and who is not. We do *not* re-measure the willingness of volunteers to compete in these lab experiments as this would introduce several confounds in our comparisons (e.g., learning, different social environment). Instead, having identified who is a lab volunteer in our population, we estimate the gender gap among them using the decisions they made in the classroom experiment.

[10] Of the 90 lab volunteers, 57 are men (i.e., 7.0% of the men in our population) and 33 are women (i.e., 9.9% of the women in our population). In Appendix C, we discuss in detail attrition by gender, showing that attrition rates are similar for men and women.

[11] To ensure that we correctly identified lab volunteers in our population, we also checked a month after our lab sessions whether there were others in our population that participated in economic experiments and classified those also as "lab volunteers."

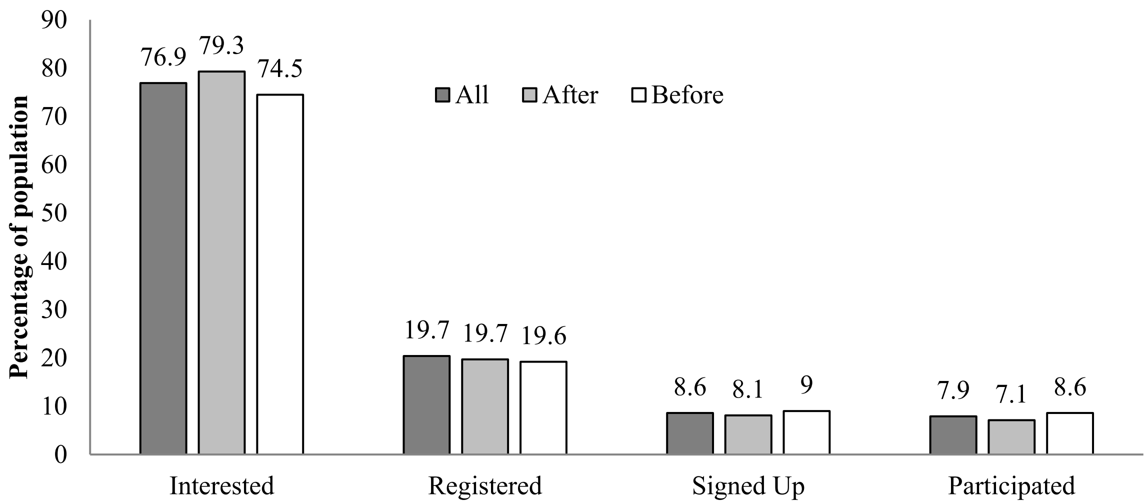


Figure 1 Attrition to the lab 'Interested' refers to the percentage of individuals in the population that indicated an interest in receiving an email with information about future experiments. 'Registered' refers to the percentage of individuals in the population that registered in the database of future volunteers. 'Signed Up' refers to the percentage of individuals in the population that signed up for one of the twenty lab sessions. 'Participated' refers to the percentage of individuals in the population that participated in an experiment – the lab volunteers. 'Before/After' refer to the order in which a question was asked during the classroom experiment about whether individuals wished to receive information about future experiments.

3 Results

We divide the analysis of our data into two subsections. The first subsection addresses our two main research questions. The second subsection presents auxiliary findings related to gender differences in risk attitudes and the propensity of men to "overcompete."

3.1 Main findings

Is there a gender gap in competitiveness among economics majors? Does self-selection cause us to overestimate it? We define the gender gap in competitiveness in the population of economics majors as $\Delta_P^{t_0} = m_P^{t_0} - w_P^{t_0}$, where $m_P^{t_0}$ and $w_P^{t_0}$ denote the fraction of men and women in the population, respectively, choosing to compete in Task 3 at time t_0, i.e., in the classroom experiment. Similarly, the gender gap in competitiveness among lab volunteers is defined as $\Delta_V^{t_0} = m_V^{t_0} - w_V^{t_0}$, where $m_V^{t_0}$ and $w_V^{t_0}$ denote the fraction of men and women lab volunteers, respectively, that choose to compete in Task 3 in the classroom experiment. We will say that selection bias causes us to overestimate the gender gap if $\Delta_V^{t_0} > \Delta_P^{t_0}$.

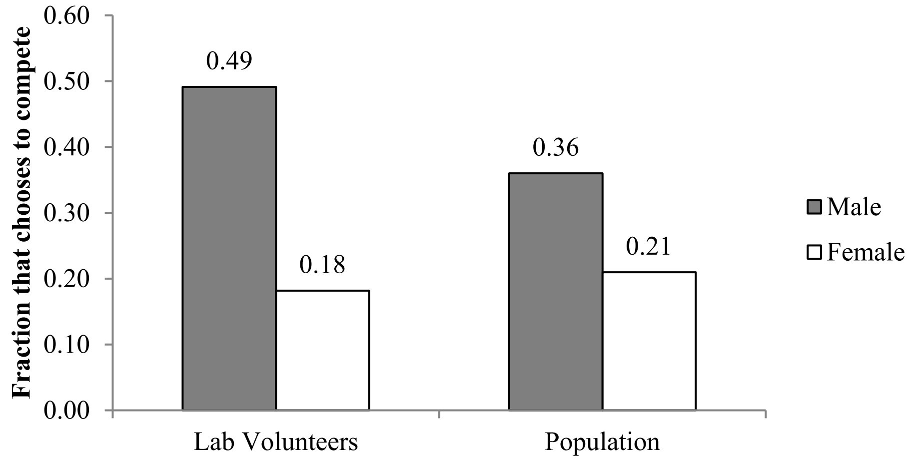


Figure 2 Gender gap in competitiveness Fraction of participants that chooses to compete in Task 3 of the experiment among volunteers and the population from which they were recruited.

Figure 2 shows the fraction of men and women that chose to compete in Task 3. In line with past studies, we find that men clearly compete more often than women in the sample of lab volunteers (49% vs. 18%; Fisher exact test, two-tailed, $N = 90$, $p < 0.01$), i.e., $\Delta_V^{t_0} = 0.31$. A significant gender difference is also observed in the population of economics majors (36% vs. 21%; Fisher exact test, two-tailed, $N = 1{,}145$, $p < 0.01$), i.e., $\Delta_P^{t_0} = 0.15$. The size of the gender gap is approximately twice as large among lab volunteers than among men and women in the population from which the volunteers were recruited, i.e., $\Delta_V^{t_0} > 2 * \Delta_P^{t_0}$.

	(I)	(II)
Female	-0.15***	-0.14***
	(0.03)	(0.03)
Volunteer		0.14**
		(0.07)
Female × Volunteer		-0.17**
		(0.08)
Constant	0.36***	0.35***
	(0.02)	(0.02)
Observations	1,145	1,145
R-squared	0.02	0.03

Table 1 Selection bias in competitiveness Linear probability regression. The dependent variable is a dummy variable indicating whether an individual competed in Task 3. Standard errors are clustered at the session level. ***/**/*: significant at the 1%/5%/10%-level.

Table 1 presents evidence from linear probability models to explore the significance of the bias in statistical terms. As a benchmark for comparison, Model (I) shows that women are 15 percentage points less likely than men to compete in the population (p <0.01). Model (II) shows that the *Female* variable is largely unchanged in the sample of non-volunteers. Most importantly, the interaction term *Female* × *Volunteer* in Model II shows that the bias in our estimate of the gender gap due to selection is substantial and statistically significant (p < 0.05). Model (II) also presents some first evidence suggesting that the bias is driven by competitive men who are more likely to volunteer for lab experiments (*Volunteer*, p = 0.04). Evidence of this can also be seen in Figure 2: the fraction of competitive women is similar in the subset of volunteers and in the population, but not the fraction of competitive men. Men who choose to compete are more common in our sample of lab volunteers, i.e., a greater fraction of competitive men selects into our sample of volunteers, i.e., $m_V^{t_0} > m_P^{t_0}$. For additional analysis on the determinants of selection into the lab, see Appendix D.

	Volunteers (I)	Volunteers (II)	Population (III)	Population (IV)
Female	-0.21**	-0.23**	-0.09***	-0.07***
	(0.09)	(0.09)	(0.03)	(0.03)
Tourn. score	0.08***	0.07***	0.05***	0.04***
	(0.02)	(0.02)	(0.01)	(0.01)
Tourn. – Piece rate score	-0.04	-0.03	-0.01	-0.01
	(0.03)	(0.03)	(0.01)	(0.01)
Confidence		0.24*		0.32***
		(0.14)		(0.04)
Constant	-0.04	0.01	0.01	0.05
	(0.14)	(0.13)	(0.03)	(0.04)
Observations	90	90	1145	1145
R-squared	0.21	0.24	0.09	0.17

Table 2 The determinants of competition Linear probability regression replicating the analysis in NV2007 (Table 6, p.1089). The dependent variable is a dummy variable indicating whether an individual competed in Task 3. 'Tourn. score' measures the numbers of correct answers in Task 2, 'Tourn. – Piece rate score' measures the difference between a subject's correct answers in Task 1 and 2, 'Confidence' is a dummy variable with value 1 if a subject believes he or she has a higher score in Task 2 than all competitors. Standard errors are clustered at the session level. ***/**/*: significant at the 1%/5%/10%-level.

The analysis in Table 1 does not control for covariates that could explain both the gender gap in competitiveness and self-selection into the lab. For instance, NV2007 show that individual performance in the summation tasks and beliefs about the probability of winning both affect positively the probability of choosing to compete in Task 3. To that end, Table 2 replicates the main statistical analysis in NV2007 (Table 6, p. 1089), separately for our sample of lab volunteers and for our population.[12]

The estimates in Table 2 reveal that, rather intuitively and in line with past studies, the propensity to select into competition increases with an individual's performance in the tournament and with their belief that they will win the tournament. This is reassuring as it suggests we can replicate past findings with our data. In line with that, once we control for performance and beliefs, we find that the estimated gender gap among lab volunteers is comparable to that reported in NV2007. Specifically, in Model (II), we find a point estimate for *Female* of -0.225 which is not dissimilar to the estimate of -0.278 reported in NV2007. More importantly, the additional controls further increase the estimated bias due to selection. Comparing the coefficient of *Female* in Model (II) and Model (IV) we observe that the absolute value of the coefficient for *Female* decreases from 0.225 (Model II) to 0.074 (Model IV). That is, when controlling for covariates, the gender gap in the sample of lab volunteers is found to be 3 times larger than that in the population, i.e., $\Delta_V^{t_0} = 3 * \Delta_P^{t_0}$. Nevertheless, as Model (III) and Model (IV) reveal, a statistically significant gender gap in competitiveness exists in the population of economics majors.

3.2 Other findings

The gender literature in behavioral economics suggests that there are robust differences in risk attitudes between men and women (Croson and Gneezy, 2009). In addition, the evidence suggests that men tend to "overcompete" and women to "undercompete" (Niederle 2016). Our data allows us to investigate whether these discrepancies are found in the population of economics majors at Erasmus University Rotterdam.

3.2.1 Are there gender differences in risk attitudes among economic majors?

Although there is extensive research documenting the existence of gender differences in risk attitudes, there is no evidence whether this is the case among economics majors as a population. Recall that we asked participants to choose one of six lotteries that vary in their expected payoff and level of risk. We can, therefore, answer this question. In addition, our data allow us to explore

[12] In Appendix E, we present summary statistics for these and other control variables. Note, that in Table 6 in NV2007 the authors present *Probit* estimates, whereas we present linear probability results to remain consistent with the remainder of our analysis. The results of Probit regressions, however, are virtually identical.

whether self-selection biases our estimates of gender differences in risk attitudes. We follow a similar approach as in the previous section.

Figure 3 shows how risk tolerant men and women in our study are. As can be seen, men are more risk tolerant than women both in our sample of lab volunteers (Mann-Whitney U test, two-tailed, $N = 90$, $p < 0.01$), as well as in our population of economics majors (Mann-Whitney U test, two-tailed, $N = 1{,}145$, $p < 0.01$). The data in Figure 3 reveal that self-selection into the lab causes us to *over*estimate the differences in risk attitudes between men and women by 63%. This appears to be driven by less risk tolerant women being more likely to volunteer for lab experiments. For additional analysis on the determinants of selection into the lab, see Appendix D.

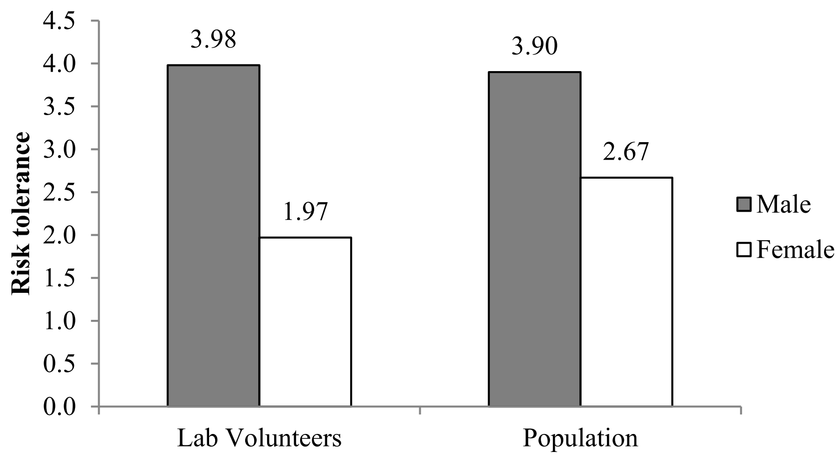


Figure 3 Mean risk tolerance

Table 3 presents evidence from linear regressions exploring the significance of the bias in statistical terms. As a benchmark for comparison, Model (I) shows that women are less risk tolerant than men in our population. Model (II) shows that the *Female* variable is similar in the sample of non-volunteers. Importantly, the interaction term in Model (II) illustrates that selection bias causes us to significantly overestimate gender differences in risk attitudes.

Given the uncertainty inherent in tournaments, it is natural to ask whether the observed *selection bias* in gender competitiveness seen in the previous section is driven by risk attitudes. To answer this question, we extend the analysis presented in Table 2 to control for risk attitudes. The findings presented in Table 4 reveal three interesting findings. First, in line with past studies, we observe that risk helps explain the willingness to compete in the lab (Model I), but also the willingness to compete in the population (Model II). Second, in line with the results in Gillen, Snowberg and Yariv (2019) and van Veldhuizen (2018), we find that risk attitudes fully account for the gender gap in competitiveness (captured by *Female*) both among lab volunteers and in the population.[13] Finally, despite this, the coefficient for the interaction term *Volunteer* × *Female* in Model (III) is unchanged from that presented in Table 2 suggesting that risk attitudes cannot account for the selection bias in competitiveness seen in the previous subsection. In other words, risk attitudes and attitudes toward competition seem to affect separately the decision to volunteer for lab experiments. This intuition is supported by our analysis presented in Appendix D.

	(I)	(II)
Female	-1.23***	-1.15***
	(0.12)	(0.12)
Volunteer		0.09
		(0.25)
Female × Volunteer		-0.87***
		(0.31)
Constant	3.90***	3.90***
	(0.08)	(0.08)
Observations	1,145	1,145
R-squared	0.0846	0.0890

Table 3 Selection bias in risk attitudes OLS regression. The dependent variable increases with the riskiness of the lottery chosen by participants. Standard errors are clustered at the session level. ***/**/*: significant at the 1%/5%/10%-level.

3.2.2 Do men compete too much and women too little among economics majors?

One of the most intriguing findings in NV2007 was that high-performing women tended to compete "too little," whereas low-performing men to compete "too much." We conclude our analysis of the data by exploring to what extent this tendency is also found in the population of economics majors.

Since individuals compete in groups of four, the probability of winning the tournament that equalizes the expected earnings of an individual from the tournament and the piece rate is 25%. Therefore, following NV2007, we refer to people with a performance in Task 2 that gives them a probability of at least 25% chance of winning the tournament as "high performers"; if their performance translates to a probability below 25%, we refer to them as "low performers." We

[13] It is worth noting that the discussion about the role of risk attitudes is ongoing as other studies have found risk attitudes do not fully account for the gap in competitiveness, e.g., Datta Gupta, Poulsen and Villeval 2013, Buser, Niederle and Oosterbeek 2014, Wozniak, Harbaugh and Mayr 2014, Dariel et al. 2017, and Lowes 2021.

follow NV2007 (Table III, p. 1085) in classifying individuals into two categories: (*i*) low performers that choose to compete ("overcompetitors"), and (*ii*) high performers that choose not to compete ("undercompetitors").

	Volunteers (I)	Population (II)	Population (III)
Female	-0.05	-0.02	-0.01
	(0.10)	(0.03)	(0.03)
Tourn. score	0.05**	0.03***	0.03***
	(0.02)	(0.01)	(0.01)
Tourn. – Piece rate score	-0.04	-0.00	-0.00
	(0.03)	(0.01)	(0.01)
Confidence	0.24*	0.31***	0.31***
	(0.13)	(0.04)	(0.04)
Risk tolerance	0.09***	0.05***	0.05***
	(0.03)	(0.01)	(0.01)
Volunteer			0.15**
			(0.06)
Volunteer × Female			-0.17**
			(0.08)
Constant	0.25	0.51***	0.51***
	(0.31)	(0.09)	(0.09)
Observations	90	1,145	1,145
R-squared	0.32	0.20	0.21

Table 4 Risk attitudes and competition OLS regression. The dependent variable a subject's choice to compete in Task 3. Standard errors are clustered at the session level. 'Tourn. score' measures the amount of correct answers in Task 2, 'Tourn. – Piece rate score' measures the difference between a subject's score in Task 1 and 2, 'Confidence' is a dummy variable taking a value of 1 if a subject believes he or she has a higher score in Task 2 than all competitors, 'Risk tolerance' is a variable that increases in an individual's willingness to take on risk, on a scale 0-5. ***/**/*: significant at the 1%/5%/10%-level.

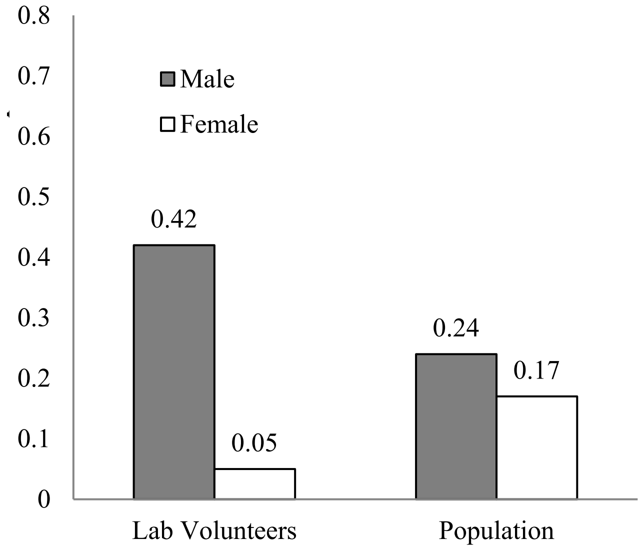


Figure 4A *Low performers*: subjects with a score that gives them a chance below 25% of winning the tournament.

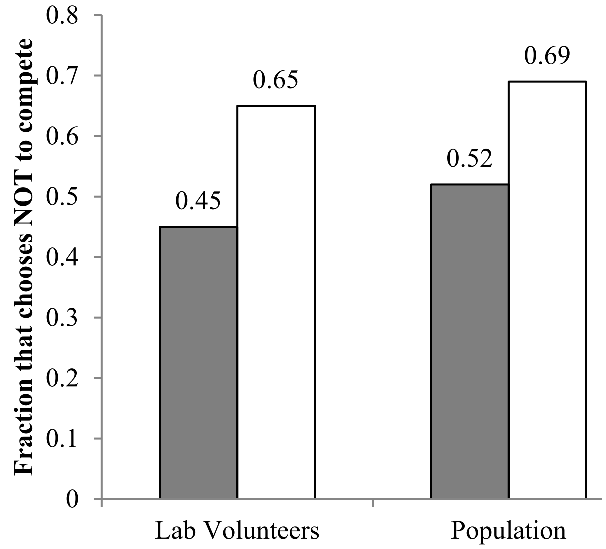


Figure 4B *High performers*: subjects with a score that gives them a greater than 25% chance of winning the tournament.

Figure 4 shows the fraction of overcompetitors (Panel A) and the fraction of undercompetitors (Panel B). Panel A reveals that, in line with NV2007, we find a substantial gender difference in overcompeting among lab volunteers (Fischer Exact test, two-tailed, $N = 45$, $p < 0.01$). The gender difference in overcompeting is also observed in the population of economics majors (Fischer Exact test, two-tailed, $N = 659$, $p = 0.05$), although it is about five times smaller than among lab volunteers. Panel B shows that selection does not seem to affect our conclusions regarding the tendency of high performers to avoid competition: In line with NV2007, high-performing women are more likely to abstain from competition than men to a similar extent in the population (Fischer Exact test, two-tailed, $N = 486$, $p < 0.01$) and among lab volunteers (Fischer Exact test, two-tailed, $N = 45$, $p = 0.34$), but the latter is not statistically significant due to the small number of high performers volunteering for experiments. Therefore, our data indicate that selection bias overestimates the willingness of low-performing men to compete relative to women, but not the tendency of high-performing women to avoid competition relative to men.

	Low performers			High performers		
	(I)	(II)	(III)	(IV)	(V)	(VI)
Female	-0.07**	-0.04	0.00	0.17***	0.18***	0.05
	(0.03)	(0.04)	(0.03)	(0.06)	(0.07)	(0.07)
Volunteer		0.20*	0.22**		-0.07	-0.08
		(0.10)	(0.09)		(0.09)	(0.08)
Volunteer × Female		-0.33***	-0.31***		0.01	-0.02
		(0.12)	(0.11)		(0.15)	(0.15)
Confidence			0.42***			-0.26***
			(0.05)			(0.04)
Risk tolerance			0.04***			-0.06***
			(0.01)			(0.01)
Constant	0.24***	0.23***	0.03	0.52***	0.52***	0.89***
	(0.02)	(0.02)	(0.04)	(0.02)	(0.02)	(0.05)
Observations	659	659	659	486	486	486
R-squared	0.01	0.02	0.17	0.02	0.02	0.15

Table 5 The propensity to overcompete (low performers) and undercompete (high performers) Linear probability model. The dependent variable is a dummy capturing a subject's choice to compete in Task 3 for 'low performers', and a subject's choice *not* to compete for 'high' performers. Standard errors are clustered at the session level. ***/**/*: significant at the 1%/5%/10%-level.

To explore the statistical significance of the bias, Table 5 presents the results from a regression analysis, similar to that in the previous subsection. Model (I) indicates that low-performing men are 7 percentage points more likely to compete than low-performing women in the population. Model (II) shows that this difference is significant among volunteers but not among non-volunteers. The interaction term (*Volunteer × Female*) in that model reveals that the selection bias is large and statistically significant. Model (III) shows that the bias cannot be fully explained by individual beliefs and attitudes toward risk. For high performers, we find women are more likely to *abstain* from competition (Model IV) and no evidence of selection bias affecting our estimates (Model V).[14] Finally, Model (VI) reveals that the relative unwillingness of high-performing women to compete can be explained by their lower tolerance to risk and confidence.

4. Discussion

Our data reveal that four gender differences commonly found in behavioral experiments with self-selected samples of volunteers are also observed in a population of economics majors. Specifically,

[14] As can be seen in Table 5, 486 individuals are classified as high performers, i.e., 42.4% of the population. The reason this number is higher than 25% seems to be that there are learning effects across tasks which, although small in magnitude, exist for many individuals. We do not find evidence of different learning effects for men and women.

we find that, compared to their female counterparts, male economics majors are (*i*) more likely to select into competition, (*ii*) more risk tolerant, (*iii*) more likely to compete when they are low-performing, and (*iv*) less likely to compete when they are high-performing as women shy away from competition. These gender differences suggest a mechanism that may help explain the persistent underrepresentation of women in economics (Bayer and Rouse 2016; Buckles 2019; Lundberg and Stearns 2019).

The data also reveal that self-selection into behavioral experiments causes us to substantially overestimate the gender differences in the underlying population in three of these four cases. Specifically, self-selection leads us to overestimate (*i*) the gender gap in competitiveness by a factor of 2 to 3 depending on the econometric model, (*ii*) the gender gap in risk attitudes by a factor of 0.63, and (*iii*) the relative tendency of men to overcompete by a factor of 5. This evidence implies that behavioral economists cannot (and should not) ignore the possible risks that self-selection poses to a study's external validity. Whenever the variable of interest is not randomized by the experimenter (e.g., gender, race, ethnicity, academic background) researchers should strive to either reduce self-selection into the sample or correct for selection bias using appropriate econometric techniques.[15] Unless this is done, one cannot rule out that a null result in a behavioral experiment is due to selection bias causing us to underestimate an actual difference in the population from which the sample was recruited or, perhaps worst, that a significant result is a false positive due to selection bias. Given this, an important implication of our study would appear to be that, despite the strong selection bias in our sample of lab volunteers, we do not find evidence that would cause us to question what appears to be the existing consensus among scholars concerning the existence of gender differences in competitiveness and risk attitudes.

[15] Although the academic background of subjects is typically not reported in papers in this literature, by virtue of using economic experiments to study attitudes toward competition, one might assume that these studies oversample economics majors.

Appendix A

The reliance of studies on self-selected samples raises the possibility that selection bias affects our estimate of the gender gap in competitiveness. Selection bias need not cause us to overestimate the gender gap in competitiveness; it could also cause us to underestimate it. As mentioned, this could happen if women in the population are less willing to compete than men, but only men and women with a certain tolerance for competition volunteer for experiments. As the latter seems plausible, we wished to look for evidence of selection bias in past studies that would allow us to form a hypothesis about the potential direction of selection bias.

Dariel et al. (2017) provide a list of all the studies using designs similar to that of NV2007 that were published (or accepted for publication) in the ten years following the publication of NV2007. We use the survey of Dariel et al. (2017) and classify previous studies into those using self-selected samples and those that do not. The classification can be found in Table A.1.

Study	Year	% selecting competition Male	Female	Gender Gap	Subjects	Self-selected
Niederle and Vesterlund	2007	73%	35%	38%	Students	Yes
Gneezy et al.	2009	50%	26%	24%	Other	Yes
Gneezy et al.	2009	39%	54%	-15%	Other	Yes
Healy and Pate	2011	81%	28%	53%	Students	Yes
Balafoutas and Sutter	2012	64%	30%	34%	Students	Yes
Balafoutas et al.	2012	59%	31%	28%	Students	Yes
Cardenas et al.	2012	44%	19%	25%	Children	No
Cardenas et al.	2012	39%	27%	12%	Children	No
Cardenas et al.	2012	35%	32%	3%	Children	No
Cardenas et al.	2012	26%	29%	-3%	Children	No
Dargnies	2012	85%	51%	34%	Students	Yes
Kamas and Preston	2012	41%	23%	18%	Students	Yes
Mayr et al.	2012	56%	36%	20%	Other	Yes
Mueller and Schwieren	2012	42%	26%	16%	Students	Yes
Price	2012	66%	49%	17%	Students	Yes
Shurchkov	2012	39%	30%	9%	Students	Yes
Shurchkov	2012	44%	19%	25%	Students	Yes
Andersen et al.	2013	52%	49%	3%	Children	Yes
Andersen et al.	2013	51%	39%	12%	Children	Yes
Cadsby et al.	2013	36%	9%	27%	Students	Yes
Datta Gupta et al.	2013	60%	34%	26%	Students	Yes
Niederle et al.	2013	74%	31%	43%	Students	Yes
Samak	2013	77%	83%	-6%	Children	No
Buser et al.	2014	49%	23%	26%	Children	No
Dreber et al.	2014	36%	17%	19%	Children	No
Dreber et al.	2014	33%	28%	5%	Children	No
Lee et al.	2014	30%	22%	8%	Children	Yes
Wozniak et al.	2014	54%	31%	23%	Students	Yes
Wozniak et al.	2014	50%	30%	20%	Students	Yes
Apicella et al.	2015	45%	30%	15%	Other	Yes
Apicella et al.	2015	52%	37%	15%	Other	Yes
Apicella et al.	2015	67%	29%	38%	Other	Yes
Brandts et al.	2015	59%	30%	29%	Students	Yes
Khachatryan et al.	2015	54%	52%	2%	Children	No
Khachatryan et al.	2015	57%	56%	1%	Children	No
Sutter and Rutzler	2015	40%	19%	21%	Children	No
Almås et al.	2016	52%	32%	20%	Children	Yes
Berlin and Dargnies	2016	63%	35%	28%	Students	Yes
Cassar et al.	2016	36%	26%	10%	Other	Yes
Sutter et al.	2016	44%	21%	23%	Children	No
Apicella et al.	2017	58%	38%	20%	Students	Yes
Buser et al.	2017	52%	28%	24%	Students	Yes
Buser et al.	2017	68%	51%	17%	Children	No
Dariel et al.	2017	50%	54%	-4%	Students	No
Halko and Saakvuori	2017	74%	54%	20%	Students	Yes
Reuben et al.	2017	54%	27%	27%	Students	Yes
Banerjee et al.	2018	22%	16%	6%	Other	Yes
Bönte et al.	2018	56%	45%	11%	Other	Yes
Buser et al.	2018	42%	26%	16%	Students	Yes
Zhong et al.	2018	49%	25%	24%	Students	Yes
Zhang et al.	2019	63%	48%	15%	Children	No
Zhang et al.	2019	60%	38%	22%	Children	No
Zhang et al.	2019	75%	48%	27%	Children	No

Table A.1 Summary and classification of studies surveyed by Dariel et al. (2017). Studies listed more than once use different samples. For details on the tasks used in these studies and sample sizes, see Dariel et al. (2017). Studies are presented in chronological order of publication. The publication year for studies that were 'forthcoming' at the time Dariel et al. (2017) was published has been updated.

Figure 1 presents the mean gender gap in competitiveness in studies relying on self-selected samples and otherwise. As can be seen in Figure 1, the mean gender gap is 21.3 percentage points in studies using self-selected samples and 12.1 percentage points in studies using non-self-selected samples. Therefore, although a gender gap in competitiveness is observed in non-self-selected samples, it is 76% *larger* in studies using self-selected samples. This suggests that selection bias causes us to substantially *over*estimate the gap in competitiveness between men and women. Specifically, men are 68.8% more likely to compete than women in the self-selected samples, and 31.6% more likely in studies in which subjects do not self-select into the experiment (Table A.1). However, one should note that the data in Figure 1 do not permit us to rule out alternative explanations for the observed difference. In fact, there is a near perfect confound in the data which prevents us from drawing safe conclusions: all but one of the studies in which samples do not self-select into the experiment, use samples of pre-university participants. Similarly, all but one of the studies using student samples rely on volunteer samples.

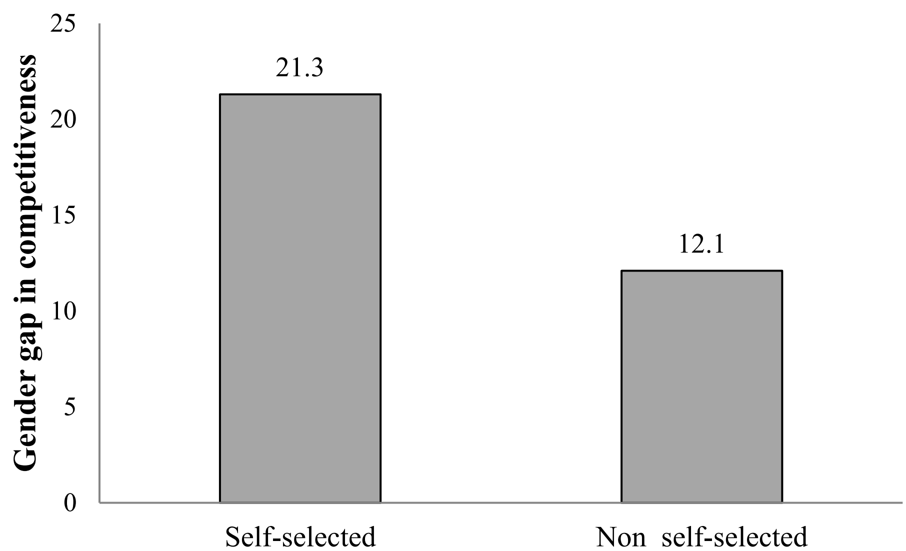


Figure A.1 Gender gap in competitiveness (in percentage points) in previous studies surveyed in Dariel, Kephart, Nikiforakis and Zenker (2017) using designs similar to that in NV2007. *Self-selected* samples (N = 36) are those that could self-select into the experiment. *Non-self-selected* samples (N = 17) are typically those in classroom experiments when participants did not receive prior information about the experiment. The gender gap is calculated as the unweighted average across different samples of the difference in the fraction of men and women choosing to compete in a design similar to that in NV2007.

Appendix B

One might be concerned that exposure to a specific classroom experiment can affect who selects into the lab. For instance, one may hypothesize that individuals that enjoy competing may be more likely to volunteer for lab experiments after participating in the classroom experiment. First, we note that, even if this hypothesis was true, it could not explain the selection bias we observe in our study. Second, we note that the evidence in Cleave et al. (2013) does not support this hypothesis. The authors of that study randomly assigned a group of individuals in a condition in which they did not participate in the classroom experiment; they were simply invited to sign up for lab experiments. Cleave et al. (2013) found that these individuals were as likely to volunteer for experiments as others in groups that had participated in classroom experiments. Of course, there could be an interaction effect that this analysis overlooks. Namely, those who enjoy (dislike) risk taking are more (less) likely to select into the lab than those in the control group.

To test the hypothesis that individuals that enjoy competing may be more likely to volunteer for lab experiments after participating in the classroom experiment, half of the participants were randomly selected to be asked this question at the very first screen of the experiment, whereas the remaining half was asked the same question on the very last screen. If the hypothesis is true, we would expect that an individual who enjoys competing to be more likely to express an interest in experiments when the question was posed *after* the experiment. Conversely, we would expect that an individual that dislikes competition would be more likely to express an interest in experiments when the question was posed *before* the experiment.

Table B.1 shows we do not find support for this hypothesis (Model I and Model III). The same holds if we control for a self-reported measure of one's leisure time (*Leisure*), a self-reported measure of one's income (*Income*), and, finally, a control for whether the individual was selected to be paid for their participation in the classroom experiment (*Paid*) (Model II and Model IV). Tables B.2 and B.3 show that we do not find support for this hypothesis if, instead of the willingness to express an interest for experiments, we consider the willingness to sign up to the database for volunteers or show up for the experiment.

	(I)	(II)	(III)	(IV)
Before	-0.024	-0.023	0.011	0.010
	(0.049)	(0.049)	(0.029)	(0.030)
Leisure		0.027		0.004
		(0.028)		(0.017)
Income		-0.006		-0.004
		(0.011)		(0.006)
Paid		-0.003		0.090^{**}
		(0.052)		(0.039)
Constant	0.249^{***}	0.197	0.172^{***}	0.159^{**}
	(0.037)	(0.123)	(0.020)	(0.068)
Sample: Only those that chose …	To compete	To compete	Not to compete	Not to compete
Observations	362	362	783	783
R^2	0.0008	0.0044	0.0002	0.0107

Table B.1 The effect of sequencing on the probability of registering to the database of volunteers. Linear probability models. The dependent variable is a dummy about whether a subject registered in the database of volunteers. Standard errors are clustered at the session level. ***/**/*: significant at the 1%/5%/10%-level.

	(I)	(II)	(III)	(IV)
Before	-0.020	-0.019	0.022	0.021
	(0.035)	(0.035)	(0.024)	(0.025)
Leisure		0.007		-0.002
		(0.019)		(0.013)
Income		-0.008		-0.003
		(0.008)		(0.004)
Paid		0.012		0.030
		(0.033)		(0.024)
Constant	0.109^{***}	0.133^{*}	0.068^{***}	0.087^{*}
	(0.026)	(0.071)	(0.015)	(0.050)
Sample: Only those that chose …	To compete	To compete	Not to compete	Not to compete
Observations	362	362	783	783
R^2	0.0011	0.0050	0.0016	0.0044

Table B.2 The effect of sequencing on the probability of signing up for a lab experiment. Linear probability models. The dependent variable is a dummy about whether a subject signed up for a lab experiment (irrespective of whether they showed up for the experiment on the day). Standard errors are clustered at the session level. ***/**/*: significant at the 1%/5%/10%-level.

	(I)	(II)	(III)	(IV)
Before	-0.010	-0.008	0.027	0.026
	(0.032)	(0.032)	(0.022)	(0.023)
Leisure		0.002		0.000
		(0.018)		(0.011)
Income		-0.010		-0.004
		(0.007)		(0.004)
Paid		0.021		0.026
		(0.036)		(0.023)
Constant	0.098^{***}	0.148^{**}	0.058^{***}	0.077
	(0.022)	(0.071)	(0.015)	(0.047)
Sample: Only those that chose …	To compete	To compete	Not to compete	Not to compete
Observations	362	362	783	783
R^2	0.0003	0.0068	0.0027	0.0054

Table B.3 The effect of sequencing on the probability of participating in a lab experiment. Linear probability models. The dependent variable is a dummy about whether a subject participated in a lab experiment. Standard errors are clustered at the session level. ***/**/*: significant at the 1%/5%/10%-level.

Appendix C

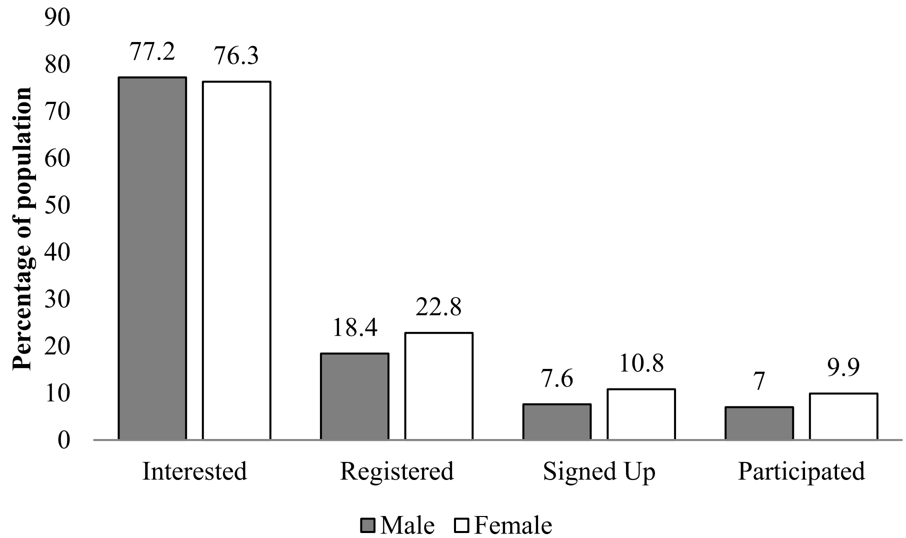


Figure C.1 Population attrition by gender 'Interested' refers to the percentage of the population who indicated interest to receive an email with more information about the ORSEE database. 'Registered' refers to the percentage of the population who registered in the ORSEE database. 'Signed Up' refers to the percentage of the population who signed up for one of the lab sessions. 'Participated' refers to the percentage of the population who showed up in one of the sessions.

Appendix D

The evidence of selection bias in sections 3.1 and 3.2 raises the question of what factors influence the decision of individuals to volunteer for experiments and how these differ for men and women. Already in sections 3.1 and 3.2, we saw evidence suggesting that different factors may account for the willingness of men and women to select into the lab. Here, we expand our analysis to consider the effect of competitiveness and risk attitudes on selection, while controlling for a number of other covariates.

Table D.1 presents the results from a linear probability model where the dependent variable is a participant's choice to volunteer for lab experiments. The independent variables include our two main behavioral measures – that of an individual's competitiveness (*Compete*) and that of their willingness to take on risk (*Risk tolerance*) – a self-reported measure of one's leisure time (*Leisure*), a self-reported measure of their income (*Income*), a control for whether they were asked to express their interest for future lab experiments before the classroom experiment (*Before*), and, finally, a control for whether they were selected to be paid for their participation in the classroom experiment (*Paid*).

The results in Table D.1 reveal that different factors drive the decision of men and women in our population to volunteer for experiments. For men, only the willingness to compete is a significant predictor of their choice to volunteer; risk attitudes appear to have no influence in their decision. For women, on the other hand, it is competitiveness that seems to have no influence on their willingness to volunteer; only their risk attitudes appear to determine selection as well as whether they were selected for payment. These findings are important for experimental economists, not only because we find evidence that variables we are often interested in measuring such as competitiveness and risk attitudes affect selection, but also because men and women appear to select differently into the lab raising challenging questions for experimental studies on gender differences.

	Male	Female
Compete	0.04**	0.00
	(0.02)	(0.04)
Risk tolerance	-0.00	-0.03***
	(0.01)	(0.01)
Leisure	-0.00	0.03
	(0.01)	(0.02)
Income	-0.01	-0.00
	(0.00)	(0.01)
Before	0.01	0.02
	(0.02)	(0.03)
Paid	0.00	0.08**
	(0.02)	(0.04)
Constant	0.10**	0.07
	(0.05)	(0.10)
Observations	811	334
R-squared	0.09	0.04

Table D.1 The determinants of selecting into the lab for men and women Linear probability models. The dependent variable is a dummy whether a subject volunteered for lab experiments. Standard errors are clustered at the session level. ***/**/*: significant at the 1%/5%/10%-level.

Appendix E

Table E.1 presents summary statistics for the other variables collected during the experiment. In the population, males solve a higher number of math problems and are more willing to take risks (all are significant at $p < 0.01$ using Mann-Whitney U tests). The same holds for the sample of volunteers (Mann-Whitney U test at $p < 0.01$ for task 1 and task 4, and $p = 0.09$ for task 2). Men in the population also report to be more confident than women, and to have more leisure and higher income (Mann-Whitney U tests, all significant at $p < 0.01$). For the sample of volunteers, however, we do not find significant differences between men and women for confidence, leisure and income (Mann-Whitney U tests, $p > 0.29$).

	Males		**Females**	
	Population ($N = 811$)	Lab volunteers ($N = 57$)	Population ($N = 334$)	Lab volunteers ($N = 33$)
Task 1 (Piece rate)	5.84	6.18	4.68	4.66
	(2.55)	(2.34)	(2.30)	(2.13)
Task 2 (Tournament)	6.44	6.82	5.39	6.00
	(2.48)	(2.22)	(2.04)	(2.15)
Task 4 (Risk)	3.90	3.98	2.67	1.97
	(1.92)	(1.80)	(1.62)	(0.88)
Confidence	0.28	0.21	0.15	0.18
	(0.44)	(0.41)	(0.36)	(0.39)
Leisure	3.32	3.32	3.01	3.15
	(0.88)	(0.93)	(0.81)	(0.76)
Income	6.15	5.80	5.75	5.61
	(2.02)	(1.92)	(1.87)	(1.68)

Table E.1 Average score in Tasks 1, 2 and 4 (standard deviation in parentheses). For Tasks 1 and 2 the score reflects the amount of correctly solved math problems in 90 seconds. Task 4 is a risk task that ranges from 1 (taking no risk) to 6 (taking the maximum amount of risk). *Confident* is a dummy variable indicating whether a subject believes s/he has outperformed the group member in Task 2. *Leisure* is a self-reported measure (on a scale of 1-5) of how much leisure a subject perceives to have. Finally, *Income* is a self-reported measure (on a scale of 1-10) regarding the income bracket a subject thinks s/he belongs.

References

Almås, I., Cappelen, A.W., Salvanes, K.G., Sørensen, E., & Tungodden, B. (2016). Willingness to compete: Family matters. *Management Science*, 62(8): 2149–2162.

Andersen, S., Ertac, S., Gneezy, U., List, J.A., & Maximiano, S. (2013). Gender, competitiveness, and socialization at a young age: Evidence from a matrilineal and a patriarchal society. *Review of Economics and Statistics*, 95(4): 1438–1443.

Apicella, C.L., Demiral, E.E., & Mollerstrom, J. (2017). No gender difference in willingness to compete when competing against self. *American Economic Review*, 107(5): 136-140.

Auriol, E., Friebel, G., & Wilhelm, S. (2019). Women in european economics. *VoxEU CEPR.* Available at: https://voxeu.org/article/women-european-economics.

Balafoutas, L., Kerschbamer, R., & Sutter, M. (2012). Distributional preferences and competitive behavior. *Journal of Economic Behavior & Organization*, 83(1): 125–135.

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

Banerjee, R., Gupta, N.D., & Villeval, M.C. (2018). The spillover effects of affirmative action on competitiveness and unethical behavior. *European Economic Review*, 101: 567-604.

Bateman, V., Gamage, D.K., Hengel, E., & Liu, X. (2021). The gender imbalance in UK economics. *Royal Economic Society, Silver Anniversary Women's Committee Report.* Available at: https://www.res.org.uk/uploads/assets/575c241a-fbff-4ef4-97e7066fcb7597e0/women-in-academic-economics-report-FINAL.pdf

Bayer, A., & Rouse, C.E. (2016). Diversity in the economics profession: A new attack on an old problem. *Journal of Economic Perspectives*, 30(4): 221–242.

Becker, G. (1985). Human capital, effort, and the sexual division of labor. *Journal of Labor Economics*, 3(1, Part 2): S33–S58.

Becker, G. (1991). *A treatise on the family*. Cambridge, Massachusetts and London, England.

Berlin, N., & Dargnies, M.P. (2016). Gender differences in reaction to feedback and willingness to compete. *Journal of Economic Behavior & Organization*, 130: 320–336.

Bertrand, M. (2020). Gender in the 21st Century. *American Economic Association Papers and Proceedings*, 110 (5): 1-24.

Blau, F.D., Kahn, L.M. (2017). The gender wage gap: Extent, trends, and explanations. *Journal of Economic Literature*, 55: 789-865.

Bönte, W., Procher, V., & Urbig, D. (2018). Gender differences in selection into self-competition. *Applied Economics Letters*, 25(8): 539-543.

Brandts, J., Groenert, V., & Rott, C. (2015). The impact of advice on women's and men's selection into competition. *Management Science*, 61(5): 1018–1035.

Buckles, K. (2019). Fixing the Leaky Pipeline: Strategies for Making Economics Work for Women at Every Stage. *Journal of Economic Perspectives*, 33(1): 43-60.

Buser, T., Gerhards, L., & Van Der Weele, J. (2018). Responsiveness to feedback as a personal trait. *Journal of Risk and Uncertainty*, 56(2): 165-192.

Buser, T., Niederle, M., & Oosterbeek, H. (2014). Gender, competitiveness, and career choices. *The Quarterly Journal of Economics*, 129(3): 1409–1447.

Buser, T., Peter, N., & Wolter, S.C. (2017). Gender, competitiveness, and study choices in high school: Evidence from Switzerland. *American Economic Review*, 107(5): 125-130.

Cadsby, C.B., Servátka, M., & Song, F. (2013). How competitive are female professionals? A tale of identity conflict. *Journal of Economic Behavior & Organization*, 92: 284–303.

Cárdenas, J.C., Dreber, A., Von Essen, E., & Ranehill, E. (2012). Gender differences in competitiveness and risk taking: Comparing children in Colombia and Sweden. *Journal of Economic Behavior & Organization*, 83(1): 11–23.

Cassar, A., Wordofa, F., & Zhang, Y.J. (2016). Competing for the benefit of offspring eliminates the gender gap in competitiveness. *Proceedings of the National Academy of Sciences*, 113(19): 5201–5205.

Cleave, B.L., Nikiforakis, N., & Slonim, R. (2013). Is there selection bias in laboratory experiments? The case of social and risk preferences. *Experimental Economics*, 16(3): 372-382.

Croson, R., & Gneezy, U. (2009). Gender differences in preferences. *Journal of Economic Literature*, 47(2): 448-474.

Dargnies, M.P. (2012). Men too sometimes shy away from competition: The case of team competition. *Management Science*, 58(11): 1982–2000.

Dariel, A., Kephart, C., Nikiforakis, N., & Zenker, C. (2017). Emirati women do not shy away from competition: Evidence from a patriarchal society in transition. *Journal of the Economic Science Association*, 3(2): 121-136.

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

Eckel, C.C., & Grossman, P.J. (2008). Men, women and risk aversion: Experimental evidence. *Handbook of Experimental Economics Results*, 1: 1061-1073.

Falk, A., & Heckman, J.J. (2009). Lab experiments are a major source of knowledge in the social sciences. *Science*, 326(5952): 535-538.

Falk, A., Meier, S., & Zehnder, C. (2013). Do lab experiments misrepresent social preferences? The case of self-selected student samples. *Journal of the European Economic Association*, 11(4): 839-852.

Gillen, B., Snowberg, E., & Yariv, L. (2019). Experimenting with Measurement Error: Techniques with Applications to the Caltech Cohort Study. *Journal of Political Economy*, 127(4): 1826-1863.

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

Goldin, C. (2014). A grand gender convergence: Its last chapter. *American Economic Review*, 104(4): 1091-1119.

Greiner, B. (2015). Subject pool recruitment procedures: organizing experiments with ORSEE. *Journal of the Economic Science Association*, 1(1): 114-125.

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

Halko, M.L., & Sääksvuori, L. (2017). Competitive behavior, stress, and gender. *Journal of Economic Behavior & Organization*, 141: 96–109.

Healy, A., & Pate, J. (2011). Can teams help to close the gender competition gap? *The Economic Journal*, 121(555): 1192–1204.

Harrison, G.W., Lau, M.I., & Rutström, E.E. (2009). Risk attitudes, randomization to treatment, and self-selection into experiments. *Journal of Economic Behavior & Organization*, 70(3): 498-507.

Ip, E., Leibbrandt, A., & Vecci, J. (2019). How do gender quotas affect workplace relationships? Complementary evidence from a representative survey and labor market experiments. *Management Science*, 66(2): 805-822.

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

Khachatryan, K., Dreber, A., Von Essen, E., & Ranehill, E. (2015). Gender and preferences at a young age: Evidence from Armenia. *Journal of Economic Behavior &*

Organization, 118: 318–332.

Kocher, M.G., Schindler, D., Trautmann, S.T., & Xu, Y. (2019). Risk, time pressure, and selection effects. *Experimental Economics*, 22(1): 216-246.

Lee, S., Niederle, M., & Kang, N. (2014). Do single-sex schools make girls more competitive? *Economics Letters*, 124(3): 474–477.

Leibbrandt, A., Wang, L.C., & Foo, C. (2017). Gender Quotas, Competitions, and Peer Review: Experimental Evidence on the Backlash Against Women. *Management Science*, 64(8): 3501–3516.

Levitt, S.D., & List, J.A. (2007). What do laboratory experiments measuring social preferences reveal about the real world? *Journal of Economic Perspectives*, 21(2): 153-174.

Lundberg, S., & Stearns, J. (2019). Women in economics: stalled progress. *Journal of Economic Perspectives*, 33(1): 3–22.

Lowes, S. (2021). Kinship structure, stress, and the gender gap in competition. *Journal of Economic Behavior & Organization*, 192: 36-57.

Mayr, U., Wozniak, D., Davidson, C., Kuhns, D., & Harbaugh, W.T. (2012). Competitiveness across the life span: The feisty fifties. *Psychology and Aging*, 27(2): 278–285.

Megalokonomou, R., Vidal-Fernandez, M., & Yengin, D. (2021). Underrepresentation of Women in undergraduate economics degrees in Europe: A comparison with STEM and business. *IZA Policy Paper No. 175*.

Müller, J., & Schwieren, C. (2012). Can personality explain what is underlying women's unwillingness to compete? *Journal of Economic Psychology*, 33(3): 448–460.

Neumark, D., Bank, R.J., & Van Nort, K.D. (1996). Sex Discrimination in Restaurant Hiring: An Audit Study. *The Quarterly Journal of Economics*, 111(3): 915–41.

Niederle, M. (2016). Gender. In the *Handbook of Experimental Economics*. Second edition. Eds. John Kagel and Alvin E. Roth. Princeton University Press.

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

Price, C.R. (2012). Gender, competition, and managerial decisions. *Management Science*, 58(1), 114–122.

Reuben, E., Wiswall, W., & Zafar, B. (2017). Preferences and Biases in Educational Choices and Labor Market Expectations: Shrinking the Black Box of Gender. *The Economic Journal*, 127: 2153-2186.

Samak, A.C. (2013). Is there a gender gap in preschoolers' competitiveness? An experiment in the US. *Journal of Economic Behavior & Organization*, 92: 22–31.

Shurchkov, O. (2012). Under pressure: Gender differences in output quality and quantity under competition and time constraints. *Journal of the European Economic Association*, 10(5): 1189–1213.

Sutter, M., & Glätzle-Rützler, D. (2015). Gender differences in the willingness to compete emerge early in life and persist. *Management Science*, 61(10): 2339–2354.

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

van Veldhuizen, R. (2018). Gender differences in tournament choices: Risk preferences, overconfidence or competitiveness? *LMU Discussion Paper No. 14*.

Wozniak, D., Harbaugh, W.T., & Mayr, U. (2014). The menstrual cycle and performance feedback alter gender differences in competitive choices. *Journal of Labor Economics*, 32(1): 161-198.

Zhang, Y.J. (2019). Culture, institutions and the gender gap in competitive inclination: Evidence from the communist experiment in China. *The Economic Journal*, 129(617): 509-552.

Zhong, S., Shalev, I., Koh, D., Ebstein, R.P., & Chew, S.H. (2018). Competitiveness and stress. *International Economic Review*, 59(3): 1263-1281.