Economic Inquiry

GENDER DIFFERENCES IN EXAMINATION BEHAVIOR

LENA NEKBY, PETER SKOGMAN THOURSIE and LARS VAHTRIK*

A unique examination strategy in first-year microeconomics courses is used to test for gender differences in examination behavior. Students have the possibility of attaining a seminar bonus on the final exam for near-perfect seminar attendance and are given two voluntary initial quizzes during the semester. At the final exam, the scores received on initial quizzes can either be accepted as is, or students can attempt to improve their marks by answering similar quiz questions on the exam. Results suggest that female students are more likely to take initial quizzes and receive a seminar bonus but are less likely to re-take quiz questions on the final exam. These results suggest higher risk aversion, less overconfidence, and more self-discipline or less procrastination among female students relative to male students. Our estimated behavioral differences may have important implications in terms of final grades on the course. (*JEL* I21, J16, A12, A14)

I. INTRODUCTION

Although women have made major educational advances over time, significant gender differences in labor market outcomes persist. Indeed gender gaps have been remarkably stable over the last decade or so (Altonji and Blank 1999; Blau 2012; Blau and Kahn 2007). While the early literature focused on discrimination and supply-side differences to explain labor market gaps, the more recent literature has instead analyzed the possible role of gender differences in preferences (see Bertrand 2011 and Croson and Gneezy 2009 for excellent overviews of this literature). Women, in this largely experimental literature, have been found to be more risk-averse on average, less willing to compete, and less overconfident than likewise men. As these studies often find large effects, gender differences in preferences have the potential of explaining a significant proportion of gender gaps in the labor market. If women are, for example, more reluctant to compete, they may be less likely to seek promotions or negotiate for higher wages within their occupation. They may also shy away from lucrative competitive occupations like those within banking and finance. A parallel literature that focuses on non-cognitive traits emphasizes conscientiousness as an important predictor of labor market outcomes and social behavior (Almlund et al. 2011; Heckman, Pinto, and Savelyev 2013; Heckman, Stixrud, and Urzua 2006). Specifically, Duckworth and Seeligman (2006) show that a facet of this trait such as self-discipline can explain the fact women typically receive higher grades than predicted by their performance on ability or achievement tests, and men receive lower than expected grades.[1]

Despite convincing laboratory evidence of gender differences in preferences and personality traits, there is a "striking" lack of research in real settings establishing the empirical relevance of these factors for actual outcomes (Bertrand 2011).[2] The few real world studies available focus primarily on competitive environments and

*The authors are grateful for comments from Antonio Cabrales, Lennart Erixon, Erik Grönqvist, Jonas Karlsson, Tuomas Pekkarinen, Arthur Schram, Björn Öckert, and seminar participants at the 2009 EAEPE annual conference, the 2010 COSME-FEDEA Annual Workshop on Gender Economics, the 2013 SOLE annual meeting, and the 2013 ESPE conference.

Nekby: Department of Economics, Stockholm University, S-10691 Stockholm, Sweden. Phone 468164481, Fax 468159482, E-mail lena.nekby@ne.su.se

Skogman Thoursie: Department of Economics, Stockholm University, S-10691 Stockholm, Sweden. Phone 46768232842; 468163048, Fax 468159482, E-mail peter.thoursie@ne.su.se; pthoursie@gmail.com

Vahtrik: Department of Economics, Stockholm University, S-10691 Stockholm, Sweden. Phone 468163486, Fax 468159482, E-mail lv@ne.su.se

1. Duckworth and Seeligman (2006) use the terms self-discipline and self-control interchangeably. Subsequently we will use self-discipline rather than self-control to label the trait factor relevant to our study.

2. Notably exceptions are Duckworth and Seeligman (2006) and Heckman, Pinto, and Savelyev (2013), which are studies on personal traits in real settings.

Economic Inquiry
(ISSN 0095-2583)
Vol. 53, No. 1, January 2015, 352–364

doi:10.1111/ecin.12130
Online Early publication July 24, 2014
© 2014 Western Economic Association International

show mixed results.[3] To what extent gender differences prevail in real settings, and especially in non-competitive environments, is therefore still an unresolved question.

There are several reasons why laboratory results may differ from real world settings. Sample sizes in experiments are typically small implying that large effects are needed in order to show significant results. Experiments may thus be designed to magnify differences and in the process eliminate other potentially mitigating factors (Lazear, Malmendier, and Weber 2004; Manning and Saidi 2010). In the real world, most tasks are not completed instantaneously but rather over a longer duration and workers have time to plan, receive feedback, observe co-workers, learn, adjust strategy, get advice, talk to mentors, and so forth. Another factor concerns the size of the stakes involved, which have has also been found to be important for outcomes. In the laboratory, stakes tend to be small (Antonovics, Arcidiacono, and Walsh 2005, 2009). Finally, it is not always clear how participants are selected into laboratory experiments. Participants are not typically randomly selected or necessarily representative of the underlying population, putting into question to what degree results can be extrapolated to a larger population. This might explain why a similar experimental design can yield different results in different rounds, at different time periods, or in other contexts.

In this article, we study gender differences in examination behavior among undergraduate students in a first-year economics course at Stockholm University. Students in this course face a number of decisions regarding examination strategy, both prior to and during the actual final examination, with potential repercussions for actual educational outcomes. Studying gender differences in this academic setting has several advantages. First, we depart from a representative sample of all undergraduate students who have taken this course over a period of five semesters. Second, we examine gender differences in behavior in an environment that involves risk and judgment about own ability but is otherwise non-competitive in the sense that all grading is based on absolute and not relative performance. Finally, students in our sample have already completed the course and are therefore not influenced by our study. The choices students made during the course are choices based only on the incentives developed for this course and offered to all course participants.

Students in the first-year microeconomic courses at Stockholm University are offered two voluntary quizzes, the scores of which are credited to the final exam. Quiz-takers have the opportunity of re-taking corresponding quiz questions on the final exam to improve (or lower) their earlier established scores. Students can also be awarded a seminar bonus on the final examination for seminar attendance. This set-up allows us to investigate gender differences in examination behavior in three ways. First, taking a quiz is one way to increase the probability of higher exam scores because quiz-takers can redo the quiz on the exam (i.e., have a second chance). Second, seminar attendance yields both an examination bonus and more time on the final exam. Third, quiz-takers have the opportunity of improving their scores by redoing one or both of the quizzes on the exam.

On the basis of individual information on 2,121 students who completed the course during the five semesters from 2006 to 2008, female students are found to be more likely to take one or both of the initial quizzes offered during the course, all else equal, and are also more likely to receive a seminar bonus. During the actual exam, female quiz-takers are less likely to re-take corresponding quiz questions than male quiz-takers. All three strategies (seminar attendance, quiz-taking, and quiz re-taking) are correlated with higher final exam scores implying that female students win via their higher propensity to take initial quizzes and their higher seminar attendance, but lose due to their lower propensity to re-take quiz questions on the final exam.

This paper contributes to the literature on gender differences in behavior in several ways. First, our study analyzes gender differences in behavior in a non-competitive setting in the sense that grading is based on absolute performance only. Although we may not be able to completely separate between different possible mechanisms behind observed gender differences such as risk aversion, overconfidence, self-discipline, and procrastination, we discuss to what degree these mechanisms are consistent with reported results.

3. Flory, Leibbrandt, and List (2010) use a field experiment showing gender differences in the propensity to seek jobs as pay becomes increasingly based on individual performance. Gender differences in competitive university admissions are found by Örs, Palomino, and Peyrache (2013) and Jurajda and Münich (2011). On the other hand, Manning and Saidi (2010) and Lavy (2013) find small gender differences in incidence of performance pay/bonuses and also in earnings and work effort under these contracts.

Second, we have rich information on students, including grades in the high school math and English courses that are a pre-requisite for this undergraduate course and can therefore account for any systematic underlying selection in relevant (for the course) skills by gender. Third, our study complements the recent literature on behavior in real world settings. While laboratory experiments generally find large effects—for example Dohmen and Falk (2011) find that women are 15% less likely to enter a variable pay scheme—we find gender differences in the magnitude of 3% to 8%. Our results are therefore in line with Manning and Saidi (2010) suggesting that gender differences in behavior are considerably smaller than those generally found in laboratory experiments. Our study is also related to a broader literature on gender differences in higher education (Goldin, Katz, and Kuziemko 2006).

The rest of the paper is organized as follows. In Section II we describe the exam procedure and in Section III the data and the empirical set-up. Results are reported in Section IV followed by a discussion in Section V. The article is concluded in Section VI.

II. THE EXAM PROCEDURE

The examination procedure in the microeconomics course involves a number of choices that students must make, each of which can be influenced by a number of personal traits and preferences such as risk aversion, overconfidence, procrastination, and self-discipline. In this section we describe each of these choices and the underlying factors which may influence them.

Students that register for microeconomics make a number of sequential choices during the course. The first two choices (after 2 and 4 weeks, respectively in a 9-week course) are whether or not to take two voluntary quizzes. The choice to take the second quiz is independent of the choice to take the first quiz. Each quiz consists of ten multiple-choice questions with a maximum possible score of ten points per quiz. The scores of these initial quizzes are credited to the final exam (scores of the quizzes taken during the course were announced shortly after they were conducted and thus, prior to the final exam). An additional advantage of taking one or both quizzes is that quiz-takers have the opportunity to re-take quizzes on the final exam in an attempt to improve their initial score.

The cost of studying for these quizzes must be considered low as students know that initial scores are not binding. There is a potential opportunity cost of taking quizzes in terms of time spent traveling to the university and time spent on actually taking the quiz. This opportunity cost will naturally vary between students but given the same opportunity cost, rational risk averse students should be more prone to take one or both of the quizzes. If one in addition considers the benefits of quiz-taking—sure points and more time for other exam questions—the dominant strategy should be to take initial voluntary quizzes. More (over)confident students may, however, forgo the second chance that initial quiz-taking provides. Likewise procrastinating students, or students lacking self-discipline, may find that the expected value of attending quizzes is relatively low due to low investments in studying during the early weeks of the course.[4]

Independent of the choice of taking quizzes, students also have the choice to participate in a seminar series. Active participation (passed assignments and high attendance) is rewarded with an additional ten points on the final exam. The cost, in terms of time and effort, of participating in the seminar series and attaining the seminar bonus must be considered higher than that of taking quizzes.[5] Students who do not attend seminars have the option of answering an additional question on the exam with a maximum possible score of ten points. Again, given the same opportunity cost, rational risk averse students will be more prone to participate in the seminar series. More (over)confident and procrastinating students may instead choose to abstain from seminar participation and focus on the ten-point "seminar" question on the final exam. The difference is that seminar attendees automatically get ten points (and more time to devote to other questions) while non-seminar attendees must spend time answering this question and risk receiving less than ten points.

Quiz-takers have the option to re-take the quiz question on the final exam in an attempt to improve initial scores. Students who did not

4. Note that psychological attributes such as risk aversion, overconfidence and procrastination are not always separable. A student may procrastinate due to overconfidence about his/her ability to pass the course. Likewise, a student may be more risk averse due to low confidence about his/her abilities.

5. The benefit of seminar attendance may be higher than ten points on the final exam as other exam questions are often on topics covered in the seminar series.

take initial quizzes are required to answer the quiz questions on the final exam or receive zero points. If eligible students chose to re-take a quiz question, the scores received on the exam are final, regardless of whether scores are improved or not. As such, re-taking quizzes on the final exam implies a risk of lowering initial quiz scores and more risk averse students may naturally shy away from this opportunity.[6] Confident and over-confident students will instead to a larger extent take the opportunity. Procrastination (or lack of self-discipline) may also play a role as such students may have studied more in the latter weeks of the course implying that their knowledge may be more "fresh."[7]

In summary, the final exam consists of two ten-point multiple-choice questions akin to the two initial quizzes, one ten-point open-ended seminar credit question (not answered by students who attained the seminar bonus), and seven ten-point open-ended questions to be answered by all students. The maximum possible score on the final exam is 100 points. Students have 5 hours to complete the final exam and grading is based on absolute performance only. Note that grading of final exams is blind in the sense that all personal information on students is removed before grading. This is important as there is a literature suggesting that differences in educational outcomes may be attributed to differences in non-cognitive skills (Cornwell, Mustard, and Van Parys 2013; Jacob 2002). In our setting, teachers lecture to a large group of students and are not involved in seminars (which are organized by teaching assistants). There is therefore little opportunity for personal contact with students. This together with the anonymous grading protocol implies that there is little or no opportunity for differences in non-cognitive skills among students to directly influence grading in our setting. This together with the anonymous grading protocol implies that there is little or no opportunity for differences in non-cognitive skills among students to directly influence grading in our setting. However, it should be noted that non-cognitive skills and especially self-discipline could still affect learning and thereby indirectly exam results. We would expect this effect to be larger for female students than for male students.[8]

III. DATA AND EMPIRICAL SETUP

A. *Data*

Data stems from individual information on quizzes and exams taken during first-year microeconomics courses at the Department of Economics, Stockholm University, from the fall term 2006 to the spring term 2008. In total, 2,349 students were enrolled in this course during this time period and took the final exam. We restrict the analysis to the 2,111 students (47% of which are female) with background information on the high school math and English courses that are a pre-requisite for this course. Students at Stockholm University have the possibility of taking the first-year microeconomics course via one of nine different academic programs or by registering independently for the course. In our sample, 58% of the female students and 56% of the male students are enrolled through programs, the remainder independently.[9]

In the empirical analysis two different samples are used. The first sample consists of all students who participated in the course and took the final exam. With this sample we can study gender differences in seminar attendance and the probability of taking one or both of the initial quizzes offered during the course. The second sample consists of student-quiz observations based on those students who took at least one quiz. With this sample, we can study gender differences in the probability of re-taking a quiz on the final exam.

Sample means on all students are presented in Table 1, by gender. Male and female students, on average, have the same scores on the final exam (63 points). Sample means also indicate that a significantly greater proportion of female students than male students take one or both of the initial quizzes. In addition, a greater proportion of female students than male students attend seminars. Female students are under-represented

6. Note that the decisions students make concerning exam strategies have private consequences only implying that behavioral differences are not driven by gender differences in care and concern for others.

7. There may be a slight disadvantage in re-taking as the time spent could instead be devoted to other exam questions. However, as students are allotted 5 hours for the final exam, the time constraint is not likely to be binding for the majority of students.

8. In a study of eighth-grade students Duckworth and Seeligman (2006) show that girls are more self-disciplined and that this will yield higher grades but only marginally higher scores on achievement tests.

9. The academic programs are business and administration, accounting, retailing, mathematics and economics, economics and statistics, social planning, economics and political science, and social science educators. The largest proportion of students enrolled via an academic program is in a business school program.

TABLE 1
Sample Means, Individual Observations

	Female Students	Male Students	Female–Male Difference
Birth year	1983.9 (0.13)	1984.1 (0.11)	−0.24 (0.17)
Total exam score (0–100)	62.9 (0.59)	62.6 (0.57)	0.26 (0.82)
Wrote at least one quiz	0.944 (0.007)	0.899 (0.009)	0.046*** (0.012)
Wrote two quizzes	0.836 (0.012)	0.787 (0.012)	0.048*** (0.017)
Seminar attendance	0.876 (0.010)	0.811 (0.012)	0.065*** (0.016)
Math level (1–4)			
Level 1 (lowest)	0.038 (0.006)	0.024 (0.005)	0.015 (0.007)
Level 2	0.180 (0.012)	0.119 (0.010)	0.062*** (0.015)
Level 3	0.524 (0.016)	0.427 (0.015)	0.097*** (0.022)
Level 4 (highest)	0.257 (0.014)	0.431 (0.015)	−0.174*** (0.020)
Math score (1–4)	2.48 (0.026)	2.35 (0.025)	0.126*** (0.035)
English level (1–3)			
Level 1 (lowest)	0.042 (0.006)	0.040 (0.006)	0.002 (0.009)
Level 2	0.691 (0.015)	0.744 (0.013)	−0.053*** (0.020)
Level 3 (highest)	0.267 (0.012)	0.216 (0.012)	0.051*** (0.019)
English score (1–4)	2.90 (0.025)	2.92 (0.023)	−0.027 (0.034)
No. of individuals	992	1,119	2,111

Notes: Math scores are averaged for all students regardless of math level. Standard errors in parentheses.
**Significant at 5% level; ***significant at 1% level.

TABLE 2
Sample Means, Quiz Observations

	Female Students	Male Students	Female–Male Difference
Birth year	1983.9 (0.10)	1984.2 (0.09)	−0.23 (0.13)
Total exam score (0–100)	64.5 (0.42)	65.1 (0.40)	−0.59 (0.58)
Initial quiz score	6.82 (0.046)	7.04 (0.047)	−0.22*** (0.066)
Re-take quiz	0.287 (0.011)	0.279 (0.010)	0.008 (0.015)
Math level (1–4)			
Level 1 (lowest)	0.035 (0.004)	0.024 (0.004)	0.011 (0.006)
Level 2	0.180 (0.009)	0.123 (0.008)	0.056*** (0.012)
Level 3	0.525 (0.012)	0.431 (0.011)	0.094*** (0.016)
Level 4 (highest)	0.260 (0.010)	0.422 (0.011)	−0.162*** (0.015)
Math score (1–4)	2.49 (0.019)	2.36 (0.019)	0.131*** (0.027)
English level (1–4)			
Level 1	0.041 (0.006)	0.042 (0.007)	0.001 (0.009)
Level 2	0.682 (0.016)	0.738 (0.014)	−0.056*** (0.021)
Level 3 (highest)	0.277 (0.015)	0.221 (0.013)	0.056*** (0.020)
English score (1–4)	2.90 (0.018)	2.92 (0.018)	−0.021 (0.026)
No. of quiz observations	1,766	1,888	3,654
No. of individuals	889	950	1,839

Notes: Math scores are averaged for all students regardless of math level. Standard errors in parentheses.
**Significant at 5% level; ***significant at 1% level.

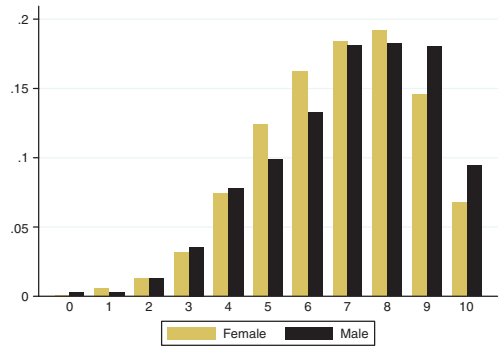
at the highest high school math levels and math scores also differ somewhat by gender.

Sample means based on student-quiz observations are shown in Table 2. Among quiz-takers, approximately 28% of both female and male quiz-takers re-take a quiz question on the exam.[10] It is difficult to interpret this as indication of no behavioral differences between males and females in re-take propensity. Rational behavior suggests that the probability of re-taking a quiz should be lower for those with higher initial quiz scores because attempting to improve one's score becomes more difficult with higher initial scores, all else equal. As shown in Table 2, female students have lower average initial quiz scores than male students.

The distribution of initial quiz scores by gender is shown in Figure 1. Female students have

10. Among students who took both initial quizzes, 25% of both male and female students re-took one quiz question and 14% chose to re-take both quiz questions.

FIGURE 1
Distribution of Initial Quiz Scores, by Gender



initial quiz scores of 5 and 6 to a relatively larger extent than male students while male students have initial quiz scores of 9 and 10 to a relatively larger extent than female students. Thus, in the empirical analysis, one should at the very least, compare gender differences in the probability of retaking the quiz conditioning on initial quiz scores.

B. *Empirical Setup*

To test for systematic gender differences in the propensity to take one or both of the initial quizzes offered during the course, we estimate the following linear probability model:

$$(1)\quad \textit{At least one quiz}_i = \beta_1 \textit{female}_i + \gamma_1 \boldsymbol{X}_i + \mu_1 \textit{female}_i \boldsymbol{X}_i + \epsilon_{1i}$$

where *At least one quiz* is a dummy variable equal to one if student i took at least one of the initial voluntary quizzes offered during the course and zero otherwise. We also use *two quizzes* as an alternative dependent variable, equal to one if student i took both of the initial quizzes and zero otherwise. The variable of interest is the *female* dummy variable which takes the value one if student i is a female and zero otherwise and β_1 measures the average differences in behavior between females and males. The variable $\boldsymbol{X}_i$ represents a full set of categorical controls for semester effects (five semesters), course code effects (ten categories), birth year effects (*1952, …,1990*), quiz question effects (two categories), high school math grades (four grades) and math levels (four levels) effects as well as high school English grades (four grades) and levels (three levels). We use a full gender interaction model captured by the interactions between the female dummy variable and $\boldsymbol{X}_i$.

An analogous regression for the probability of attaining the seminar bonus (ten points) on the exam is also estimated. The equation is specified in the same way as Equation (1) except that no control for quiz question is included in estimation.

To test for gender differences in the propensity to re-take quiz questions on the exam, the following linear probability model, based on student-quiz level data, is estimated:

$$(2)\quad \textit{Retake quiz}_{iq} = \beta_2 \textit{female}_i + \gamma_2 \boldsymbol{X}_i + \mu_2 \textit{female}_i \boldsymbol{X}_i + \epsilon_{2i}$$

where *Retake quiz* is a dummy variable equal to one if student i re-takes the corresponding quiz question, q, on the exam and zero otherwise. Over and beyond the controls described above, estimation of re-take propensities includes a full set of controls for initial quiz scores (*0–10*) and a dummy for seminar attendance.

The aim is to capture gender differences in behavior that do not stem from any other non-observable gender differences, such as underlying ability. The strategy used is inevitably based on selection on observables. We argue, however, that we control for the most crucial factors possible given the setting. By using information on math grades in high school we can be reasonably certain that our estimated gender differences in behavior do not reflect gender differences in the underlying skills relevant for the course in question. Likewise by including a measure for English grades in high school, we control for underlying differences in verbal skills which may influence testing depending on the structure of the final exam (open-ended vs. multiple-choice questions). Moreover, entrance into a given program within this economics course is competitive and course code effects should capture any systematic differences in entrance requirements.

IV. RESULTS

A. *Gender Differences in the Probability of Taking Initial Quizzes*

Results based on estimations of Equation (1) are reported in Table 3. In Column 1 the dependent variable is whether the student took at least one quiz and in Column 2 the dependent variable is whether the student took two quizzes. Coefficient estimates show that female students are associated with a 3.3 (3.4) percentage point

TABLE 3
Gender Differences in the Probability of Taking a Quiz and Attending Seminars

	The Probability of Taking at Least One Quiz	The Probability of Taking Two Quizzes	The Probability of Getting a Seminar Attendance Bonus
Female	0.033*** (0.012)	0.034** (0.017)	0.052*** (0.016)
Take quiz(zes) baseline	0.920	0.810	0.842
Percent effect	3.6	4.2	6.5
No. of observations	2,111	2,111	2,111

Notes: Linear probability models on dummy variables indicating whether or not individuals took at least one quiz, took both quizzes and attended seminars (received seminar bonus). All estimations control for semester of study (1–5), course code (1–10), birth year (1952–1990), high school math level (1–4) and math grade (1–4) as well as high school English level (1–3) and grade (1–4), all fully interacted with the gender dummy variable.
**Significant at 5% level; ***significant at 1% level.

higher probability of taking at least one quiz (both quizzes) than male students. These effects correspond to a 3.6 (4.1) percentage point higher probability for female students to take at least one quiz (both quizzes) given the baseline. Gender differences in the propensity to attend seminars are also estimated and results reported in Column 3 (Table 3). Results indicate that female students are more likely to attend seminars and therefore receive the seminar bonus on the exam than likewise male students.

B. Gender Differences in the Probability of Re-taking Quizzes on the Exam

Results from estimations on the probability to re-take quizzes on the exam, i.e., Equation (2), are reported in Table 4. Results show that female students are significantly less likely than male students to re-take quiz questions on the exam. In percentage terms, female students are almost 9% less likely to re-take a quiz on the final exam than male students, all else equal.

Average differences can mask variation in re-take propensities across the skills distribution. There are a number of studies that suggest that women who choose to compete, especially in male-dominated environments, are as competitive as the men in these environments.[11] Figure A1 in the Appendix plots the coefficient estimates of the interaction between the female dummy variable and each initial quiz score, i.e., gender differences in re-take propensities across the distribution of initial quiz scores.[12] Results suggest that gender differences in re-take propensities are driven by those with an initial quiz score of five. Female students with this initial quiz score (five) are significantly less likely to re-take the quiz on the exam than male students with the same initial quiz score. Re-take propensities are similar for female and male students with higher initial quiz scores confirming results in previous studies that suggest smaller gender differences in behavior at the high end of the skills distribution.

TABLE 4
Gender Differences in the Probability of Re-taking Quizzes on the Final Exam (Quiz Observations)

	Re-take Propensity
Female	−0.025** (0.013)
Re-take (quiz) baseline	0.283
Percent effect	−8.8
No. of observations	3,654

Notes: Linear probability models on dummy variables indicating whether or not individuals re-took at least one quiz. Controls for quiz question (1–2), seminar attendance, semester (1–5), course code (1–10), birth year (1952–1990), high school math level (1–4) and math grade (1–4) as well as high school English level (1–3) and grade (1–4), quiz question and initial quiz score are included in estimation. All controls are fully interacted with the gender dummy variable. Robust standard errors in parentheses.
**Significant at 5% level.

C. Potential Costs of Gender Differences in Exam Behavior

Before discussing potential explanations behind observed gender differences in examination behavior, let us first look at how these

11. See for example: Atkinson, Baird, and Frye (2003), Birley (1989), Garratt, Weinberger, and Johnson (2013), Datta Gupta, Poulsen, and Villeval (2005), Dwyer, Gilkeson, and List (2002), Johnson and Powell (1994), Master and Meier (1988), and Nekby, Vahtrik, and Skogman Thoursie (2008).

12. Initial quiz scores of zero to three are grouped together due to a low number of observations in these cells.

TABLE 5
Correlation between Quiz-Taking and Final Exam Scores

	(1)	(2)	(3)	(4)
At least one quiz	12.2*** (1.59)	12.3*** (1.70)	—	—
Two quizzes	—	—	10.7*** (1.00)	10.5*** (1.01)
Female	−1.53** (0.73)	−2.20 (3.24)	−1.49** (0.72)	0.75 (1.83)
Female* at least one quiz	—	0.73 (3.30)	—	—
Female* two quizzes	—	—	—	−2.74 (1.96)
Final exam baseline		62.8		
Percent effect	19.4	19.3	17.4	16.7
No. of observations		2,111		

Notes: Ordinary least squares estimation of exam scores (0–100) on quiz-taking, defined as dummy variables equal to one if an individual has taken at least one quiz, alternatively both quizzes, and zero otherwise. Control for quiz question (1–2), seminar attendance, semester (1–5), course code (1–10), birth year (1952–1990), math level (1–4) and math grade (1–4) including full interaction between math level and math grade. Robust standard errors in parentheses.
**Significant at 5% level; ***significant at 1% level.

behavioral differences (quiz-taking, seminar attendance, and quiz re-taking) correlate with other examination outcomes. The difference between quiz-takers and non-quiz-takers in final exam scores is shown in Table 5. Quiz-taking is associated with higher final exam scores. Estimation controlling for all relevant covariates yields results showing that taking at least one quiz (or two quizzes) is associated with approximately 12 higher points on the final exam in comparison to not taking any quizzes (this corresponds to 17%–20% higher final exam scores for quiz-takers).

Thus, quiz-takers do better on final exams, either due to a systematically different type of study behavior which voluntary quizzes give rise to (including the possibility of re-taking quiz questions on the final exam) or because quiz-takers are selected among the more ambitious and scholarly students enrolled in the course. As we control for differences in relevant math and verbal skills, more weight should perhaps be given to the first explanation. Female quiz-takers who take at least one initial quiz do not appear to differ from likewise male quiz-takers in terms of the quiz premium on final exams as indicated by the insignificant coefficient on the gender-quiz interaction terms (Column 2). The quiz premium for those who take two initial quizzes appears to differ somewhat between female and male students (Column 4) but this difference is not significant at conventional levels.

Turning instead to an examination of the correlation between re-taking quizzes on the exam and final exam scores (based on quiz observations), we see from results reported in Table 6 that quiz re-takers are associated with significantly higher final quiz scores in comparison with those that abstain from re-taking the quiz. Quiz re-takers are associated with, on average, 1.3 higher points than those that do not re-take the quiz on the final exam. No gender differences in quiz improvement are noted as shown by the insignificant estimated coefficient on the interaction between the female dummy variable and the re-take dummy variable.

Re-taking quizzes on the exam implies a time investment. Students must weigh in to what degree they will benefit from re-taking quiz

TABLE 6
Correlation between Quiz Re-Taking and Final Exam Scores (Quiz Observations)

	Quiz Score:	
Quiz re-take	1.26*** (0.063)	1.26*** (0.083)
Female	−0.01 (0.033)	−0.03** (0.014)
Quiz re-take* female	—	0.07 (0.11)
	Other exam question scores:	
Quiz re-take	6.44*** (0.659)	6.47*** (0.658)
Female	0.39 (0.541)	0.80 (0.620)
Quiz re-take* female	—	−1.42 (1.010)
No. of observations	3,654	3,654

Notes: Ordinary least squares estimation of quiz scores (0–10) and exam scores (0–100) on quiz re-taking, defined as dummy variable equal to one if an individual that has taken at least one of the initial voluntary quizzes and re-taken it on the final exam. All estimation controls for quiz question (1–2), seminar attendance, semester of study (1–5), course code (1–10), birth year (1952–1990), math level (1–4) and math grade (1–4) as well as a full interaction between math level and math grade. Robust standard errors reported in parentheses.
**Significant at 5% level; ***significant at 1% level.

questions given that less time can be invested in the remaining mandatory questions on the exam. Students are, however, allotted a maximum 5 hours to complete final exams implying that the time constraint may not be binding for the majority of students. Abstaining from re-taking quiz questions on the exam may, nonetheless, be a rational strategy for students who feel that their time is better invested in answering other final exam questions or who may prefer open-ended questions to multiple-choice (quiz) questions. If such is the case, we would expect that non-re-takers have higher scores on the other exam questions than re-takers, especially for female students. Results in the second panel of Table 6, however, show that re-takers are associated with higher points (a 6–7 point advantage) on the remaining exam questions and that there is no gender difference in this association. As such, these correlations suggest that it is not time constraints driving female quiz-takers to abstain from re-taking quizzes on the exam nor a rational preference for open-ended questions, rather there appears to be spill-over effects on other exam questions implying that the selection of students who prepare for the possibility of re-taking quizzes on the exam improve their scores over and beyond the improvement on the quiz questions alone. Again no gender differences in this correlation are found.

We emphasize that these regressions only reflect associations. Even if we control for a large set of relevant covariates that are available to us (see Section B), re-taking the quiz on the exam is still potentially endogenous in the final grades equation due to selection on unobservables. Note that the same set of control variables is used in the selection equation (i.e., the re-take probability equation given by Equation (2)) and in the final grades equation. As such, we rely on a regression-control strategy. To make a causal interpretation of the effect of re-taking on final grades we need an exclusion restriction, i.e., a variable that affects the re-take probability but not final grades. Such a variable is not available to us.

As a robustness check we estimate the selection equation using a Probit model and use the non-linearity as an exclusion restriction. The result from such a strategy reveals that re-taking the quiz is still associated with significantly higher final quiz scores but that the significant correlation between re-taking and scores on the remaining questions disappears.[13] That the significant correlation between re-taking the quiz and scores on the remaining questions disappears suggests that there is positive selection on unobservable characteristics among re-takers. In other words, quiz re-takers seem to have unobserved characteristics that yield higher points on the remaining final exam questions. Therefore, if we rely on the results from the selection model, the benefits in terms of improving final grades from re-taking the quiz are alleviated compared to the results presented in Table 6. We can still conclude, however, that quiz re-taking is associated with higher final grades through higher grades on final quiz scores.

Taken together, results show that quiz re-takers are associated with higher final grades on the exam, at least with respect to final quiz scores.

V. POTENTIAL EXPLANATIONS FOR GENDER DIFFERENCES IN EXAMINATION BEHAVIOR

There are a number of potential mechanisms behind our observed behavioral differences in examination strategies having to do with gender differences in confidence, risk-aversion, competitiveness, and/or procrastination, which may be difficult to tease out in the type of real world setting examined here. Below we discuss the credibility of some of these mechanisms in explaining observed results.

Seminar attendance and quiz-taking are two ways of gaining safe points on the final exam as well as more time to devote to other exam questions. Quiz-takers also have a second chance to improve their quiz scores. As such, seminar attendance and quiz-taking must be considered risk-averse strategies in examination behavior. Our results showing higher propensities among female students to take initial quizzes and attain a seminar bonus are therefore consistent with women being, on average, more risk avert than men.

Re-taking a quiz on the final exam is associated with a risk of lowering initial quiz scores. Our result that female students are less inclined to re-take the quizzes, all else equal, is therefore also consistent with an interpretation that female students are more risk averse than male students. Re-take propensities are estimated on a sample of students that took initial quizzes. If quiz-taking is associated with a higher degree of risk aversion,

13. Results are available from the authors upon request.

then we underestimate gender differences in risk aversion since there must be a stronger selection of risk avert males in the sample.[14]

An alternative interpretation of results is that male students are more (over)confident than female students. Overconfidence is a relative concept. Men could be confident (and correct) in their interpretation of their abilities whereas women are under-confident, or men could be overconfident (i.e., have an incorrect estimation of their ability) whereas women have a correct estimation of their ability. Empirical evidence for male overconfidence can be found in, for example, Barber and Odean (2001), who investigate gender differences in overconfidence among financial investors. They find that male financial traders trade more and reduce portfolio returns more than female traders.[15] Our results may reflect male perceptions (correct or not) of a higher capacity to receive high final exam scores without initial quiz scores or the seminar bonus. Higher confidence, on average, can also explain why male students re-take quizzes to a large extent, all else equal.

One way to explore if gender differences in examination behavior are due to male overconfidence is to compare initial quiz scores with the quiz scores on the final exam for those that re-take quizzes on the exam. If men are overconfident then there should be a greater spread of male scores due to re-taking as overconfident male students are more likely to win and lose than female students. If not, than female risk aversion is a more likely explanation for differences in re-take propensities.

The distribution of the difference in final and initial quiz scores is shown in Figure A2 in the Appendix. In this figure, we depart from an initial quiz score of five as results showing gender differences in re-take probabilities is driven by students with this initial quiz score. Results show that male students are more likely to win from re-taking than female students.[16] This suggests that male overconfidence is not part of the story behind gender differences in re-take behavior.

Another potential explanation for behavioral differences concerns gender differences in time-use.[17] Male students may be more prone to procrastinate, due to hyperbolic discounting or lack of self-discipline, implying that they study less at the beginning of the course and more intensely close to the date of the final exam. Gender differences in procrastination may then explain why male students take initial quizzes to a lesser extent and have lower seminar attendance. Higher male procrastination is also consistent with higher re-take propensities if male students study more intensively close to the final exam date and therefore have more "fresh" knowledge. Male procrastination is also supported by the fact that male students register for the final exam to a lower extent than female students (registration is open approximately 2 weeks before the final exam).

Taken together, the observed gender differences in examination behavior found in this study seem consistent with conclusions from the experimental literature that women are more risk averse, on average, than men. We cannot however rule out procrastination as a potential explanation for observed results.

VI. DISCUSSION AND CONCLUSION

This study has used information on examination behavior among first-year microeconomics students at Stockholm University to analyze gender differences in examination behavior. Students in this course have the possibility of attaining a seminar bonus on the final exam and are given two voluntary quizzes during the course. At the final exam, the scores received on initial quizzes can either be accepted, or students can attempt to improve their marks by answering similar quiz questions. Results suggest that female students are more likely to take initial quizzes and receive

14. A lower propensity to re-take quiz questions on the exam may also be explained by a female preference for open-ended questions. Female students may prefer to invest more time in the other exam questions than in the multiple-choice quiz questions. Given a higher preference for open-ended questions among female students, taking one or both of the initial quizzes during the term is still a risk-averse strategy as it allows female students a larger chance to act on this preference during the final exam.

15. Other studies on male overconfidence include Beyer (1990), Beyer and Bowden (1997), Lundeberg, Fox, and Puncchohar (1994), Niederle and Vesterlund (2007), and Soll and Klayman (2004).

16. This also means that the average difference in initial and final quiz scores is higher for male students given that we conditioned on an initial quiz score of five. Note that there are no gender differences in the correlation between re-taking a quiz and final quiz score if we do not condition on a given initial quiz score (see Table 6).

17. Previous research has found that there are gender differences in procrastination behavior and that male and female students procrastinate for different reasons (see, e.g., Uzun Özer, Demir, and Ferrari 2009 and Warner and Pleeter 2001).

a seminar bonus but are less likely to re-take quiz-questions on the final exam. We argue that results are due to higher levels of risk aversion among female students and/or a greater degree of procrastination among male students.

Despite convincing laboratory experimental evidence of gender differences in behavior, there is a lack of research in real settings. The few real world studies available focus primarily on competitive environments. This study adds to this literature by analyzing gender differences in behavior in a real setting that involves risk and judgment about own ability but is otherwise non-competitive in the sense that all grading is based on absolute and not relative performance.

Our results provide conclusions on how gender differences in preferences or personal traits have implications for real world outcomes. First, our results tentatively show that gender differences in behavior affect gender differences in educational outcomes. Attending seminars, taking initial quizzes and re-taking quizzes on the final exam are all correlated with higher final exam scores. On the other hand, female students could improve their results on the final exam by behaving more like men when it comes to re-taking quizzes on the final exam.

Second, we find behavioral differences among young economics students already before they enter the labor market. Female students that shy away from risk taking may then, for example, be less inclined to accept positions with lower job security and/or may delay negotiating for higher wages or seeking promotions until they are more secure about the value of their merits than likewise men. Thus, our results suggest that gender differences in behavior may explain some of the gender differences observed in the labor market.

Finally, in line with Manning and Saidi (2010) we find gender differences in behavior that are considerably smaller than those generally found by laboratory experiments. This suggests that previous research using laboratory experiments has overstated the importance of gender differences in behavior when explaining gender differences in the labor market.

APPENDIX

FIGURE A1

Female/Initial Quiz Score Interaction in the Propensity to Re-Take Quizzes on the Final Exam (Quiz Observations)

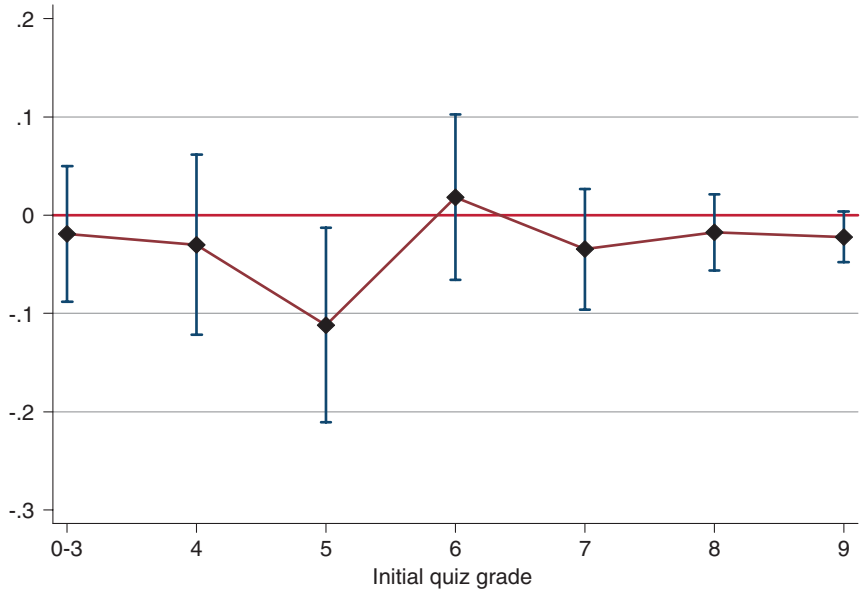
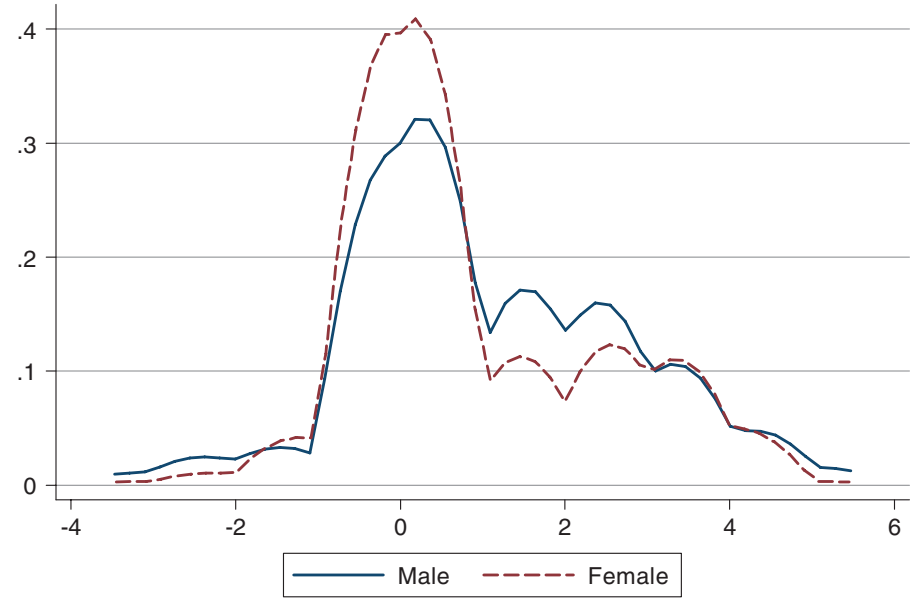


FIGURE A2

Distribution of Difference in Initial and Final Quiz Score (Quiz Re-Takers)



Note: The above distribution is calculated for an initial quiz score of five.

REFERENCES

Almlund, M., A. L. Duckworth, J. Heckman, and T. D. Kautz. "Personality Psychology and Economics." NBER Working Paper Series No. 16822, 2011.

Altonji, J. G., and R. Blank "Race and Gender in the Labor Market," in *Handbook of Labor Economics*, Vol. 3C, edited by O. Ashenfelter and D. Card. Amsterdam: Elsevier, 1999, 3144–259.

Antonovics, K., P. Arcidiacono, and R. Walsh. "Games and Discrimination: Lessons from the Weakest Link." *Journal of Human Resources*, 40(4), 2005, 918–47.

———. "The Effects of Gender Interactions in the Lab and the Field." *Review of Economics and Statistics*, 91(1), 2009, 152–62.

Atkinson, S., S. Baird, and M. Frye. "Do Female Fund Managers Manage Differently?" *Journal of Financial Research*, 26, 2003, 1–18.

Barber, B. M., and T. Odean. "Boys Will Be Boys: Gender, Overconfidence, and Common Stock Investment." *Quarterly Journal of Economics*, 116, 2001, 261–92.

Bertrand, M. "New Perspectives on Gender, " in *Handbook of Labor Economics*, Vol. 4B, edited by O. Ashenfelter and D. Card. Amsterdam: Elsevier, 2011, 1545–92.

Beyer, S. "Gender Differences in the Accuracy of Self-Evaluations of Performance." *Journal of Personality and Social Psychology*, 59, 1990, 960–70.

Beyer, S., and E. M. Bowden. "Gender Differences in Self-Perceptions: Convergent Evidence from Three Measures of Accuracy and Bias." *Personality and Social Psychology Bulletin*, 23, 1997, 157–72.

Birley, S. "Female Entrepreneurs: Are They Really Any Different?" *Journal of Small Business Management*, 27, 1989, 32–37.

Blau, F. D. *Gender, Inequality, and Wages*. Oxford: Oxford University Press, 2012.

Blau, F., and L. Kahn. "Changes in the Labor Supply Behavior of Married Women: 1980–2000." *Journal of Labor Economics*, 25(3), 2007, 393–438.

Cornwell, D., D. Mustard, and J. Van Parys. "Noncognitive Skills and the Gender Disparities in Test Scores and Teacher Assessments: Evidence from Primary School." *Journal of Human Resources*, 48(1), 2013, 239–66.

Croson, R., and U. Gneezy. "Gender Differences in Preferences." *Journal of Economic Literature*, 47(2), 2009, 448–74.

Datta Gupta, N., A. Poulsen, and M.-C. Villeval. "Male and Female Competitive Behavior: Experimental Evidence." IZA Working Paper No. 1833, Institute for the Study of Labor, 2005.

Dohmen, T., and A. Falk. "Performance Pay and Multi-Dimensional Sorting: Productivity, Preferences and Gender." *American Economic Review*, 101(2), 2011, 556–90.

Duckworth, A. L., and M. L. P. Seeligman. "Self-Discipline Gives Girls the Edge: Gender in Self-Discipline, Grades, and Achievement Test Scores." *Journal of Educational Psychology*, 98(1), 2006, 198–208.

Dwyer, P. D., J. H. Gilkeson, and J. A. List. "Gender Differences in Revealed Risk Taking: Evidence from Mutual Fund Investors." *Economics Letters*, 79, 2002, 151–58.

Flory, J., A. Leibbrandt, and J. A. List. "Do Competitive Work Places Deter Female Workers? A Large-Scale Natural Field Experiment on Gender Differences in Job-Entry Decisions." NBER Working Paper Series w16546, 2010.

Garratt, R. J., C. Weinberger, and N. Johnson. "The State Street Mile: Age and Gender Differences in Competition Aversion in the Field." *Economic Inquiry*, 51, 2013, 806–15.

Goldin, C., L. Katz, and I. Kuziemko. "The Homecoming of American College Women: The Reversal of the College

Gender Gap." *Journal of Economic Perspectives*, 20(4), 2006, 133–56.

Heckman, J., J. Stixrud, and S. Urzua. "The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior." *Journal of Labor Economics*, 24(3), 2006, 411–65.

Heckman, J., R. Pinto, and P. Savelyev. "Understanding the Mechanisms through Which an Influential Early Childhood Program Boosted Adult Outcomes." *American Economic Review*, 103(6), 2013, 2052–86.

Jacob, B. "Where the Boys Aren't: Non-Cognitive Skills, Returns to School and the Gender Gap in Higher Education." *Economics of Education Review*, 21, 2002, 589–98.

Johnson, J. E. V., and P. Powell. "Decision Making, Risk and Gender: Are Managers Different?" *British Journal of Management*, 5(2), 1994, 123–38.

Jurajda, S., and D. Münich. "Gender Gap in Performance under Competitive Pressure: Admissions to Czech Universities." *American Economic Review*, 101(3), 2011, 514–18.

Lavy, V. "Gender Differences in Market Competitiveness in a Real Workplace: Evidence from Performance-Based Pay Tournaments among Teachers." *Economic Journal*, 123(569), 2013, 540–73.

Lazear, E., U. Malmendier, and R. Weber. "Sorting in Experiments." Unpublished Manuscript, Stanford University, 2004.

Lundeberg, M. A., P. W. Fox, and J. Puncchohar. "Highly Confident but Wrong: Gender Differences and Similarities in Confidence Judgments." *Journal of Educational Psychology*, 86, 1994, 114–21.

Manning, A., and F. Saidi. "Understanding the Gender Pay Gap: What's Competition Got to Do with It?" *Industrial and Labor Relations Review*, 63(4), 2010, 681–98.

Master, R., and R. Meier. "Sex Differences and Risk Taking Propensity of Entrepreneurs." *Journal of Small Business Management*, 26, 1988, 31–35.

Nekby, L., L. Vahtrik, and P. Skogman Thoursie. "Gender and Self-Selection into a Competitive Environment: Are Women More Overconfident than Men?" *Economics Letters*, (100), 2008, 405–7.

Niederle, M., and L. Vesterlund. "Do Women Shy Away from Competition? Do Men Compete Too Much?" *Quarterly Journal of Economics*, 122(3), 2007, 1067–101.

Örs, E., F. Palomino, and E. Peyrache. "Performance Gender Gap: Does Competition Matter Performance Gender Gap: Does Competition Matter?" *Journal of Labor Economics*, 31(3), 2013, 443–99.

Soll, J., and J. Klayman. "Overconfidence in Interval Estimates." *Journal of Experimental Psychology: Learning, Memory, and Cognition*, (30), 2004, 299–314.

Uzun Özer, B., A. Demir, and J. Ferrari. "Exploring Academic Procrastination among Turkish Students: Possible Gender Differences in Prevalence and Reasons." *Journal of Social Psychology*, 149(2), 2009, 241–57.

Warner, J., and S. Pleeter. "The Personal Discount Rate: Evidence from Military Downsizing Programs." *American Economic Review*, 91(1), 2001, 33–53.