UNIVERSITY OF GOTHENBURG
SCHOOL OF BUSINESS, ECONOMICS AND LAW

WORKING PAPERS IN ECONOMICS

No 461

Are Men Really More Overconfident than Women? A Natural Field Experiment on Exam Behavior

Olof Johansson Stenman and Katarina Nordblom

August 2010

ISSN 1403-2473 (print)
ISSN 1403-2465 (online)

Department of Economics
School of Business, Economics and Law at University of Gothenburg
Vasagatan 1, PO Box 640, SE 405 30 Göteborg, Sweden
+46 31 786 0000, +46 31 786 1326 (fax)
www.handels.gu.se   info@handels.gu.se

Are Men Really More Overconfident than Women? A Natural Field Experiment on Exam Behavior*

Olof Johansson-Stenman and Katarina Nordblom

Department of Economics, School of Business, Economics and Law,

University of Gothenburg, Box 640, SE 40530 Gothenburg, Sweden.

E-mail: Olof.Johansson@economics.gu.se; Katarina.Nordblom@economics.gu.se

Telephone: +46 31 7861000; Fax: + 46 31 7861053

Abstract This paper reports from a simple natural field experiment based on an economics exam. Part of the exam consisted of 30 multiple choice questions, where the students obtained 1 point per correct answer while 1 point was deducted for each incorrect answer. We find no significant gender differences in overconfidence, irrespective of how we measure it, i.e., whether we regard the number of questions answered, the number of questions answered incorrectly or the answering patterns for the most difficult questions. Hence, the data provides no support for the frequently proposed hypothesis that men are more overconfident than women.

Keywords: overconfidence, gender differences, exam behavior, field experiment.

JEL classifications: A22, D80, J16.

* We wish to thank Fredrik Carlsson, Peter Martinsson, Ola Olsson, Måns Söderbom, and Joakim Westerlund for helpful comments and suggestions. We are also grateful to the Swedish Research Council for financial support.

1. Introduction

The aim of the present paper is to test for gender differences in overconfidence based on a natural field experiment (using the terminology of Harrison and List, 2004) in terms of an exam at a Swedish university. Specifically, we analyze the gender-based response patterns to multiple-choice questions, where a correct answer is worth one point while one point is deducted for each incorrect answer.

Overconfidence, i.e., that we tend to systematically overestimate our own abilities in various respects, is according to psychologists more common than most people realize. DeBondt and Thaler (1995) argue that, “Perhaps the most robust finding in the psychology of judgment is that people are overconfident,” see e.g. Taylor and Brown (1994) and Baumeister (1998) for overviews of the large psychological literature on the subject. Within economics, Odean (1999) presents empirical evidence on excessive stock market trading due to overconfidence, while Malmendier and Tate (2005, 2008) provide evidence that CEOs tend to be overconfident about their ability to manage a company. With respect to education, Gibbons and Silva (2007) found that 96% of secondary school pupils in UK believe that they are average or above average when asked how good they are at their school work, and Thorpe et al. (2007) found that 90% of first year university students consider themselves to be average or above.

There is also some evidence that firms realize and profit on the fact that people are overconfident. For example, Grubb (2009) notes that consumers overestimating the precision of their demand forecasts creates an incentive for firms to offer tariffs with included quantities at zero marginal cost, followed by steep marginal charges. He then concludes that this matches the observed cellular phone service pricing patterns in many markets.

While there are obvious negative instrumental effects of overconfidence, there are also possible positive effects, e.g., in terms of overcoming imperfect willpower (Benabou and Tirole, 2002) and increasing the chance of success in risky activities (Compte and Postlewaite, 2004).[1] Hirschleifer et al. (2010) found that overconfident CEOs in innovative industries are more effective innovators than their less confident counterparts. With respect to education, Chevalier et al. (2009) found that high school students with a more positive view of their academic abilities are more likely to expect to continue to higher education even after controlling for observable measures of ability and students' characteristics.

Why is it important to investigate gender-based differences in overconfidence? One reason is that, due to the above mentioned implications, possible gender differences in overconfidence constitute a potential explanation to observed gender differences in, e.g., wage bargaining and entrepreneurial activities, which in turn are possible reasons for why women in virtually all countries have substantially lower average wages than men. For example, Bertrand and Hallock (2001) found that only 2.5 percent of the top five executives for a large group of U.S. firms were women.

Turning to the empirics of gender differences in overconfidence – the subject under investigation in the present paper – it is often stated, and sometimes found, that men on average are more overconfident than women (e.g., Lundeberg et al., 1994). However, although there are several results in that direction when overconfidence is measured by individual responses in surveys (e.g., Lundeberg et al., 1994; Prince, 1993), the empirical evidence regarding gender-related patterns when it comes to ac-

[1] The latter can be the case, e.g., since people tend be unreasonably risk averse with respect to small stakes gambles; cf. Rabin (2000), Rabin and Thaler (2001), and Johansson-Stenman (2010). Köszegi (2006) provides an illuminating analysis of the tradeoffs between the "ego-utility" of overconfidence and various instrumental effects.

tual behavior is less clear. The importance of using real incentives when measuring overconfidence is highlighted by Cesarini et al. (2006), who found that the degree of overconfidence were largely reduced when introducing monetary incentives in several confidence interval estimation tasks. In addition, although not their main task of investigation, they found no indication based on their sample that men are more overconfident than women.

Barber and Odean (2001) constitute a much-cited study on excessive stock market trading. Due to transaction costs, it is generally not rational to trade unless you have better information than the market (Barber and Odean, 2000; Odean, 1999). Barber and Odean (2001) found that men trade substantially more than women and interpret this finding as support for men being more overconfident. In other words, they *assume* that men's observed higher trading activities are driven by their higher overconfidence, based on the existing (primarily survey-based) evidence of such a gender-based difference in overconfidence. However, there are other possible explanations. For example, maybe men on average experience lower disutility with spending time thinking about the stock market development than women do. Recent evidence from asset-market experiments also calls into question the interpretation by Barber and Odean. Biais et al. (2005) found that higher over-confidence implies more trade and also that men trade more than women, yet they did not find male students to be more overconfident in their miscalibration measures. Deaves et al. (2008) found that overconfidence drives trading, but obtained no gender-based differences in either trading activity or overconfidence. Deaves et al. (2010) analyzed market forecasters and found that they, as a group, are overconfident. Yet, again no gender-based differences in overconfidence were found.

Bengtsson et al. (2005) is, in a sense, the paper most strongly related to the present paper as it also deals with exam behavior. They measured the fraction of men and women who voluntarily answered an exam question for which the result only mattered if the student had very good results on the other answered questions and found that women answered this question to a significantly lower extent than men. However, although differences in overconfidence certainly constitute a possible explanation, it is again not the only one. For example, one might conjecture that there are differences in the disutility of answering this question; perhaps men to a larger extent are thrilled by the induced competitive atmosphere, consistent with previous findings that women tend to be more averse to competition than men (Gneezy et al., 2003).[2] There are also a number of experimental studies showing that men tend to prefer a tournament setting, where the best performers are getting an unproportionally large share of the payoffs, than women do, see e.g. Niederle and Vesterlund (2007), Vandegrift and Yavas (2009) and Cason et al. (2010).

In the present paper, we test for gender differences in overconfidence based on a very simple natural field experiment, presented in Section 2, in terms of a university exam. Half of the exam consists of multiple-choice questions where a correct answer is worth one point, while one point is deducted for each incorrect answer and not answering implies no point. Clearly, for a given knowledge level of the student, the higher the number of questions answered, the higher the level of overconfidence (or the lower the level of underconfidence). Yet, e.g., since we cannot measure the knowledge level perfectly for each individual, there is no unique ideal measure of gender-

[2] However, see also Gneezy et al. (2009), who show the pattern that men living in patriarchal societies behave more competitively in experiments than women, while women living in matrilineal societies behave more competitively than men. This suggests that cultural explanations of such gender differences are important.

based overconfidence differences. We will therefore, in Section 3, present results for a number of different measures. Yet, for each measure presented, we obtain no significant gender differences. In Section 4, we then present the results from regression analysis, where we explicitly correct for other potential variables that might drive the results. Again, we find no significant gender effects. Section 5 concludes the paper.

2. The Experiment on Gender Differences in Overconfidence

The purpose of the experiment is to identify gender differences in overconfidence (or possibly underconfidence). Hence, our purpose is not to measure the absolute extent to which men and women actually are overconfident, but solely to measure the gender-based *difference* in overconfidence.

The experiment was conducted during an exam in elementary macroeconomics given as part of the business program at the University of Gothenburg; the sample contains 86 females and 88 males, i.e., in total 174 students. The maximum test score was 60 points – 30 for the longer questions and 30 for the multiple-choice questions, where one out of four alternatives was correct. This is a commonly used exam format. We have undertaken two manipulations in order to measure gender differences in overconfidence.

First, the standard format is that a correct answer to a multiple-choice question is worth one point, while an incorrect answer or not answering a question results in zero points. In contrast, we deducted one point for each incorrect answer.[3] Second, three of the questions were constructed to be more difficult than the others, and in fact so difficult that we did not expect the students to know the answers to them.[4] The cu-

[3] In total, though, one could not score below zero points on the multiple-choice part.

[4] One question was impossible to know the answer to even after having taken the course (Q10 in Table 1). The second required quite a lot of (for the level) advanced mathematics (Q19), and to the third diffi-

toff to “pass” was 27 points, while a “high pass” required 42 points. Of course, the students did not know that they were participating in a scientific study.

Ignoring the time cost associated with answering a specific multiple-choice question, the choice rule that a respondent answers a question if and only if the subjective probability of a correct answer exceeds 50% is in most cases a sensible one. In order to see this, let p_{26} be the subjective probability that a student has 26 points (i.e., just below the cutoff) before answering an additional question and similarly that p_{27} is the subjective probability that the student has 27 points (i.e., just above the cutoff); moreover, let p^R be the subjective probability that the answer to the next question is right and p^L that it is wrong.[5] The probability that the student passes the exam will then increase when answering one additional question if and only if $p_{26} p^R > p_{27} p^L$, which in turn implies that

$$p^R > \frac{p_{27}}{p_{26} + p_{27}}.$$

Then it clearly follows that the likelihood to pass the exam when $p_{26} = p_{27}$ increases when answering an additional question if and only if $p^R > 0.5$. In other words, when a student, after having answered a number of questions, perceives to be equally likely that he/she is currently just below the pass cutoff as that he/she is just above it, the subjective probability to pass the exam increases if and only if the subjective probability that the answer to the additional question is correct exceeds 50%. Similarly, when the perceived likelihood of being just below the high pass cutoff is exactly the same as

---

cult question, the answer could be found in the textbook but was certainly not central to the course (Q18).

[5] Note that it will always be rational to answer the easiest remaining question, implying that a student will never answer a subjectively more difficult question than any of the remaining ones.

the perceived likelihood of being just above it, the subjective probability of receiving a high pass increases if and only if the subjective probability that the next answer will be correct is larger than 50%. However, that the perceived likelihoods of being below and of being above a limit are exactly equal is of course in most cases not correct. Yet, taking into account also that the students answered the longer questions (for which there is no incentive not to answer), it is presumably a reasonable approximation in most cases.[6] For example, if a student perceives the probability that he/she has 26 points is 10% larger than that he/she has 27 points, such that $p_{26}/p_{27} = 1.1$ and hence $p_{27}/(p_{26} + p_{27}) \approx 0.48$, then the subjective probability of passing the exam increases if the subjective probability of providing a correct answer is larger than 0.48.

Consequently, if men end up answering the multiple-choice questions to a larger extent than women, and at the same time are not doing better at these or at other questions, then we can see it as an indication of gender-based difference in overconfidence.

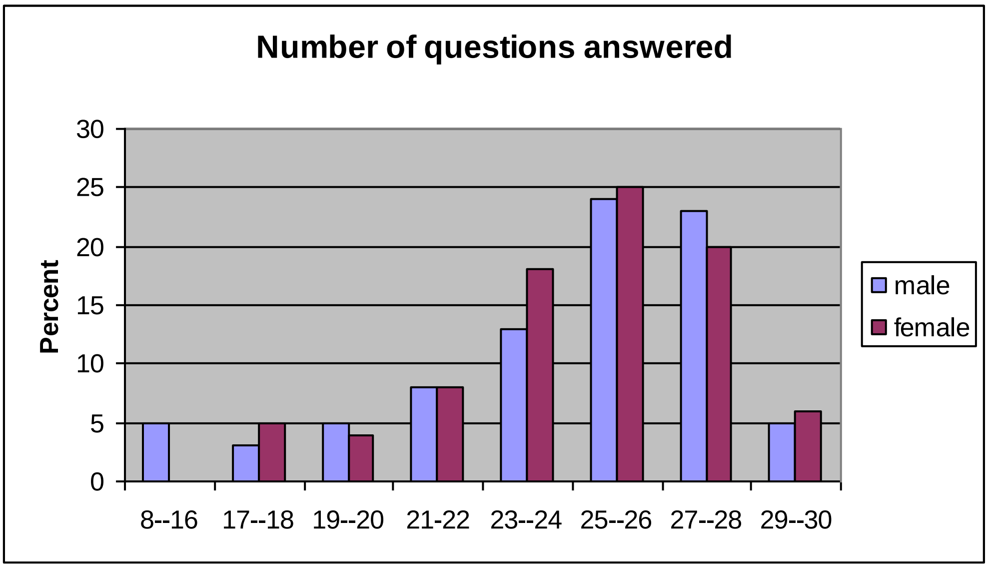
3. Results

Our first measure reflecting a gender difference in overconfidence is the difference in the total number of multiple-choice questions answered, reported in the first set of columns in Table 1 below. On average, women answered 24.9 multiple-choice questions out of 30, while men answered 24.1 questions; the difference is not statistically

[6] One may question the above reasoning since it presupposes that the students dealing with the risky choices of how many questions to answer act based on standard theory. There is indeed ample evidence that standard theory cannot accurately explain complex choices under risk, and that people in such situations often use simple heuristics (e.g., Kahneman et al., 1982). Yet, in the present case, the most obvious and salient choice rule is presumably to answer when it is more likely that the answer is correct than that it is not, and not to answer otherwise, which is hence the same choice rule as the one derived from standard theory.

significant based on a t-test (p = 0.23). The distributions of questions answered, by gender, are shown in Figure 1. Clearly, the distributions are very similar between men and women, although a larger share of male students answered very few questions. The difference between the distributions, based on a Wilcoxon-Mann-Whitney (WMW) test, is not significant (p = 0.63).

Figure 1. Distributions of number of questions answered for men and women.



However, from these patterns it can of course not be concluded that there is no gender difference in overconfidence. For example, if women answer correctly to a much larger degree than men, then such a difference can still exist.

To the far right in Table 1, we present the net incorrect answers, i.e., the number of incorrect answers minus the number of correct answers. A higher value could reflect a higher degree of overconfidence. These answers are not significantly different between men and women either; the means are not significantly different according to a t-test (p=0.27) and we cannot reject that the two distributions are from the same underlying distributions according to a WMW test (p=0.67).

Table 1. Number of answers to the multiple-choice questions (out of 30).

	All answers				Incorrect answers				Incorrect minus correct answers			
	Mean	Std. dev.	Min	Max	Mean	Std. dev.	Min	Max	Mean	Std. dev.	Min	Max
Total	24.4	3.9	8	30	3.9	2.8	0	16	-16.6	5.9	-26	4
Male	24.1	4.5	8	30	4.0	3.1	0	16	-16.2	6.8	-25	4
Female	24.8	3.1	17	30	3.8	2.4	0	13	-17.2	4.8	-26	3

In Figure 2 below, we plot the share of correct answers against the share answering the question for men and women. Again, the pattern is roughly the same for men and women; based on both simple t-tests where the means are compared (p=0.35) and WMW tests where the overall distributions are compared (p=0.89), the shares of correct and incorrect answers, respectively, do not differ significantly between male and female students.

Figure 2. Share of correct answers based on share answering the question for men and women.

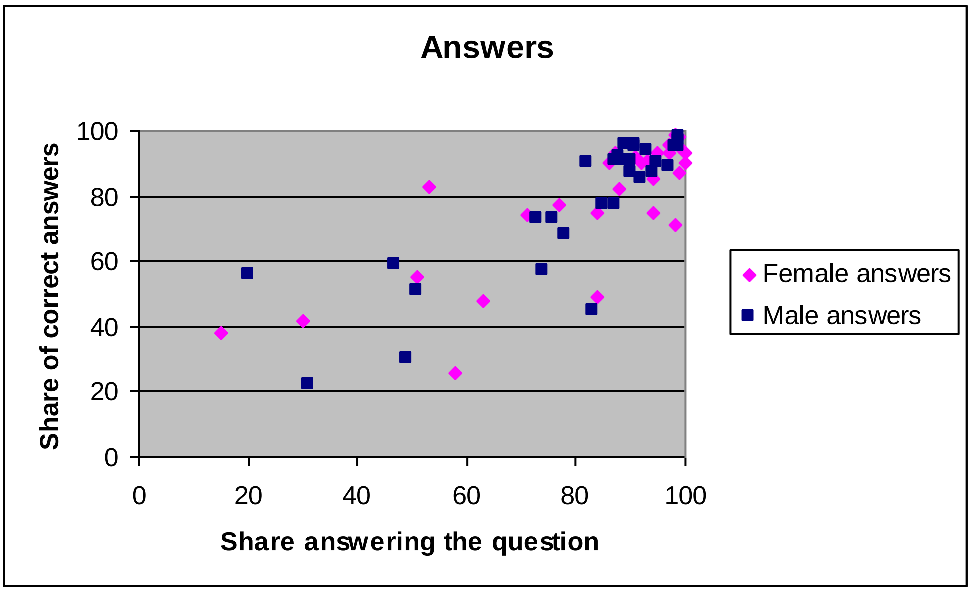


Table A1 in Appendix shows the response rate as well as the share of correct answers for each individual question, both in total and by gender. Not surprisingly, we found that some of the differences between the genders are significant, but there is no clear gender-based pattern. In brief, two questions were significantly (at the 5% level) more often answered among women than among men, while we found the opposite for two other questions.

Figure 2 and Table A1 in Appendix also reveal that some questions were perceived as easy and some as much more difficult. For example, 99 percent answered Question 1, and 98 percent of those who answered did so correctly, whereas Question 10 was answered by only 23 percent, of which 27 percent answered correctly, implying that only six percent of all students answered the question correctly. One may argue that when studying overconfidence, we should focus on the difficult questions. Following this argument, we define difficult questions in two different ways:

Ex ante difficult questions: Includes the three questions specifically constructed to be difficult.

Ex post difficult questions: Includes the questions that fewer than 50 percent of all students gave a correct answer to (whether they answered or not).

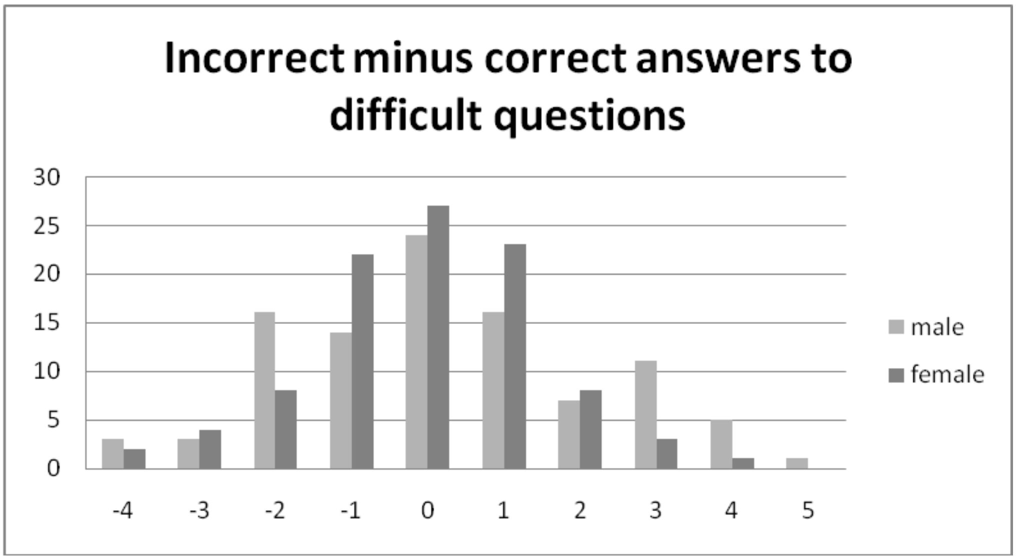
Six questions were considered difficult based on the ex post definition, and the ex ante difficult questions are a subsample of those. In Table 2, we present the number of incorrect answers minus the number of correct answers to the questions that were difficult according to the two definitions, respectively. A positive value can be interpreted as overconfidence. The distribution of the overall results for the ex post difficult questions are depicted in Fig. 3; the pattern for the ex ante difficult questions is similar.

Table 2. Respondents' overall results on the ex ante and ex post difficult questions.

	Ex ante difficult questions (out of 3)		Ex post difficult questions (out of 7)	
	Mean	Std. dev.	Mean	Std. dev.
	Number of answers			
Total	1.11	0.88	3.49	1.68
Male	1.22	0.87	3.50	1.68
Female	1.00	0.88	3.49	1.68
	Incorrect minus correct answers			
Total	0.17	1.10	0.06	1.81
Male	0.02	1.16	0.18	2.03
Female	0.33	1.01	-0.07	1.56

Men answered the ex ante difficult questions to a larger extent than women, as seen in the first set of columns in Table 2. The distributions differ significantly with p=0.073 according to a WMW-test and the means are different with p=0.088 according to a t-test. However, this difference cannot be attributed to overconfidence since the bottom part of the table reveals that men actually do better than women given that they answer these questions (p=0.043 in WMW for the distributions and p=0.069 for the t-test of means).

Figure 3. Incorrect minus correct answers on the ex post difficult questions.



Turning to the ex post difficult questions, i.e., fewer than 50 % answered correctly, we no longer find any gender differences. The distributions of incorrect minus correct answers for men and women do not differ significantly according to a WMW test p=0.56), and nor do the mean values according to a t-test (p=0.).

Consider next the two definitions of overconfidence directly related to our definitions of difficult questions: A student is *ex ante overconfident* if he/she had more incorrect than correct answers to the three ex ante difficult questions. Similarly, a student is *ex post overconfident* if he/she had a negative net result on the six questions that ex post turned out to be difficult. Table 3 reports the shares of overconfident males and females according to the two definitions, respectively.

Table 3. Percent of respondents who behaved overconfidently according to the two overconfidence definitions.

	Ex ante overconfident		Ex post overconfident	
	Mean	Std. dev.	Mean	Std. dev.
Total	39.7	49.1	37.9	48.7
Male	35.2	48.0	39.8	49.2
Female	44.2	50.0	36.0	48.3

Roughly 40 percent of the students were categorized as overconfident according to both definitions, and a larger share of women than men show ex-ante overconfidence. However, the difference is not statistically significant (p=0.23 according to a t-test). Nor are the shares of ex post overconfident men and women significantly different (p=0.61), although women are overconfident to a slightly lower degree.

4. Regression Analysis

Next, we run regressions corresponding to each of the measures reflecting overconfidence. In addition to gender, we control for the results from the longer questions as a proxy for knowledge. We also correct for whether or not the student took the exam as a re-exam, the chosen specialization of the education (analytical versus language-based specialization within the program), and whether or not the student was particularly young, i.e., younger than 24 years. Table A2 in Appendix presents descriptive statistics of the explanatory variables for males and females, respectively. The regression results are presented in Table 4.

For the total number of multiple-choice questions answered, it can be observed that the only significant parameter is the one corresponding to the points on the longer questions, reflecting knowledge, where the parameter as expected is positive. Similarly, in the second OLS regression, where the dependent variable is the number of incorrect answers minus the number of correct answers on the multiple-choice questions, the only significant parameter reflects that the more points a student scored on the longer questions, the better the results in the multiple-choice section.

Table 4. Regression analysis based on different measures reflecting overconfidence (standard errors in parentheses).

	Number of answers, all questions (OLS)	Incorrect minus correct answers, all questions (OLS)	Number of answers *ex ante* difficult questions (OLS)	Incorrect minus correct answers, *ex ante* difficult questions (OLS)	*Ex ante* overconfident (Probit, marginal effects)	Number of answers *ex post* difficult questions (OLS)	Incorrect minus correct answers, *ex post* difficult questions(OLS)	*Ex post*-overconfident (Probit, marginal effects)
Male	0.02 (0.56)	-0.82 (0.69)	0.25* (0.14)	-0.37** (0.17)	-0.12 (0.08)	0.15 (0.23)	0.13 (0.28)	-0.02 (0.08)
Points on longer questions	0.28*** (0.05)	-0.65*** (0.06)	0.01 (0.01)	-0.02* (0.01)	-0.01 (0.01)	0.03 (0.02)	-0.03 (0.02)	-0.01 (0.01)
Took the exam as a reexam	0.12 (0.73)	0.49 (0.91)	0.08 (0.18)	0.16 (0.23)	0.02 (0.10)	0.27(0.30)	0.39 (0.37)	0.16 (0.11)
Analytical specialization	-0.62 (0.57)	0.44 (0.69)	-0.14 (0.14)	0.10 (0.17)	0.03 (0.08)	-0.29 (0.23)	0.65** (0.28)	0.25*** (0.08)
Younger than 24 years	-0.74 (0.57)	0.97 (0.71)	-0.06 (0.14)	0.24 (0.18)	0.02 (0.08)	0.05 (0.23)	0.51* (0.29)	0.09 (0.08)
Constant	20.15*** (1.11)	-5.44*** (1.38)	0.88*** (0.28)	0.56* (0.34)		2.47*** (0.45)	-0.24 (0.57)	
Adj. R^2, pseudo-R^2	0.18	0.45	0.01	0.02	0.02	0.00	0.03	0.06
Number of observations	174	174	174	174	174	174	174	174

Notes: Ex ante overconfident refers to having more wrong than right answers to the three questions that were constructed to be extra difficult. Ex post overconfident refers to having more wrong than right answers to the six questions that less than 50 % answered correctly conditional on answering.

Next, we analyze the behavior concerning the ex ante and ex post difficult questions, respectively.[7] As in the nonparametric tests, we find that while men answered more of the ex ante difficult questions, women had a higher share of incorrect answers. In the probit of ex ante overconfidence, neither gender nor any other explanatory variable turns out significant. For the ex post difficult questions, we again find no significant effect of gender in any of the regressions. We obtain no significant effect of knowledge as measured by the score for the longer questions here. Perhaps this reflects that in a situation where there is a slight general tendency toward overconfidence, those with poorer results found it more obvious that they should not answer these difficult questions, which in turn offsets the effects of knowledge differences in how well they answered these questions given that they answered them.[8] The only significant effects in terms of net results are found for analytical specialization and for being young.[9] In general, the fact that students with the analytical specialization have slightly better scores and that it is harder to be admitted to that specialization than to the language one suggests that students with analytical specialization score higher on knowledge.

---

[7] Given the discrete nature of the data and the fewer possible alternatives one could argue in favor of ordered probit (or logit) instead of OLS models. However, for ease of comparison with the previous regressions, we present the OLS. Moreover, the results in terms of parameter significance are consistently the same based on models with ordered probit instead of OLS.

[8] As a sensitivity analysis, we have altered the definitions for difficulty in order to include more or fewer questions. The main results, however, remain robust to different specifications.

[9] We have no good explanation for this finding, but we can speculate that these students are more used to competitiveness and therefore more prone to being overconfident. There is also some evidence from psychological studies that experts are more likely to be overconfident than more inexperienced subjects (e.g. Heath and Tversky, 1991; Glaser et al. 2004, 2007). However, see also Menkhoff et al. (2006), who did not obtain any clear relation between experience and overconfidence. There is no significant difference in the gender share between the specializations (see Table A2), nor are any interaction effects between gender and specialization significant.

The main finding in this section is that the results from Section 2 are fully reinforced when we correct for other variables through regression analysis: Irrespective of how we measure overconfidence, we find no evidence of a gender difference.

5. Conclusions

Based on a natural field-experiment on exam behavior, this paper fails to identify any gender-based differences in overconfidence, although we measured it in several alternative ways. Unlike several previous studies based on survey analysis, the present study is fully consequential, since the subjects' exam results depended on their choices. Moreover, unlike laboratory experiments on overconfidence, the subjects here did not know that they were part of a scientific study, implying that there is no corresponding potential bias associated with such an awareness; see Levitt and List (2007) for discussions on this and additional advantages of field experiments compared to lab experiments.

Yet, one should of course still be careful when making generalizations based on a single experiment. In particular, our student-based sample is obviously not representative of the Swedish population, and the Swedish population may in turn not be representative of people elsewhere. First, we cannot rule out that women who self-select into business education are different than other women in terms of overconfidence (and perhaps more different than the targeted males are compared to other men in this respect). For example, Croson and Gneezy (2008) conclude that women selecting into, e.g., managerial positions, are not different from men in terms of risk behavior, whereas women tend to be more risk averse in general. Similarly, Nekby et al. (2008) show that women who select into male-dominated fields tend to be more competitively oriented. However, the students in our study are in an early and still very broad stage of their educational program (their third semester), and they have not yet selected a higher-level specialization, which could be more or less competitive. The program itself is rather broad and leads to both highly competitive and not very competitive jobs. Moreover, the program is not male dominated; in fact, the gender distribution is just about even. Overall, we can certainly not rule out potential selection effects through the choice of education, but our conjecture is that such effects are not very large. Second, there are also some indications that gender-based differences might be smaller in Sweden than in other countries.[10]

In conclusion, while the present study calls into question the general stereotype that men are more overconfident than women, we clearly need more research with respect to under which conditions, if any, this stereotype is correct in terms of actual behavior when real incentives are at stake. Further research based on different empirical strategies as well as samples is therefore encouraged.

[10] Based on studies on children in Sweden, Dreber et al. (2009) found smaller gender differences in terms of competitiveness than found in other studies. The Global Gender Gap Report 2008 ranks Sweden as the third most gender equal country in the world. Cesarini et al. (2009) provide a recent attempt, based on a sample of twins, to disentangle how much cultural versus genetic factors contribute to overconfidence.

Appendix

Table A1. Frequency of answers to the multiple-choice questions (in percentage)

	Percentage answering			Percentage answering correctly conditional on answering		
	Total	Male	Female	Total	Male	Female
Q1	99	99	99	98	98	98
Q2	83	83	84	47	45	49
Q3	97	95	99	93	90	95
Q4	99	99	100	95	97	93
Q5	50	49	51	43	30**	55**
Q6	97	97	97	91	89	93
Q7	94	91	97	96	95	96
Q8	94	89***	99***	89	91	87
Q9	50	47	53	71	59***	83***
Q10	23	31**	15**	27	22	38
Q11	94	91	97	96	96	96
Q12	76	76	77	75	73	77
Q13	99	98	100	92	95	90
Q14	81	78	84	72	68	75
Q15	87	87	87	92	91	93
Q16	72	73	71	74	73	74
Q17	98	99	98	97	95	99
Q18	66	74**	58**	43	57***	26***
Q19	25	20	30	48	56	42
Q20	57	51	63	49	51	48
Q21	91	87	94	76	77	75
Q22	95	93	98	96	94	99
Q23	90	89	91	94	96	92
Q24	92	92	92	88	85	90
Q25	91	85***	98***	74	77	71
Q26	94	94	94	86	87	85
Q27	91	90	93	89	87	91
Q28	93	90	95	92	91	93
Q29	84	82	86	90	90	90
Q30	88	88	88	87	92*	82*

Note: *, **, and *** indicate that differences between men and women are significant at the 10, 5, and 1 percent levels, respectively.

The ex ante difficult questions include Q10, Q18, and Q19. When we define questions as difficult ex post, i.e., that the questions are deemed difficult when fewer than 50 percent of the students give a correct answer, we include Q2, Q5, Q9, Q10, Q18, Q19, and Q20. As a sensitivity analysis, we broaden the definition to include the questions that fewer than 60 percent know the correct answer to, and thus also include Q12, Q14, and Q16. We also use a narrower definition and include only the six questions that fewer than 50 percent gave a correct answer to conditional on answering, i.e., Q2, Q5, Q10, Q18, Q19, and Q20.

Table A2 Mean values and shares of explanatory variables divided by gender (standard deviations in parentheses).

	Male	Female
Points on longer questions, out of 30 points possible	16.6*** (6.7)	19.4*** (5.3)
Took the exam as a re-exam	0.26**	0.13**
Analytical specialization	0.55	0.47
Younger than 24 years	0.60*	0.72*
Number of observations	88	86

Notes:
*, **, and *** indicate that differences between men and women are significant at the 10, 5, and 1 percent levels, respectively.

References

Barber, B. M., and T. Odean (2000), "Trading Is Hazardous to Your Wealth: The Common Stock Investment Performance of Individual Investors," *Journal of Finance*, *55*, 773–806.

Barber, B. M., and T. Odean (2001), "Boys will be boys: Gender, overconfidence, and common stock investment", *Quarterly Journal of Economics* 116, 261-292.

Baumeister, R., (1998). The self. In: Gilbert, D., Fiske, S., Lindzey, G. (Eds.), *Handbook of Social Psychology*. McGraw Hill, Boston.

Benabou, R. and Tirole, J. (2002). 'Self-confidence and personal motivation', *Quarterly Journal of Economics*, vol. 117, pp. 871–915.

Bengtsson, C., Persson, M., and P. Willehag (2005), "Gender and overconfidence", *Economics Letters* 86, 199-203.

Bertrand, M., Hallock, K., 2001. The gender gap in top corporate jobs. *Industrial and Labor Relations Review* 55, 3–21.

Biais, B., D. Hilton, and K. Mazurier (2005), "Judgemental Overconfidence, Self-Monitoring, and Trading Performance in an Experimental Financial Market", *Review of Economic Studies* 72(2), 287-312.

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

Cesarini, D., M. Johannesson, P. Lichtenstein and B. Wallace (2009), "Heritability of Overconfidence", *Journal of the European Economic Association*, 7, (2-3), 617-627.

Cesarini D., Sandewall, O. and Johannesson M. (2006). Confidence interval estimation tasks and the economics of overconfidence. *Journal of Economic Behavior and Organization*, 61, 453-470.

Chevalier, A., S. Gibbons, A. Thorpe, M. Snell and S. Hoskins (2009). Students' academic self-perception. *Economics of Education Review* 28, 716–727.

Compte, O. and Postlewaite, A. (2004). 'Confidence-enhanced performance', *American Economic Review,* 94(5), 1536–57.

Croson, R. and U. Gneezy (2008), "Gender differences in preferences", *Journal of Economic Literature* 47(2), 448–74.

De Bondt, W. and R. H. Thaler, (1995), Financial decision-making in markets and firms: a behavioral perspective., in (R. A. Jarrow, V. Maksimovic and W. T. Ziemba, eds), Finance, *Handbooks in Operations Research and Management Science,* vol. 9, pp. 385.410. Amsterdam: North Holland.

Deaves, R., E. Lüders, and G. Y. Luo (2008), "An Experimental Test of the Impact of Overconfidence and Gender on Trading Activity", *Review of Finance*, 1-21.

Deaves, R., E. Lüders, and M. Schröder (2010), The dynamics of overconfidence: Evidence from stock market forecasters. *Journal of Economic Behavior & Organization*, forthcoming.

Dreber, A., von Essen, E., and E. Ranehill (2009) "Outrunning the gender gap – boys and girls compete equally", SSE/EFI Working Paper 709, Stockholm School of Economics.

Gibbons, S., and Silva, O. (2007). Enjoyment, choice and achievement. mimeo: London School of Economics, Centre for Economic Performance.

Glaser, M., T. Langer, and M. Weber (2007), "On the Trend Recognition and Forecasting Ability of Professional Traders", *Decision Analysis* 4(4), 176-193.

Glaser, M., M. Nöth, and M. Weber (2004), "Behavioral Finance", in: Derek J. Koehler and Nigel Harvey (eds.), *Blackwell Handbook of Judgment and Decision Making*, 527-546.

Gneezy, U., Niederle, M., Rustichini, A. (2003) "Performance in Competitive Environments: Gender Differences", *Quarterly Journal of Economics*, 118, 1049-1074.

Gneezy, U., Leonard, K. L and List, J. A. (2009) "Gender Differences in Competition: Evidence from a Matrilineal and a Patriarchal Society", *Econometrica*, 77, 1637-1664.

Grubb, M. D. (2009) "Selling to Overconfident Consumers", *American Economic Review*, 99, 1770–1807.

Harrison, G. and J. List (2004) Field Experiments, *Journal of Economic Literature* 42, 1009-1055.

Heath, C. and A. Tversky (1991), "Preference and belief: Ambiguity and competence in choice under uncertainty", *Journal of Risk and Uncertainty* 4, 5–28.

Hirschleifer, D, Low, A.,Teoh, S. H. (2010), “Are Overconfident CEOs Better Innovators?”, MPRA Paper No. 22425.

Johansson-Stenman, O. (2010) "Risk Aversion and Expected Utility of Consumption Over Time", *Games and Economic Behavior*, 68, 208-219.

Kahneman, D., Slovic, P. and Tversky,A. (Eds.) (1982). *Judgement under uncertainty: Heuristics and biases.* Cambridge: Cambridge University Press.

Köszegi, B., (2006), .Ego Utility, Overconfidence, and Task Choice, *Journal of the European Economic Association*, 4(4), 673-707.

Levitt, Steven D. and John A. List (2007), “What do Laboratory Experiments Measuring Social Preferences Tell Us About the Real World,” *Journal of Economic Perspectives*, 21 (2): 153-174.

Lundeberg, M.A., Fox, P.W., Puncochar, J. (1994), "Highly confident but wrong: gender differences and similarities in confidence judgments", *Journal of Educational Psychology*, Vol. 86 pp.114–21.

Malmendier, U. and G. Tate (2005). “CEO Overconfidence and Corporate Investment.” *Journal of Finance*, 60(6): 2661–2700.

Malmendier, U. and G. Tate (2008). “Who Makes Acquisitions? CEO Overconfidence and the Market’s Reaction.” *Journal of Financial Economics*, 89(1): 20–43.

Menkhoff, L., U. Schmidt and T. Brozynski (2006), “The Impact of Experience on Risk Taking, Overconfidence, and Herding of Fund Managers: Complementary Survey Evidence”, *European Economic Review* 50(7), 1753-1766.

Nekby, L., Skogman Thoursie, P., and L. Vahtrik (2008), “Gender and self-selection into a competitive environment: Are women more overconfident than men?”, *Economics Letters* 100, 405-407.

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

Odean, T. (1999). "Do Investors Trade Too Much?" *American Economic Review*, 89, 1279–1298.

Prince, M. (1993), "Women, Men, and Money Styles," *Journal of Economic Psychology*,14, 175–182.

Rabin, M., (2000). "Risk aversion and expected utility theory: A calibration theorem", *Econometrica* 68, 1281–1292.

Rabin, M., Thaler, R.H., (2001), "Anomalies: risk aversion", *Journal of Economic Perspectives*, 15, 219–232.

Taylor, S.E. and Brown, J.D., (1994). "Positive illusions and well-being revisited: Separating fact from fiction", *Psychological Bulletin* 116, 21-27.

Thorpe, A., Snell, M., Hoskins, S., & Bryant, J. (2007). False uniqueness: The self-perception of new entrants to higher education in the UK and its implication for access—A pilot study. Higher Education Quarterly, 61, 3–22.

Vandegrift, D. and Yavas, A. (2009). Men, women, and competition: An experimental test of behavior. *Journal of Economic Behavior and Organization*, 72(1): 554-570.