Gender bias in risk aversion: evidence from multiple choice exams

Marín, Carmen and Rosa-García, Alfonso

Universidad de Murcia

23 December 2011

Gender bias in risk aversion: evidence from multiple choice exams*

Carmen Marín[†‡] Alfonso Rosa-García[§¶]

December 23, 2011

Abstract

We provide evidence on a gender bias in risk aversion among students of economics in Spain. In a sample of 1947 multiple choice exams with penalization for errors, we find that women consistently answer less questions, while differences in marks are not significant. These empirical results are consistent with a recent theoretical prediction of the effect of risk aversion in test exams. This finding shows that women can suffer a disadvantage with this kind of exams widely used in education.

Keywords: gender differences, risk aversion, multiple choice

JEL Classification: A22, I21, J16

*We are indebted to Josep-Oriol Escardibul and Hubert Janos Kiss for helpful comments. This paper has benefitted also from suggestions provided by participants of the XXXVI Simposio Asociación Economía Española (SAEe), the III Jornadas de Docencia en Economía and of the II Jornada sobre la Dimensión Social de la Educación Universitaria en España. Alfonso Rosa-García kindly acknowledge the financial support from the Spanish Ministry of Science and Innovation under the project ECO2010-19830.

[†]Corresponding author. Email: carmenma@um.es; Tel.: +34 868883754; Fax: +34 868883758

[‡]Address: Dpto. de Fundamentos del Análisis Económico. Facultad de Economía y Empresa. Universidad de Murcia, Campus de Espinardo. 30100 Murcia.

[§]Email: alfonso.rosa@um.es; Tel.: +34 868883728

[¶]Address: Dpto. de Fundamentos del Análisis Económico. Facultad de Economía y Empresa. Universidad de Murcia, Campus de Espinardo. 30100 Murcia.

1 Introduction

Multiple choice tests are widely used as an easy and objective mean to grade students. However, different studies have remarked that test performance is not independent of its design. Usual penalties, in which wrong answers decrease scores, could induce a gender bias as men and women have shown different risk patterns. The literature has also emphasized that test scrambling could yield different results, making the exams more difficult to understand for students. In this paper we intend to contribute to this literature, studying how the test structure affected our students' behavior and performance.

We analyze several exams completed by undergraduate students for the subject Political Economy, taught in the Business Degree of the Universidad de Murcia. Our sample consists of 1947 individual multiple choice tests, in which the existence of a gender difference together with the possibility of a test scrambling effect are alternatively explored. Consistently across all the exams, males tend to answer a higher number of questions than females, which is in line with the empirical and experimental evidence on the higher risk aversion observed in women. However, with regard to a possible bias in score output, the data show only mixed evidence. Test scrambling has an effect when the whole sample is analyzed and only over decisions on the number of answered questions. These outcomes are all consistent with previous evidence on scores, and we add two new insights: a link between test behavior and the higher risk aversion of females as well as partial effects of test scrambling.

The possibility of a gender bias in Economy test scores has been studied for years (Siegfried, 1979; Ballard and Johnson, 2005). This literature has usually identified a higher score for males,[1] which is also in line with the gender gap in favor of men reported on math scores (Hedges and Nowell, 1995). Empirical evidence shows that this bias is correlated with the level of gender equality in the country, and seems to decrease across years (Guiso et al., 2008).[2] In our sample, we find that males improve upon the scores of girls in 4 of the 7 exams, although the difference was significant only in one case. This outcome agrees with the weakness of a gender bias in scores that would favor males facing this kind of exams.

Our main contribution is to report evidence of a higher degree of risk aversion in females, measured as the amount of answered questions: females answered around 4% less questions than males. Moreover, we find it to be significant in each of the sub-samples. Our exams are designed with a penalty that gives an expected zero score to a random answer. This widely used method seems to penalize those agents who are more risk averse (Burgos, 2005), so our observed differences can therefore be explained by a gender bias in

[1] But Makridou-Bossiou (2003) provides the opposite evidence for Greek students.

[2] Although differences in mean seem to be correlated with cultural factors, variability seems to be worldwide larger in the case of males (Machin and Pekkarinen, 2008).

risk aversion. Thus, we provide new data that support the evidence on a higher risk aversion for women (Croson and Gneezy, 2009). Our results also support the theoretical analysis by Espinosa and Gardeazábal (2010) that predicts high differences in number of answered questions but low differences in scores for agents with different risk aversion.

The literature has also explored scrambling test effects. Typically, non-significant (or very slight) differences are found between those students who solve exams with questions in a randomized order against those students who solve them according with the sequence of topics taught in class (Gohmann and Spector, 1989). We also do not find a significant effect in score, but we do provide some evidence on the effect over the number of answered questions: people doing the scrambled version of the exam answered less questions. This suggests that when individuals confront a less uncertain type of exam, they become more conservative.

2 The Data

We analyze a set of seven end-of-term examinations of the subject Political Economy across a five years period. The sample includes only those periods when we were able to recover all the marks (June and September 2005, 2008 and 2009, and September 2007).

Our target subject was taught in the first of the 3-years Degree in Business of the Universidad de Murcia and was a typical introductory course to Microeconomics. The final exam included a multiple choice test with 20 questions that accounted for the 40% of the final mark, and that the student had to pass in order to be evaluated. Each of the test questions was comprised of three possible options, being correct only one of them. Each wrong answer penalized 0.5 correct answers, so that the expected score of a random answer was zero. There was no penalty for unanswered questions. In this type of exam, if students assign to each option a positive probability of being correct, they have in fact to decide whether to participate or not in a lottery with positive expected value (if the true answer is usually a higher probability). In that case risk aversion matters, since more averse students are expected to answer less questions.

Exams were distributed to students in four versions that only differed in the order of questions. Type 1 exams presented questions in the same order they had been taught during the semester. Rest of types presented questions in a randomized order. Table 1 reports some descriptive statistics of the exams.

	J05	*S05*	*S07*	*J08*	*S08*	*J09*	*S09*	*Average*
Amount of exams	**485**	**161**	**195**	**343**	**210**	**318**	**235**	*278*
Females	**295**	**99**	**110**	**179**	**113**	**167**	**133**	*157*
Female score (Max=10)	6.10	5.17	4.54	4.45	3.59	4.12	4.78	*4.68*
Male score (Max=10)	6.13	5.04	5.02	4.81	3.28	4.08	4.89	*4.75*
Female answers (Max=20)	16.3	15.8	14.9	14.4	14.5	14.9	14.9	*15.1*
Male answers (Max=20)	16.9	16.3	15.9	15.5	15.5	15.6	15.6	*15.9*
Type 1	**122**	**51**	**57**	**99**	**60**	**84**	**59**	*76*
Type 1 score (Max=10)	6.02	5.10	4.39	4.49	3.63	4.24	4.81	*4.67*
Other types score (Max=10)	6.15	5.13	4.89	4.67	3.38	4.06	4.83	*4.73*
Type 1 answers (Max=20)	16.41	15.59	15.16	14.93	14.72	15.14	14.93	*15.27*
Other type answers (Max=20)	16.58	16.14	15.38	14.87	15.10	15.25	15.28	*15.52*

Table 1. Descriptive statistics of the sample

There is a total of 1947 individual tests, ranging from a minimum of 161 in the sub-sample of September 2005 and to a maximum of 485 in June 2005. From the total, we have 1096 done by women (56.29%) and 532 Type 1 exams (27.32%). We observe that differences in score are very small, while differences between males and females in answered questions are persistent. Differences in the number of answered questions between individuals facing Type 1 exam and the rest are not so clear.

3 Empirical Evidence

In order to disentangle if students' behavior is dependent on gender, we have regressed the number of blank (unanswered) questions against a set of explicative variables. In particular, the variables woman (0 for men, 1 for women) and Type (0 for the exam in a randomized order, 1 for the ordered exam) were included. The specific characteristics of the examinations were controlled using the variable Mean, that consisted of the average number of blank questions in each sub-sample. We have also controlled for the individual's knowledge introducing the score quartile to which the students have belonged in each examination (the 3rd quartile is the one with the highest level of blank answers and we take it as the baseline). Table 2 shows the estimation results.

Dependent Variable: Blank			Included observations: 1947	
Variable	**Coefficient**	**Coefficient**	**Coefficient**	**Coefficient**
C	4.012***	4.026***	-0.080	-0.074
	(0.078)	(0.085)	(0.319)	(0.320)
Woman	0.701***	0.676***	0.773***	0.764***
	(0.097)	(0.113)	(0.096)	(0.102)
Type	0.220**	0.168	0.154	0.135
	(0.108)	(0.164)	(0.096)	(0.147)
Type*Woman		-0.091		0.035
		(0.218)		(0.194)
1st			-1.858***	-1.858***
			(0.123)	(0.123)
2nd			-0.414***	-0.414***
			(0.114)	(0.114)
4th			-0.369**	-0.369**
			(0.128)	(0.128)
Mean			1.051***	1.051***
			(0.068)	(0.068)
R-squared	*0.133629*	*0.028579*	*0.225932*	*0.225945*
Adjusted R-squared	*0.130949*	*0.027079*	*0.223538*	*0.223151*

Table 2. Variables are significant at ***1%, **5% or *10%

Women answered significantly less questions than men. The same result holds for the type of exam, since individuals doing the non-scrambled version answered significantly less questions. Individuals with the non-scrambled exam answered 0.22 questions less (around 1%) while women left 0.7 (around 3.5%) unanswered questions more than men. Interestingly, people in the fourth quartile answered more questions than those in the third one. The effect of Type is not significant when we interact it with Woman.

The most consistent result is that the woman variable is strongly significant across all the sample. In fact, estimating for each of the seven examinations, we find that gender is always a significant explanatory variable, while Type is not significant in any of the seven sub-samples. Table 3 shows these results.

	JUN 05	SEP 05	SEP 07	JUN 08	SEP 08	JUN 09	SEP 09	Full sample
Women	*Yes****	*Yes**	*Yes****	*Yes****	*Yes****	*Yes****	*Yes****	*Yes****
Type	*No*	*No*	*No*	*No*	*No*	*No*	*No*	*Yes**

Table 3. Exams in which blank answers depend significantly on each variable is sginficant at ***1%, **5% or *10%

This evidence suggests a significant difference in men and women's behavior in our test exams. Given the increasing evidence of differences in risk aversion between genders, we argue that risk aversion is a potential explanation for this divergence. If it is the case, following Espinosa and Gardeazábal (2010) we

should observe only small differences in marks. Table 4 shows the results of regressing marks against gender, controlling for the student's quartile and the sub-sample mean. In this case, women obtained lower scores than men, although the difference was not significant, and type had no effect. This is consistent with the explanation of risk aversion as the cause of the gender gap.

Dependent Variable: Mark			**Included observations: 1947**	
Variable	**Coefficient**	**Coefficient**	**Coefficient**	**Coefficient**
C	4.872***	4.860***	-0.507***	-0.512***
	(0.077)	(0.083)	(0.092)	(0.093)
Woman	0.028	0.050	-0.048	-0.037
	(0.095)	(0.112)	(0.030)	(0.036)
Type	-0.103	-0.057	0.016	0.040
	(0.106)	(0.162)	(0.034)	(0.052)
Type*Woman		-0.080		-0.042
		(0.214)		(0.068)
1st			3.095***	3.095***
			(0.043)	(0.043)
2nd			1.428***	1.428***
			(0.040)	(0.040)
4th			-2.126***	-2.126***
			(0.045)	(0.045)
Mean			0.963***	0.963***
			(0.018)	(0.018)
R-squared	*0.000527*	*0.000599*	*0.898730*	*0.898750*
Adjusted R-squared	*-0.000502*	*-0.000944*	*0.898416*	*0.898384*

Table 4. Variables are significant at ***1%, **5% or *10%

Table 5 shows that, across all the sub-samples, there is a significant effect for gender (with men obtaining a higher score than women) only in one of the seven cases. Type is not significant in any of them.

	JUN 05	SEP 05	SEP 07	JUN 08	SEP 08	JUN 09	SEP 09	Full sample
Women	*No*	*No*	*No*	*No*	*No*	*Yes***	*No*	*No*
Type	*No*	*No*	*No*	*No*	*No*	*No*	*No*	*No*

Table 5. Exams in which marks depend significantly on each variable is sginficant at ***1%, **5% or *10%

Despite the weak gender effect in test scores, we still find some significant differences in some groups of individuals. In particular, some differences appear in answered questions and scores among the students with the lowest marks. Table 6 shows the regressions including all the control variables and the interacts

between Woman and Type with the two lowest quartiles in scores.

observations: 1947	Dependent Variable: Blank		Dependent Variable: Mark	
Variable	Coefficient	Coefficient	Coefficient	Coefficient
C	-0.160		-0.544***	-0.530***
	(0.326)		(0.095)	(0.091)
Woman	0.718***	0.765***	0.021	
	(0.129)	(0.084)	(0.049)	
Type	-0.030		0.040	
	(0.171)		(0.052)	
Type*Woman	0.030		-0.043	
	(0.194		(0.068)	
Woman*(3rd&4th)	0.102		-0.107*	-0.097**
	(0.173)		(0.061)	(0.041)
Type*(3rd&4th)	0.361*	0.343**		
	(0.193)	(0.141)		
1st	-1.706***	-1.777***	3.155***	3.149***
	(0.165)	(0.127)	(0.055)	(0.049)
2nd	-0.2567	-0.328***	1.489***	1.484***
	(0.159)	(0.117)	(0.053)	(0.046)
4th	-0.371***	-0.380***	-2.125***	-2.124***
	(0.128)	(0.127)	(0.045)	(0.045)
Mean	1.051***	1.025***	0.962***	0.963***
	(0.068)	(0.022)	(0.018)	(0.018)
R-squared	*0.227492*	*0.227269*	*0.898910*	*0.898875*
Adjusted R-squared	*0.223903*	*0.225279*	*0.898492*	*0.898615*

Table 6. Variables are significant at ***1%, **5% or *10%

Mean includes the average number of blank answers (scores) in each sub-sample, while Blank (Mark) is the dependent variable. For the case of Blank answers, the gender effect is completely robust. Nevertheless we also identify additional effects for the Type 1 exams: individuals with a lower score answered less questions (more blank answers) when they had the ordered exam. This suggest that in the less uncertain environment of the ordered exam, individuals were less risky. With regard to scores, the finding is that between students with low marks, women did worse than men. This suggests that possibly a gender gap in scores did occur among the individuals with less knowledge.

4 Conclusion

In our multiple choice tests, males and females mainly differ in the level of non-answered questions, which can be the consequence of a higher risk aversion of women. Ceteris paribus, a lower degree of risk aversion

can generate a higher expected mark. For risk averse agents, Espinosa and Gardeazábal (2010) show that the level of answered questions should be relevant, although the differences in scores are expected to be small. Consistently with this prediction, we find that women answered less questions in our exams but the difference in scores between genders was very small and mostly non-significant. Another potential reason that could affect risk averse agents is the effect of test scrambling. In this respect we obtain it has very small effects and only with regard to the number of answered questions when using the full sample of exams.

Therefore, we find consistent evidence of different men and women behavior in multiple choice tests, being differences of risk aversion a robust explanation for this fact. Although the differences in scores are small, an effective discrimination against women due to their higher risk aversion can be in place. Potential ways of modifying this result are to increase the number of questions or to avoid the penalization for the wrong answers.

References

[1] Ballard, C. and M. Johnson (2005): "Gender, expectations, and grades in introductory microeconomics at a US university", *Feminist Economics*, 11(1): 95-122.

[2] Burgos, A. (2004): "Guessing and gambling", *Economics Bulletin*, 4 (4), 1-10.

[3] Croson, R. and U. Gneezy, (2009): "Gender differences in preferences", *Journal of Economic Literature*, 47 (2) (2009), pp. 1–27.

[4] Espinosa, M. P. and J. Gardeazábal (2010): "Optimal correction for guessing in multiple-choice tests", *Journal of Mathematical Psychology* 54, 415 425

[5] Gohmann, Stephan F. and Lee C. Spector (1989): "Test scrambling and student performance", *Journal of Economic Education*, 20, 235–238.

[6] Guiso, L., F. Monte, P. Sapienza and L. Zingales (2008): "Culture, gender, and math", *Science*, 320, 1164–1165.

[7] Hedges, L. V. and A. Nowell (1995): "Sex differences in mental test scores, variability, and numbers of high-scoring individuals", *Science*, 269, 41-45

[8] Machin, S. and T. Pekkarinen (2008): "Global sex differences in test score variability", *Science*, 322, 1331–1332.

[9] Makridou-Bousiou, Despina (2006): "Gender differences in economic knowledge in Greece", *International Business & Economics Research Journal*, 5, 35-41

[10] Siegfried, John (1979): "Male-Female differences in economic education: a survey", *Journal of Economic Education*, Vol. 10, 1-11.