Gender differences in multiple-choice questions and the risk of losing points

Krista Riukula*
March 14, 2023

Abstract

I study the gender differences in performance in multiple-choice questions (MCQ) in a setting where wrong answers are penalized and the objective is to score as many points as possible. I exploit data from an undergraduate level microeconomics course at a Finnish university across a six-year period of 2010 and 2012–2016. The course consists of two equally weighted exams that include both multiple-choice and open-ended questions. The results show that, when controlling for the performance in the first midterm exam, women omit more MCQ items in the second exam than men, which, in turn, translates to fewer points. Women do not do worse in the open-ended questions that are similar to the MCQ's, and neither is the probability of women answering incorrectly to the MCQ's higher. Hence, gender differences in test results might reflect differences in behavior in a very particular test setting rather than genuine differences in skills.

Keywords: *Gender difference, Exam behavior, Risk preference*
JEL classifications: *I20, J16, D81*

*Etla Economic Research, Arkadiankatu 23 B, 00100 Helsinki, Finland. Email: krista.riukula@etla.fi. I thank Kristiina Huttunen, Aino Kalmbach, Antti Kauhanen, Tuomas Pekkarinen, and Marko Terviö for useful comments.

1 Introduction

There has been a great interest in gender differences in preferences (see Croson and Gneezy (2009) and Niederle (2017b) for excellent reviews). One area of interest has been whether the genders differ in exam behavior. Previous literature has shown that some exam types favor one gender over the other, mostly that females do better in open-ended questions and males in multiple-choice questions (MCQ's) (see, e.g., Ferber et al., 1983; Lumsden and Scott, 1987; Walstad and Robson, 1997). Empirical work has shown that females, for example, skip more MCQ's, and this difference might be due to females being more unwilling to guess than males. Women and men differ in the degree of risk aversion (Croson and Gneezy, 2009; Byrnes et al., 1999). Women consider risk to be more of a threat, and may feel fear of loss, which leads to overweighting the probability of a loss, while men are in general more confident of winning (Croson and Gneezy, 2009). Hence, the differences in test performances might become even more prominent when penalties for wrong answers are implemented. It is a question of great importance as many exams are in the form of MCQ's due to, for example, efficiency in grading. If gender differences in risk preferences affect test-taking strategies, the effects might also spill over affecting educational attainment.

Gender differences in risk-taking have received a great amount of attention in the literature throughout the last decades in both psychology and economics. It has been studied in different contexts ranging from driving behavior and smoking, to gambling and choice dilemmas. Croson and Gneezy (2009) find that the published experimental work in gender differences in preferences are broadly consistent with women being more risk averse than men. Greater risk aversion is likely to have an impact on major decisions such as investment portfolio choice and choice of occupation, and this can even affect individual economic well-being (Shurchkov and Eckel, 2018). According to Shurchkov and Eckel (2018), the majority of the evidence suggests that women are more risk averse than men, although the magnitude of this difference depends on context and framing. For example, Wieland et al. (2014) find that the gender difference in risk aversion is reduced or eliminated as the context changes from tasks framed as gambles to other domains.

Risk aversion is not the only risk attitude where women and men differ (see, e.g., Shurchkov and Eckel, 2018). Women are also more loss averse and there are differences in overconfidence, optimism, perception of subjective probabilities, and ambiguity attitudes. Niederle (2017a) shows that women are less likely to enter competitions than men and that this difference in competitiveness can also help account for gender differences in education choices, such as math and science choices in school. Niederle and Vesterlund (2007) also find that women are less likely to enter tournaments and that the gap is driven by men being more overconfident and by gender differences in preferences for performing in a competition. The result is that women shy away from competition while men embrace it. These differences in risk preferences might affect test-taking strategies and hence, performance in tests.

In this paper, I study the gender differences in performance in MCQ's using data from a second-year undergraduate economics course at a Finnish university across a six-year period of 2010 and 2012–2016. The course consists of two equally weighted exams that include both MCQ's and open-ended questions. Wrong answers to MCQ's are penalized with negative points, while omission yields zero points. Hence, students that are, for example, more risk-averse might be more inclined to omit an MCQ item if they do not know the correct answer. The performance in the first exam provides a natural setting for controlling for students' ability in the subject. The open-ended questions are fairly similar to the MCQ's which also provides a natural setting for controlling for students' ability in the subject. They consist of similar calculations as the MCQ's but are lengthier and more time-consuming. However, controlling for ability with points from the open-ended questions in the same exam neglects the fact that women and men might have differing time preferences for the two types of questions. Women might simply invest more time towards the open-ended questions on the cost of MCQ's. Hence, I control for the performance in the first midterm exam. The results show that, when controlling for the performance in the first exam, women omit more MCQ items than men, which contributes to roughly one third of the difference in total points. Women do equally well compared to men in the rest of the exam, including in questions similar to the MCQ's but lengthier, suggesting that there are no gender differences in the ability in the subject. Thus, gender differences might actually reflect differences in behavior in a very particular test setting rather than genuine differences in skills. The reason why women tend to omit more items might stem from the gender differences in risk preferences: women consider risk to be

more of a threat, and may feel fear of loss that, in turn, leads to over-weighting the probability of a loss, while men are, for example, more confident of winning.

This study complements the studies of gender differences in exam performance (e.g. Ferber et al., 1983; Baldiga, 2013; Pekkarinen, 2014; Iriberri and Rey-Biel, 2021; Saygin and Atwater, 2021) that have shown that men do better in MCQ's and omit less items. One of the first studies on gender differences in exam behavior was by Ferber et al. (1983). They show that men do better in both MCQ's and essay questions but the difference is larger in the former. In competitive settings women have been proved to omit more items in MCQ's (see, e.g., Pekkarinen, 2014; Saygin and Atwater, 2021). Similarly, Baldiga (2013) shows that when there is a penalty for wrong answers, women answer significantly fewer questions than men. Iriberri and Rey-Biel (2021) show using within-participant regression analysis that female participants leave significantly more omitted questions than males when there is a reward for omitted questions. Saygin and Atwater's (2021) findings suggest that the magnitude and the sign of the gender gap in answering questions under uncertainty is context dependent. They study the Turkish college entrance exams and find that the gender gap is larger in math and when questions are more difficult while it reverses in literature.

Most of the studies have controlled for ability by performance in the class or by performance in schooling in general. Pekkarinen (2014) controls for the matriculation exam points, while Ferber et al. (1983) control for SCAT points and GPA. Baldiga (2013) controls for students' confidence and knowledge of the material by requiring students to provide answers for one part and indicate the probability of their answer being correct and uses exogenous variation in size of penalty for incorrect answers in MCQ's. Some of these controls might hide the underlying unobserved ability as the measures are from preceding times and from different fields. Women might simply do worse in these studies because their ability in that specific field is inferior. Most of these studies have been done in the field of economics and business, where men do, in general, perform better. Some of the exams studied have also been of high stakes (for example Pekkarinen (2014) and Saygin and Atwater (2021) look at entrance exams). In this study, I have the privilege of being able to control with the performance in the first midterm exam and with similar open-ended questions. Moreover, I focus on an exam that is arguably of less importance than entrance exams.

Publication bias might also affect; work with significant findings are being published with a higher probability. Hence, work finding no gender gaps might simply not be published. The replicability of some scientific findings has also recently been called into question (see, e.g. Camerer et al., 2016). One answer to this has been the use of pre-analysis plans.[1] Pre-registered pre-analysis plans (PAPs) may reduce publication bias if it implies that null results are more likely to be reported. However, Brodeur et al. (2022) find little or no evidence to suggest that pre-registration in itself reduces p-hacking or publication bias. However, they find that pre-registered studies that have a complete PAP are significantly less p-hacked.[2]

2 Empirical setup

2.1 Data

I analyze a set of 12 midterm exams in an undergraduate level microeconomics course at a Finnish university across a six-year period of 2010 and 2012–2016. Most of the students attending the course are doing their undergraduate studies in business. The course consists of two midterm exams that account together for 80% of the total course grade. Exercise sets account for the remaining 20%. Each midterm exam has 12 to 18 MCQ's followed by typically three to four open-ended questions. The MCQ's account for 36% to 45% of the total exam points and are hence a significant part of the total grade. The open-ended questions always include a question (Question 1) in which students are asked to define different terms from the field of microeconomics. This question is usually worth 8% to 20% of the total exam points. The remainder

[1] *Drazen et al. (2021) propose a journal-based replication policy to tackle with the reliability crisis of experimental research.*

[2] *This study did not have a pre-registered PAP.*

of the questions (Questions 2 to 4) include calculations similar to those in the MCQ's but lengthier and more time consuming. These questions account for the rest of the points, 44% to 56%, depending on the exam. The exam has the MCQ block first, followed by the open-ended questions.

Each of the MCQ's has five possible options, only one of them being correct.[3] Each wrong answer is penalized by one point, while a correct answer rewards two to three points depending on the exam. Hence, the expected score of randomly choosing an answer is negative and students who have no idea about the correct answer are better off by omitting the item. However, if one can draw out two wrong answers, the student is better off guessing in expected terms.

Table 1: *Midterm exam points*

	Females		Males		Diff.
	Mean	**Std. Dev.**	**Mean**	**Std. Dev.**	***p*-value**
Panel A. First midterm exam					
Total points	64.91	15.96	69.93	15.10	0.00***
Omitted items	1.84	1.62	1.20	1.47	0.00***
Share of correct answered items	0.78	0.15	0.78	0.15	0.68
Points from MCQ	23.71	8.14	26.14	7.74	0.00***
Points from Question 1	8.75	3.80	9.20	4.12	0.28
Points from Question 2	15.49	5.90	16.24	5.53	0.20
Points from Question 3	12.53	6.92	14.04	6.85	0.04**
Points from Question 4	12.06	6.27	12.95	7.47	0.47
Panel B. Second midterm exam					
Total points	62.87	22.54	67.54	20.49	0.04**
Omitted items	2.03	1.68	1.38	1.67	0.00***
Share of correct answered items	0.79	0.18	0.81	0.14	0.18
Points from MCQ	19.50	9.21	21.03	7.56	0.07*
Points from Question 1	10.29	5.85	11.41	7.01	0.11
Points from Question 2	8.91	7.58	10.81	8.58	0.03**
Points from Question 3	12.78	7.64	12.78	8.11	1.00
Points from Question 4	12.76	8.63	14.40	8.08	0.13
Observations	139		262		401

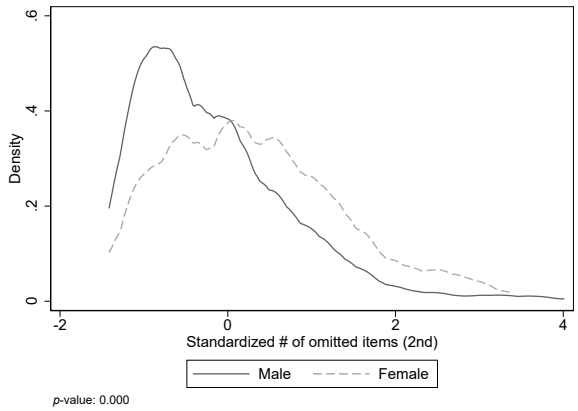
*Notes: The table provides descriptive statistics on the exam performance for the 2010–2016 (excluding year 2011) Microeconomics course's first and second midterm exams by gender and the p-values for their differences. The maximum total points is 100 and the exam consists of MCQ's and three to four open-ended questions (Questions 1-4). Omitted items refers to unanswered MCQ items and the share of correct answered items refers to the share of correctly answered MCQ items. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.*

[3] *In a few occasions, there was a mistake in the exam. If there was a mistake in the question, everyone was rewarded points.*

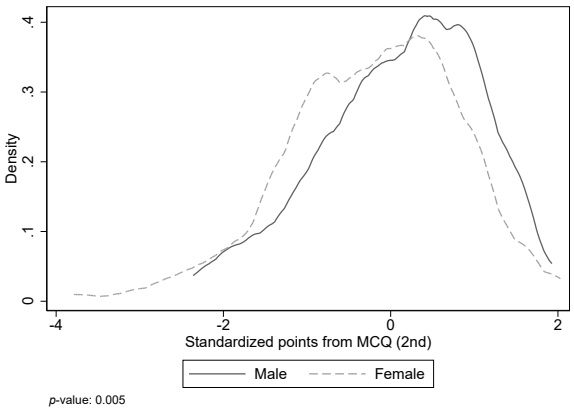
Table 1 provides the descriptive statistics for the midterm exams by gender. Male students omit on average 1.39 items in the second midterm exam, while females omit roughly 0.6 items more. Females also score, on average, 1.5 fewer points from the MCQ's, which might be at least partly explained by the fact that they omit more items than males. Out of the questions answered, both females and males get on average the same portion correct (0.79 and 0.81, respectively). Interestingly, the differences between points from the open-ended questions are, on average, not statistically different suggesting that females and males do equally well in them. This, in turn, indicates that females attending the course have the same ability in the subject as their male co-students. Females score 4.7 fewer points from the entire exam. Roughly one third (32%) of this difference derives from the difference in the MCQ's. The points in Table 1 are, however, merely descriptive as the maximum points and the ratio between female and male students differ by both exam and year. In the analysis, I standardize the number of points received and the number of omitted questions by exam since the maximum points awarded for every question and the number of MCQ's differ by exam. A standardized variable is a variable that has been re-scaled to have a mean of zero and a standard deviation of one.

Figure 1 provides graphical evidence on how the distributions of standardized points and number of omitted items differ by gender. Females seem to have a longer left tail in the distribution of MCQ points as shown in panel b), while the distributions between males and females look rather similar for the rest of the questions (panels c) to e)). Questions 2 and 3 are of similar type (usually calculations) than most of the MCQ's, but lengthier. Females do equally well, if not better in them than males, indicating that their ability in the subject does not differ from their male co-students. Moreover, there is a large spike for males omitting a very small amount of items suggesting that they omit less items than females. The exact *p*-values for the Kolmogorov-Smirnov tests of differences between male and female distributions are reported in the bottom-left corner of each respective sub-figure and show that only the differences between omitted items and points from the MCQ's are significant at the 1% significance level. The *p*-value for total points is significant at the 10% significance level and insignificant for all of the open-ended questions. Figure 1 provides evidence suggesting that males and females might be behaving differently in terms of omitting items when wrong answers are being penalized.

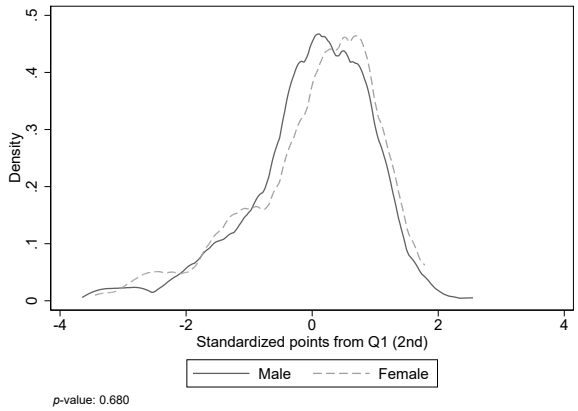
Figure 1: *Distribution of the standardized number of omitted items and points from the second midterm exam by gender for the years 2010–2016 (excluding year 2011)*



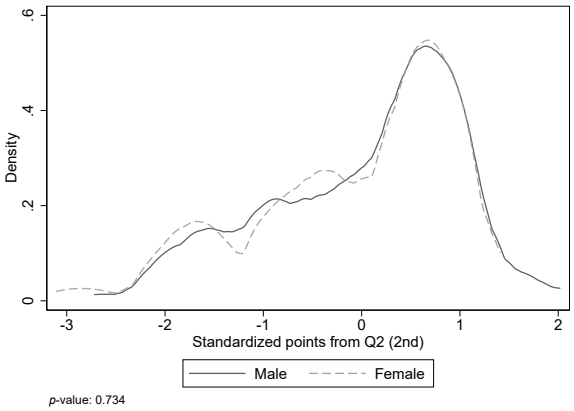
(a) Omitted items



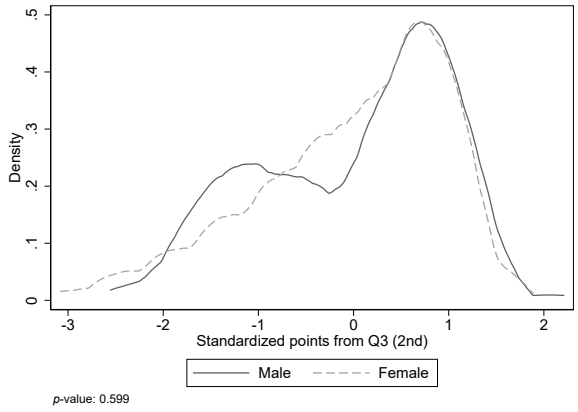
(b) Points from MCQ



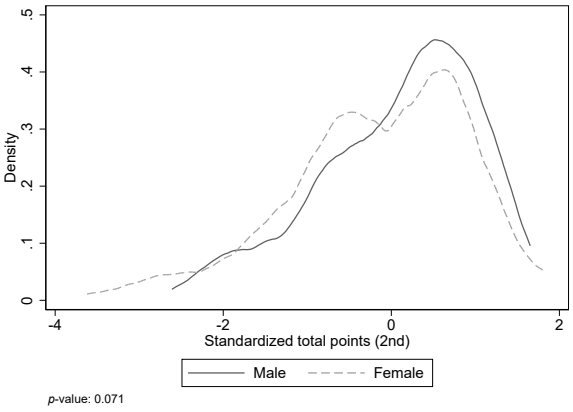
(c) Points from Question 1



(d) Points from Question 2



(e) Points from Question 3



(f) Total Points

Notes: The exact p-values for the Kolmogorov-Smirnov tests of differences between male and female distributions are reported on the bottom-left corner of each corresponding sub-figure. The number of points received and the number of omitted questions are standardized by exam. The exam consists of MCQ's and three to four open-ended questions (Questions 1–4). Omitted items refers to unanswered MCQ items.

2.2 Empirical specification

I study gender differences in exam performance, the number of omitted questions, and points from the MCQ's, with the following regression:

$$Y_i = \alpha_0 + \alpha_1 D_i + \alpha_t + \delta X_i + \epsilon_i, \quad (1)$$

where Y_i is the outcome (exam performance / number of omitted questions / points for MCQ's; all standardized) in the second midterm exam, D_i is the female dummy, α_t the year indicators, X' is a vector of control variables, i.e., points from the first midterm exam, and ϵ_i is the error term. Standard errors are heteroskedasticity robust.

3 Results

To disentangle if students' behavior in answering MCQ's is gender dependent, I have regressed the number of omitted questions against a set of explanatory variables. The results presented in Table 2 suggest that females omit more items even when I control for their ability or performance in the first midterm exam. Females score 0.19 standard deviations fewer points (≈ 1.7 points) from the MCQ's than men as shown in column (1) in panel a) when controlled for the performance in the first midterm exam. This might be partly explained by column (2) where we can see that females omit 0.35 standard deviations more items (≈ 0.59 items) than male students when controlled for the performance in the previous exam. The results remain similar, albeit larger in magnitude, when controlling only for the points from the open-ended questions (questions 1–4) from the first midterm exam as shown in panel b) in Table 2. Females score 0.25 standard deviations fewer points (≈ 2.3 points) from the MCQ's and omit 0.43 standard deviations more items (≈ 0.72 items) than male students when controlled only for the performance in the open-ended questions in the previous exam. The results are in line with previous studies showing that females omit more MCQ items when wrong answers are penalized (see, e.g., Pekkarinen, 2014; Baldiga, 2013).

Table 2: *Results*

	MCQ (1)	Omitted items (2)	Q1 (3)	Q2 (4)	Q3 (5)	Share correct (6)	Total points (7)
Panel A. Controlling for total points from the 1st midterm exam							
Female	-0.185*	0.352***	0.157	-0.024	0.093	-0.016	-0.073
	(0.101)	(0.103)	(0.108)	(0.106)	(0.101)	(0.019)	(0.100)
Points from Q1-Q4	0.303***	-0.121*	0.253***	0.191***	0.310***	0.035***	0.435***
	(0.055)	(0.066)	(0.064)	(0.067)	(0.070)	(0.010)	(0.055)
MCQ points	0.279***	-0.339***	0.230***	0.114*	0.175***	0.035***	0.291***
	(0.056)	(0.057)	(0.059)	(0.061)	(0.058)	(0.010)	(0.052)
Observations	401	401	401	401	401	332	401
R-squared	0.212	0.186	0.129	0.054	0.130	0.147	0.305
Panel B. Controlling for points from open-ended questions from the 1st midterm exam							
Female	-0.248**	0.430***	0.104	-0.050	0.053	-0.023	-0.140
	(0.101)	(0.108)	(0.106)	(0.105)	(0.100)	(0.018)	(0.098)
Points from Q1-Q4	0.444***	-0.293***	0.369***	0.248***	0.398***	0.052***	0.583***
	(0.051)	(0.064)	(0.061)	(0.058)	(0.062)	(0.009)	(0.051)
Observations	401	401	401	401	401	332	401
R-squared	0.163	0.113	0.096	0.046	0.111	0.115	0.252

*Notes: The table reports the results for equation 1 for exam performance in the second midterm exam controlling for the performance in the first midterm exam. Data includes a set of 12 midterm exams in an undergraduate level microeconomics course at a Finnish university across a six-year period of 2010 and 2012–2016. MCQ refers to standardized points from MCQ's, omitted items refers to the standardized number of omitted MCQ items, Q1(2/3) refers to standardized points from question 1(2/3), share correct refers to the standardized share of correctly answered MCQ items, and total points refers to standardized total points from the exam. Robust standard errors are reported in parenthesis. Controls include standardized total points from questions 1-4 and MCQ points from the first midterm exam in panel a) and standardized total points from questions 1-4 from the first midterm exam in panel b). Year indicators are included. Data for the share of correct answers (column (6)) is missing for the year 2016. *** p < 0.01, ** p < 0.05, * p < 0.1.*

I conduct a few robustness checks to show evidence that the results are not due to females performing worse, but might, instead, reflect differences in exam behavior. First, I show that gender is not associated with worse performance in the course subject and hence, women perform as well as men in the other parts of the exam. This holds true for all the questions (Questions 1 to 3) as shown in columns (3) to (5) in both panel a) and b) in Table 2. The coefficients for the female dummy are small and imprecisely estimated. Moreover, the coefficients are both negative and positive indicating that there are no systematic differences between male and female performance. The results for questions 2 and 3 (columns (4) and (5)) are particularly interesting as these questions are of similar type as most of the MCQ's; calculations, but lengthier and more time-consuming than in the MCQ part. The coefficients for the female dummy are small and imprecisely estimated for these questions. Hence, women do equally well in them when controlled for the past performance.[4]

Second, I show that omitting questions is not – with high probability – due to women performing worse in multiple-choice type questions but suggest that this effect is due to differences in exam behavior. I regress the share of correct answers in

[4] *The results remain similar, albeit larger in magnitude, when controlling only for the points from questions 2 and 3 from the first midterm exam (not reported here). Females score 0.24 standard deviations fewer points from the MCQs and omit 0.42 standard deviations more items than male students when controlled only for the performance in questions 2 and 3 in the previous exam.*

MCQ's on the female dummy using the same controls as above. As shown in column (6), gender is not a driving force for answering incorrectly.[5]

Third, I use a different proxy to measure ability. Namely, I use the performance (standardized points from questions 1, 2 and 3) in the same midterm exam as a proxy for the ability in the subject. I acknowledge that females might have a time preference for the open-ended questions, which could be driving the effect. If females prefer open-ended questions and devote more of the limited exam time towards them, they might do relatively better in them than in the MCQ items. The point estimates for the female dummy are now larger, 0.49 and -0.34, for the number of omitted items and MCQ points, respectively, and significant at the 1% significance level (not reported here).

Fourth, one might argue that there is a selection bias; females that attend the microeconomics course might have lower unobserved ability than women in general and hence, perform worse. However, given the mathematical nature of microeconomics, women attending the course might actually be, for example, more strategic than women in general. Hence, these individuals might be, for example, less risk averse than women on average. This, in turn, implies that the results provide only the lower boundary for the gender gap in performance and behavior in MCQ's.

Lastly, I look at whether the worse performance of females in MCQ's translates into fewer total points in the exam. Females score 0.07 standard deviations fewer total points (≈ 1.6 points) when controlled for the performance in the first midterm exam (column (7) in Table 2). However, this difference is not statistically significant. The data lacks the information on the final grade, which is given on a scale of 0–5, hence, effects on the final grade cannot be assessed. The MCQ's contributed to 36% to 45% of the total points, depending on the exam. If their share was higher, the effects of gender differences in risk preferences would very likely have a larger effect also on total points and the final grade.

4 Discussion

In this study, I analyze performance differences in multiple-choice questions in undergraduate microeconomics. The results suggest that the gender differences may actually reflect differences in behavior in a very particular test setting rather than genuine differences in skills. Females omit more items than males in MCQ's when they are penalized for wrong answers. The unobserved ability is hard to control for. This study had the privilege of being able to control with the performance in the first midterm exam and with similar open-ended questions. The reason why women tend to omit more items might stem from the gender differences in risk preferences; women consider risk to be more of a threat, and may feel fear of loss that, in turn, leads to over-weighting the probability of a loss, while men are, in general, more confident of winning.

Shurchkov and Eckel (2018) and Niederle (2017a) discuss possible policy interventions by government or business that might lessen gender differences in outcomes; Sandberg (2015) argues that the solution to the gender gap is to encourage women to "lean in", for example, to take on more risk and to engage to a greater extent in competition. Bohnet (2016), in turn, argues that "de-biasing" institutions can address the gender gap more effectively than policies that are designed to change the way women behave. Niederle (2017a) also suggests that a more cautious and prudent approach might be to address whether institutions differ and hence their potential effect on gender differences in economic outcomes. One way to do this would be to change the test-setting so that guessing is no longer penalized. The guessing penalty was removed, for example, from the SAT[6] in 2016 (Shurchkov and Eckel, 2018).

[5] *Data for the share of correct answers is missing for the year 2016.*

[6] *The SAT is a standardized test widely used for college admissions in the United States.*

References

Baldiga, K. (2013). "Gender differences in willingness to guess." *Management Science*, *60*(2), 434–448.

Bohnet, I. (2016). "What works: Gender equality by design." *Harvard university press*.

Brodeur, A., Cook, N., Hartley, J., and Heyes, A. (2022). "Do pre-registration and pre-analysis plans reduce p-hacking and publication bias?" *Available at SSRN*.

Byrnes, J. P., Miller, D. C., and Schafer, W. D. (1999). "Gender differences in risk taking: A meta-analysis." *Psychological bulletin*, *125*(3), 367.

Camerer, C. F., Dreber, A., Forsell, E., Ho, T.-H., Huber, J., Johannesson, M., Kirchler, M., Almenberg, J., Altmejd, A., Chan, T., et al. (2016). "Evaluating replicability of laboratory experiments in economics." *Science*, *351*(6280), 1433–1436.

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

Drazen, A., Dreber, A., Ozbay, E. Y., and Snowberg, E. (2021). "Journal-based replication of experiments: An application to "being chosen to lead"." *Journal of Public Economics*, *202*, 104482.

Ferber, M. A., Birnbaum, B. G., and Green, C. A. (1983). "Gender differences in economic knowledge: A re-evaluation of the evidence." *The Journal of Economic Education*, *14*(2), 24–37.

Iriberri, N., and Rey-Biel, P. (2021). "Brave boys and play-it-safe girls: Gender differences in willingness to guess in a large scale natural field experiment." *European Economic Review*, *131*, 103603.

Lumsden, K. G., and Scott, A. (1987). "The economics student re-examined: Male-female differences in comprehension." *The Journal of Economic Education*, *18*(4), 365–375.

Niederle, M. (2017a). "A gender agenda: A progress report on competitiveness." *American Economic Review*, *107*(5), 115–119.

Niederle, M. (2017b). 8. Gender. In J. Kagel & A. Roth (Ed.), The Handbook of Experimental Economics, Volume 2: The Handbook of Experimental Economics (pp. 481-562). Princeton: Princeton University Press.

Niederle, M., and Vesterlund, L. (2007). "Do women shy away from competition? Do men compete too much?" *The Quarterly Journal of Economics*, *122*(3), 1067–1101.

Pekkarinen, T. (2014). "Gender differences in behaviour under competitive pressure: Evidence on omission patterns in university entrance examinations." *Journal of Economic Behavior & Organization*.

Sandberg, S. (2015). "Lean in-women, work and the will to lead." *New York: Alfred A. Knopf.*

Saygin, P. O., and Atwater, A. (2021). "Gender differences in leaving questions blank on high-stakes standardized tests." *Economics of Education Review*, *84*, 102162.

Shurchkov, O., and Eckel, C. C. (2018). "Gender differences in behavioral traits and labor market outcomes." In *The Oxford Handbook of Women and the Economy*, Oxford University Press.

Walstad, W. B., and Robson, D. (1997). "Differential item functioning and male-female differences on multiple-choice tests in economics." *The Journal of Economic Education*, *28*(2), 155–171.

Wieland, A., Sundali, J., Kemmelmeier, M., and Sarin, R. (2014). "Gender differences in the endowment effect: Women pay less, but won't accept less." *Judgment and Decision Making*, *9*(6), 558–571.