A new and direct test of the 'gender bias' in multiple-choice questions

STAN DU PLESSIS AND SOPHIA DU PLESSIS

Stellenbosch Economic Working Papers: 23/09

KEYWORDS: GENDER BIAS, ECONOMICS EDUCATION
JEL: A22

STAN DU PLESSIS
DEPARTMENT OF ECONOMICS
UNIVERSITY OF STELLENBOSCH
PRIVATE BAG X1, 7602
MATIELAND, SOUTH AFRICA
E-MAIL: STAN@SUN.AC.ZA

SOPHIA DU PLESSIS
DEPARTMENT OF ECONOMICS
UNIVERSITY OF STELLENBOSCH
PRIVATE BAG X1, 7602
MATIELAND, SOUTH AFRICA
E-MAIL: SOPHIA@SUN.AC.ZA

UNIVERSITEIT
STELLENBOSCH
UNIVERSITY

A WORKING PAPER OF THE DEPARTMENT OF ECONOMICS AND THE
BUREAU FOR ECONOMIC RESEARCH AT THE UNIVERSITY OF STELLENBOSCH

A new and direct test of the 'gender bias' in multiple-choice questions

STAN DU PLESSIS AND SOPHIA DU PLESSIS[1]

ABSTRACT

Local and international research has identified a bias in favour of male students with MCQs. If correctly identified, this bias holds implications for reasonable assessment strategies in economics courses. A standard method used in the literature is to relate student performance to various features of the learning environment (such as the type of question) and to student-specific characteristics (such as past performance and lecture attendance). A more direct approach is possible: we set comparable questions (in three categories – graphs, quantitative and theory) in the written and MCQ sections of three tests in the introductory microeconomics course at the University of Stellenbosch. This allows a direct comparison between the performance of male and female students (overall and per question category), without the need to model overall student performance. The number of students in this course, almost 2000, offers a suitably large sample for studying this question. Our evidence does not confirm the strong claims about gender bias in the literature; indeed we find the opposite: a strong positive female gender effect, but for written questions only. We also find no evidence of higher risk-aversion by female students towards MCQ questions with negative marking.

Keywords: Gender bias, Economics education
JEL codes: A22

[1] The authors are respectivley, Professor and Senior Lecturer in the Department of Economics, Stellenbosch University. This paper was written for the biennial conference of the Economic Society of South Africa, 10–12 September 2007, Indaba Hotel, Johannesburg. We would like to thank Reta Gelderblom for compiling the data and Corné van Walbeek for very useful comments. The usual disclaimer applied.

A new and direct test of the 'gender bias' in multiple-choice questions

Introduction

In December 2009 Elinor Ostrom will become the first woman Nobel laureate in Economics. That it took 40 years since the first of these prizes were awarded to Frisch and Tinbergen for a woman to win, is at least suggestive of how difficult it has been for woman to advance in the world of economics. And it is possible that women are discouraged at the very outset by the assessment of university level courses in economics. For example, a potential gender bias in the test scores of economics students, especially with multiple-choice questions (MCQs), might dissuade women from pursuing careers in economics. Local and international research has identified a bias in favour of male students in MCQs (for example, Siegfried, 1979; Buck et al., 2002; Leaver and van Walbeek, 2006) which, if correct, holds implications for reasonable assessment strategies in economics courses.

A standard method in this literature is to relate student performance to various features of the learning environment (such as the type of question) and to student-specific characteristics (such as past performance and lecture attendance). In such a model gender bias emerges as an economically and statistically significant gender effect among the control variables (for example, Leaver and van Walbeek, 2006). But in this method identifying gender bias is dependent on the overall quality of the student performance model. A more direct approach is possible: we set comparable questions (in three categories – graphical, quantitative and theoretical) in the written and MCQ sections of three tests in the introductory microeconomics course at the University of Stellenbosch. This allows a direct comparison between the performance of male and female students (overall and per question category), without necessitating any controls for environment or other student characteristics.

1. Gender bias in economic performance

In 1979 John Siegfried published a survey of articles done on the performance of men relative to women in introductory economics courses (Siegfried, 1979). He indicates that two-thirds of the surveyed studies show that men outperform women on measures of economic knowledge. The surveyed studies converge on the conclusion that men are already ahead of women by the time they enter college. The same amount of learning takes place during college, but women never catch up with their male counterparts. Siegfried (1979) supported the thesis that the upbringing

of women may encourage dependence, with evidence showing that women perform better in smaller class groups where they receive more individual attention.

An important focus of the research on gender bias in economics during the 1980s was the particular biases associated with multiple-choice questions. Research showed that gender bias in economics was even more pronounced for multiple-choice questions, than for essay-type questions. Ferber, Birnbaum and Green (1983) and Heath (1989) reported that the gap between men and women in comprehension of economics appeared to be even more pronounced when tested by means of multiple-choice questions.

If men are better at answering multiple-choice questions and women at descriptive questions, then it holds serious consequences for test formats. It raises an important policy question when large classes (often the case in undergraduate economics) encourage the use of multiple-choice questions as a practical measure to manage assessment.

There is some difficulty with this view, per se, because it relies on the acceptance of men having a relative advantage in spatial and numerical skills, and women in verbal skills. If this is true, women should have an advantage in reading a multiple-choice question with sufficient understanding to find the correct answer. Furthermore would this not mean that men are penalised by essay-type questions?

These issues were considered in the 1990s, with extensive research done on the gender effect of multiple-choice questions. Hirschfeld, Moore and Brown (1995) tried to find out why men outperformed women on Graduate Record Exams (GREs), based on data from 1989 to 1992. They concluded that the 'willingness to guess' may be an important factor in generating this bias [i.e. gender bias]. First Hirschfeld et al. (1995) compared the GRE taken at the end of the final college year with a Scholastic Aptitude test (SAT) taken at the beginning of the year. Both tests reward a willingness to guess, but the study also showed that women gained confidence during the year and were more willing to guess at the end of the year than they were at the beginning. Second, they compared the GRE results to the MFAT (Major Field Achievement Test). The latter has no penalty for wrong answers and shows a much smaller gender gap (Hirschfeld et al., 1995). This is an important finding with direct relevance to the study being reported on, which was conducted on multiple-choice questions at the University of Stellenbosch where incorrect answers are penalised.

More recent empirical work has suggested a different picture. Goldin (2006) and Goldin, Katz and Kuziemko (2006) have shown that women have largely reversed the former gap in the United States (as they have elsewhere in the OECD). They found that women are now outperforming their male counterparts at college, even after allowing for factors such as family background, high school performance and performance in mathematics and reading. While girls have long outperformed boys at school, a number of important changes have contributed to what Goldin (2006) has called a 'quiet revolution' in women's role in the labour force and in education. These changes include the following: eliminating gender restrictions in the labour market, a rise in the age of first marriage for women (in which access to contraceptives played an important role) and a higher divorce rate, reducing the security of income offered by marriage. Meanwhile, boys continue to suffer from many behavioural problems at a young age, which contributes to female advantage in college (Goldin et al., 2006: 153-154).

There is not much empirical research on the gender gap in economics teaching among South African students, with the important exception of Leaver and van Walbeek (2006) and Parker (2006). Leaver and van Walbeek (2006) considered the introductory course in microeconomics at the University of Cape Town, and they were specifically interested in the effect of 'question type'[2] on the gender bias. They found that women performed worse than men in all categories of MCQ questions. Parker (2006) studied the performance of students in introductory microeconomics courses across five South African Universities and related this to various student characteristics, including gender. She found a consistently large and significant gender bias gap in favour of male students, ranging from 3.46 % to 7.75 % after controlling for other student characteristics. This study is directly relevant to one of the two possible explanations offered by Parker (2006: 146) for gender bias, i.e., the preponderance of MCQs in introductory microeconomics exams in South Africa. The study did not investigate student expectations of the role of economics in their future careers – Parker's second explanation for the gender gap.

2. Data and method

The data used here were collected during the first semester of 2007 in three tests for the Economics 178 course at the University of Stellenbosch. In the first test (8 March 2007) two questions in the multiple-choice part of the paper were matched with two questions in the written part of the test. Three question pairs were matched in the second test (14 April 2007) and five

[2] Based on Bloom's taxonomy and the content type of the question.

question pairs in the third test (1 June 2007). Tables 1, 2 and 3 show the relevant question pairs and indicate their category in the third column.

Table 1 about here

Table 2 about here

Table 3 about here

In total the experiment consisted of ten question pairs split into three categories: graphical, quantitative and theoretical. Table 4 shows the distribution of the question pairs across these categories.

Table 4 about here

Students are not required to write all the tests in Economics 178 but have to sit a minimum of four out of a possible five per year. Table 5 shows the number of students who elected to write the three tests included in this study and for whom we have the requisite data. The tests under consideration here were for the microeconomics section of Economics 178. The remaining tests were on macroeconomics.

Table 5 about here

Written questions and MCQs are marked on different scales, requiring a data transformation to make the results comparable. In this course at Stellenbosch the answer to an MCQs is scored on a three-point scale: -2/3 for an incorrect answer, 0 for skipping the question and 2 for the correct answer. The written questions had different totals and students could earn marks on a discrete scale with intervals of 0.5 starting at 0 and ranging to 2, 3 or 4, depending on the question.

We adopted a simple strategy to make the results comparable: for the MCQs we grouped incorrect answers and skipped questions together in category zero, leaving correct answers in

category 1[3]. Table 6 summarises the aggregation that was done for each of the written questions to divide their answers into the same 0 and 1 categories. If a mark of more than half of the potential mark was allocated, the answer was divided into the 1 category, and if less than half of the potential mark, the 0 category. Henceforth all answers in the zero category will be considered "incorrect" and all answers in the 1 category will be considered "correct".

Table 6 about here

Following the data transformation, the results can be aggregated across questions and grouped according to the three question types: graphical, quantitative and theoretical.

3. Results

Table 7 shows summary statistics for the test totals as well as for the multiple-choice and written question subsections.

Table 7 about here

Measured by the median, male students did better on average than female students in the first test but worse in the last two. However, using the mean overturns the gender ranking for the first test.

In the multiple-choice questions, male students did somewhat better in Tests 2 and 3 if the median was taken as a summary statistic, but using the mean narrows the gap for Test 3 and overturns the result for Test 2. Male students did better in the MCQs in Test 1 using the mean as summary statistic, though their median score was the same as for the female students. In the written questions male students did worse in each test when the mean was used as a summary statistic, but they only did worse in Test 3 if the median was used instead. However, the out-performance of female students on written questions in Test 3 (3.7 % on the median) is the largest gender gap for the summary statistic in Table 7.

The purpose of this study was not to investigate gender effects in test scores generally, but to investigate the following narrower hypothesis: the existence of a gender bias in multiple-choice question results relative to written exam results when equivalent questions are posed in both test

[3] In section 5 we consider the difference between incorrect and skipped answers in an investigation of risk aversion.

types. To test this hypothesis we used the ten matched pairs of questions described above. The results of this comparison are grouped by question type.

3.1 Graphical questions

Two question pairs on Test 1 and a further two question pairs on Test 3 were explicitly about graphs. Table 8 shows the proportion of students with correct answers for each of these questions and for the four questions combined. We used bold text to indicate the largest proportion in each column.

Table 8 about here

Women did better in all four written versions of the questions involving graphs, while the men did better in two of the MCQ versions. Aggregating across all four questions the women did better in the written versions of the questions and the men in the MCQ versions, but the gaps are not equally significant. With the large sample size at our disposals a two-sample z-test for the equality of the proportions is appropriate in both the written and MCQ cases[4]. Woman students outperformed men by five percentage points in the written versions of the questions on graphs, a gap that is both large and significant at levels of significance below 1%. By contrast the 1% gap between the men and women on the MCQ versions of these question is small and only significant at a 19% level of significance.

In summary, we find a gender effect for questions with graphs but not the effect mentioned in the literature. Our results suggest little gender effect on multiple-choice questions, but a positive gender effect for female students on the equivalent written questions.

3.2 Theory questions

One question pair in the second test and two question pairs in the third test concerned economic theory. Table 9 shows the proportion of students with correct answers for each of these questions and for the three questions combined. We used bold text to indicate the largest proportion in each column.

Table 9 about here

[4] It is easily demonstrated that the normal distribution can be used for the z-test and present sample: $n_m p_m = 2240 > 5$; $n_m(1-p_m) = 1850 > 5$; $n_f p_f = 2137 > 5$; $n_f(1-p_f) = 1463 > 5$.

Women did better than men in the theory questions in both the written and the MCQ versions of the same questions. This is true when aggregating the results of the three relevant questions, and is true for all three questions individually where the written questions are concerned. It is also true for two of the MCQ questions. At 5.8% on the written questions and 2.2% on the MCQ questions the performance gaps between the men and women students are not just statistically significant (at significance levels below 1%) but also large in economic terms.

In summary, we find a positive gender effect for female students on theory questions, and the effect is much stronger for the written questions.

3.3 Quantitative questions

Two questions in the second test and one in the final test were explicitly quantitative. Table 10 shows the proportion of students with correct answers for each of these questions and for the three questions combined. We used bold text to indicate the largest proportion in each column.

Table 10 about here

For the group of quantitative questions the results conformed most nearly to the gender effects observed elsewhere in the literature. Male students did consistently better in each of the MCQ versions of the same question and the gap between their total proportion of correct answers (63.1%) and that of their female peers (60.5%) is both large and statistically significant. The gap between the 46.6% correct answers by women for the written versions of these questions and the 41.6% for the men is even larger (and of course statistically significant). In summary, we find large and significant gender effects for the quantitative questions.

3.4 Combined result

The proportions of correct answers for all ten questions are shown in table 11.

Table 11 about here

Aggregating across all the questions female students outperformed their male counterparts by 5% on the written questions, a gap that is both large and statistically significant. By contrast the male students did better in the MCQ versions of the same questions but by a small margin (0.7%) and one that is not statistically significant even at a 30% level of significance.

4. Risk aversion

The risk attached to guessing in MCQs given the marking scheme at Stellenbosch introduces another aspect to the data reported above, that is the possibility that the results might be affected by different levels of risk aversion in male and female students. Comparing the outcomes of the MCQ and written versions of the same question, therefore, compares not just knowledge and understanding of economics but also risk aversion. We consider that possibility in this section.

Starting with the premise that a student that scores "incorrect" on the written version of a test does not know the answer to the question we categorized students who answered the matching MCQ question as guessing. The proportion of male and female student guessing MCQ questions in this sense is shown in table 12.

Table 12 about here

Apparently 84.2% of the male students who did not get the written version of given question correct nevertheless guessed the answer to the matching MCQ. While a lower proportion of female students guessed, the difference is in fact trivial in size and would only be statistically significant at a 55% level of significance. This experiment yields no evidence of different levels of risk aversion between the male and female students. But this does not mean that the male and female students guessed equally well, an issue informed by the proportions of correct guesses reported in table 13[5].

Table 13 about here

Though approximately the same proportion of male and female students were willing to answer the MCQ version of a question they answered incorrectly in the written part of the test, the two groups did not guess with equal accuracy. Men guessed better, at 65.4% of all guesses, which is 2.6% better than the accuracy of their female peers, and large and statistically significant difference.

[5] These proportions are the ratios of correct answers to the total number of guesses in all the MCQ questions by male and female students respectively.

5. Conclusion

This paper used an experiment of matched pairs of questions to investigate the suspected adverse gender effect of multiple-choice questions for female students. We failed to confirm the results found by modelling student performance in the literature. However, the positive results are more remarkable: we find strong evidence of a positive female gender effect on the written questions that had been paired with their multiple-choice equivalents. While there were no evidence of differences in risk aversion in this experiment, the result record a large and significant gap between the success with which male and female students guess MCQ versions of questions they had difficulty answering in written form. Male students guessed with significantly more success under these circumstances.

References

Buck, G., I. Kostin, et al. (2002). Examining the relationship of content to gender-based performance differences in advanced placement exams. College Board Report no 2002-12.

Ferber, M. A., B. G. Birnbaum, et al. (1983). "Gender differences in economic knowledge: a reevaluation of the evidence." Journal of Economic Education **14**(2): 24-37.

Goldin, C. (2006). "The Quiet Revolution That Transformed Women's Employment, Education, and Family." American Economic Review (Papers and Proceedings) **96**(2): 1-21.

Goldin, C., L. F. Katz, et al. (2006). "The Homecoming of American College Women: The Reversal of the College Gender Gap." Journal of Economic Perspectives **4**(Fall): 133-156.

Heath, J. A. (1989). "An econometric model of the role of gender in economic education." American Economic Review **79**(2): 226-230.

Hirschfeld, M., R. L. Moore, et al. (1995). "Exploring the gender gap on the GRE subject test in Economics." Journal of Economic Education **26**(1): 3-15.

Leaver, R. and C. van Walbeek (2006). "Gender Bias" in multiple choice questions: does the type of question make a difference? Cape Town, University of Cape Town, mimeograph.

Parker, K. (2006). "The Effect of Student Characteristics on Achievement in Introductory Microeconomics In South Africa." South African Journal of Economics **74**(1): 137-149.

Siegfried, J. J. (1979). "Male-female differences in economic education: a survey." Journal of Economic Education(Spring): 1-11.

Table 1 *Question pairs for Test 1*

Multiple-choice questions	*Questions in the written section*	*Question category*
Section A, question 8 Suppose a pest destroys much of the maize crop in the Free State. At the same time, suppose consumers' tastes shift from another product towards maize products. What would happen to the equilibrium price and quantity in the market for maize after both effects have been taken into account? a. Price will increase; quantity will increase b. Price will increase; quantity will decrease c. Price will decrease; quantity is ambiguous d. Price will increase; quantity is ambiguous **[2 marks, with a penalty of 2/3 of a mark for an incorrect answer]**	*Section B, question 2* Suppose the market for soccer memorabilia in South Africa is in equilibrium. Explain, with the aid of a graph, what would happen to the equilibrium price and quantity of soccer memorabilia if many soccer tourists come to South Africa, ceteris paribus. **[3 marks]**	Graphical
Section A, question 10 Figure 1 about here Refer to the given graph. Suppose that the price of a pair of socks falls from R5 to R2. The substitution effect is represented by the movement from point: a. Y to point X b. X to point Y c. Z to point X d. X to point Z **[2 marks, with a penalty of 2/3 of a mark for an incorrect answer]**	*Section B, question 3* Vladimir has the choice of spending his income on two normal products, milk and honey. Suppose Vladimir currently consumes at his optimum consumption point. Explain, with the aid of a graph, what effect a pay rise will have on his consumption bundle. **[3 marks]**	Graphical

Table 2 *Question pairs for Test 2*

Multiple-choice questions	*Questions in the written section*	*Question category*
Section A, question 11 Suppose that 50 cans of cool drink are demanded at a particular price. If the price of cool drink rises by 4 %, the quantity falls to 46. This means that the: a. Demand for cans of cool drink for this price increase is inelastic b. Price elasticity of demand for cans of cool drink is 0 c. Demand for cans of cool drink in this price range is elastic d. Demand for cans of cool drink is unit elastic **[2 marks, with a penalty of 2/3 of a mark for an incorrect answer]**	*Section B, question 5* The manager of Cartier watches appoints you as an economist. They would like to know if it is worthwhile lowering the price of their expensive watches to broaden the clientele base. You find the graph below forgotten in a drawer. Given this information, explain thoroughly what you will tell the manager by making use of the price elasticity of demand. Show all your calculations. Figure 2 about here **[4 marks]**	Quantitative
Section A, question 12 Which one of the following statements on consumer surplus is true? a. Consumer surplus shows the value producers attach to a product, given their costs b. Consumer surplus will always increase when the market price for a product increases c. Consumer surplus will decrease when a tax is imposed on sellers d. Consumer surplus will always increase when the amount paid by the buyers increases. **[2 marks, with a penalty of 2/3 of a mark for an incorrect answer]**	*Section B, question 3* Define the marginal rate of substitution between two goods. Briefly explain what causes this rate to diminish as we move downwards along an indifference curve. **[2 marks]**	Theoretical

Table 2 (continued) *Question pairs for Test 2*

Multiple-choice questions	*Questions in the written section*	*Question category*
Section A, question 14 Zach borrowed R500 000 from the bank and used it to start a flower shop. The interest rate was 4 % per year. During the first year of his business, Zach sold 12 000 roses for R3 each. Also during the first year, the business incurred costs that required outlays of money amounting to R14 000. Zach's economic profit for the year was: a. - R478 000 b. - R56 000 c. R22 000 d. R2 000 **[2 marks, with a penalty of 2/3 of a mark for an incorrect answer]**	*Section B, question 8.2* Given the following total cost function: $TC = 7 + 4q^2$ Suppose that the marginal cost (MC) currently amounts to R32. How many units are produced? Show all your calculations. **[3 marks]**	Quantitative

Table 3 Question pairs for Test 3

Multiple-choice questions	*Questions in the written section*	*Question category*
Section A, question 12 Suppose the price elasticity of demand for rugby balls is 1.20. A 15 % increase in price will result in: a. An 18 % decrease in the quantity of rugby balls demanded b. A 15 % decrease in the quantity of rugby balls demanded c. An 8 % reduction in the number of rugby balls demanded d. A 12.5 % reduction in the number of rugby balls demanded **[2 marks, with a penalty of 2/3 of a mark for an incorrect answer]**	*Section B, question 3.1* Refer to the given table of instant coffee and roasted coffee beans. The retail prices and the quantities sold at each price for a given time period are indicated. <table><tr><th>Jaar/ *Year*</th><th>Prys van produk A / *Price of product A*</th><th>Hoeveelheid verkoop / *Quantity sold*</th><th>Prys van produk B / *Price of product B*</th><th>Hoeveelheid verkoop / *Quantity sold*</th></tr><tr><td>2000</td><td>12</td><td>7</td><td>5</td><td>820</td></tr><tr><td>2001</td><td>11</td><td>10</td><td>4</td><td>850</td></tr></table> Calculate the price-elasticity of demand for both producers **[3 marks]**	Quantitative

Table 3 (continued) *Question pairs for Test 3*

Multiple-choice questions	*Questions in the written section*	*Question category*
Section A, question 14 In the figure shown, assume demand increases and as a result, the new equilibrium price is R22 and equilibrium quantity 110. The increases in producer surplus to producers already in the market would be equal to: Figure 3 a. R90 b. R210 c. R480 d. R570 **[2 marks, with a penalty of 2/3 of a mark for an incorrect answer]**	*Section B, question 4.4* The graph below represents a perfectly competitive market: Figure 4 about here Assume the initial equilibrium price (P2) holds and the equilibrium price decreases to P1 as a consequence of a decrease in demand. What is the impact on producer surplus? Explain. **[1 mark]**	Graphical
Section A, question 16 The efficient scale of a firm is the quantity of output that: a. Maximises marginal product b. Maximises profit c. Minimises average total cost d. Minimises average variable cost **[2 marks, with a penalty of 2/3 of a mark for an incorrect answer]**	*Section B, question 5.2* Give two reasons why the long-run industry supply curve may slope upwards. **[2 marks]**	Theoretical
Section A, question 17 When price is greater than marginal cost for a firm in a perfectly competitive market: a. Marginal cost is falling b. There are opportunities to increase profit by increasing production c. The firm is minimising its losses d. The firm could decrease output to maximise profit **[2 marks, with a penalty of 2/3 of a mark for an incorrect answer]**	*Section B, question 5.1* Compare the long-run equilibrium position of a perfectly competitive market with that of a monopoly in terms of price, quantity and two measures of efficiency. No graph is required. **[4 marks]**	Graphical

Table 3 (continued) *Question pairs for Test 3*

Multiple-choice questions	*Questions in the written section*	*Question category*
Section A, question 23 Public goods are: a. Rivalrous in consumption and excludable b. Non-rivalrous in consumption and excludable c. Non-rivalrous in consumption and non-excludable d. Rivalrous in consumption and excludable **[2 marks, with a penalty of 2/3 of a mark for an incorrect answer]**	*Section B, question 8.1* A group of citizens living near to an international airport value peace and quiet. Airplane noise is a disturbance to the citizens. 8.1 Which type of market failure arises here? **[1 mark]**	Theoretical

Table 4 Distribution of questions in three categories

Theory		*Graph*		*Quantitative*	
Multiple-choice question	*Written question*	*Multiple-choice question*	*Written question*	*Multiple-choice question*	*Written question*
T2, Section A, 12	T2, Section B, 3	T1, Section A, 8	T1, Section B, 2	T2, Section A, 11	T2, Section B, 5
T3, Section A, 16	T3, Section B, 5.2	T1, Section A, 10	T1, Section B, 3	T2, Section A, 14	T2, Section B, 8.2
T3, Section A, 23	T3, Section B, 8.1	T3, Section A, 14	T3, Section B, 4.4	T3, Section A, 12	T3, Section B, 3.1
		T3, Section A, 17	T3, Section B, 5.1		

Table 5 Number of students with test data

Test	Male	Female	Total
Test 1	1068	928	1996
Test 2	1033	904	1937
Test 3	977	872	1849

Table 6 Adjustments to written question bins

Question	*Original number of potential bins*	*Adjustments*
T2, Section B, 3	5	• 0 to 1 become 0 • 1.5 & 2 become 1
T3, Section B, 5.2	5	• 0 to 1 become 0 • 1.5 & 2 become 1
T3, Section B, 8.1	3	• 0 to 0.5 become 1 • 1 stays 1
T1, Section B, 2	7	• 0 to 1.5 become 0 • 2 to 3 become 1
T1, Section B, 3	7	• 0 to 1.5 become 0 • 2 to 3 become 1
T3, Section B, 4.4	3	• 0 to 0.5 become 1 • 1 stays 1
T3, Section B, 5.1	9	• 0 to 2 become 0 • 2.5 to 4 become 1
T2, Section B, 5	9	• 0 to 2 become 0 • 2.5 to 4 become 1
T2, Section B, 8.2	5	• 0 to 1.5 become 0 • 2 to 3 become 1
T3, Section B, 3.1	7	• 0 to 1.5 become 0 • 2 to 3 become 1
Multiple-choice questions	3	• -⅔ becomes 0 • 0 becomes 0 • 2 becomes 1

Table 7 *Descriptive statistics for the three tests*

	Median	Mean	Std. deviation	Skewness	Kurtosis[a]
			Test 1		
Total	**76**	**74.5**	**16.97**	**-0.7034**	**3.2**
Male	78	74.4	17.3	-0.7874	3.4
Female	76	74.6	16.5	-0.5896	2.83
Multiple-choice	**73.3**	**70.8**	**25.68**	**-.64029**	**2.8**
Male	73.3	71.5	25.8	-0.6996	2.9
Female	73.3	70	25.5	-0.5735	2.7
Written	**85**	**79**	**19.3**	**-1.2144**	**4.35**
Male	85	77.7	20.51	-1.2197	4.2
Female	85	80.5	17.67	-1.1163	4.1
			Test 2		
Total	**55.5**	**55.1**	**17.43**	**0.0014**	**2.5**
Male	54.5	54.7	17.48	-0.0584	2.5
Female	55.5	55.7	17.37	0.0724	2.5
Multiple-choice	**63.3**	**60.9**	**23.34**	**-0.3296**	**2.6**
Male	63.3	60.6	23.89	-0.36521	2.6
Female	61.7	61.2	22.71	-0.27748	2.6
Written	**51.7**	**52.4**	**17.04**	**0.11843**	**2.6**
Male	51.7	52.1	16.97	0.04506	2.65
Female	51.7	52.8	17.12	0.19931	2.57
			Test 3		
Total	**43.1**	**44.9**	**16.91**	**0.21738**	**2.7**
Male	43.1	44.7	16.91	0.21738	2.7
Female	43.8	45.4	18.37	0.33717	2.6
Multiple-choice	**48.9**	**49.9**	**21.79**	**-0.04115**	**2.6**
Male	51.1	50.7	21.18	-0.078179	2.7
Female	48.9	49	22.43	0.007131	2.4
Written	40	42.2	17.66	0.44974	2.7
Male	38.8	40.53	16.8	0.436049	2.8
Female	42.5	44.1	18.4	0.420051	2.6

[a] A positive value indicates a leptokurtic distribution, i.e., with more weight around the mean and fatter tails than the normal distribution.

Table 8 *Proportions of correct answers for questions involving graphs*

Gender	*Question 1*		*Question 2*		*Question 3*		*Question 4*		*Total*	
	MCQ	*Written*	*MCQ*	*Written*	*MCQ*	*Written*	*MCQ*	*Written*	*MCQ*	*Written*
Male	**91.4%**	75.7%	65.9%	78.7%	55.7%	44.3%	**57.1%**	16.2%	**68%**	54.8%
Female	90.4%	**81.9%**	**69%**	**83.1%**	**56.5%**	**45.1%**	48.9%	**24.4%**	67%	**59.4%**

Table 9 *Proportions of correct answers for theory questions*

Gender	*Question 1*		*Question 2*		*Question 3*		*Total*	
	MCQ	*Written*	*MCQ*	*Written*	*MCQ*	*Written*	*MCQ*	*Written*
Male	**67.3%**	32.2%	36.1%	4.4%	72%	45.6%	58.6%	27.2%
Female	64.5%	**34.2%**	**42.9%**	**7.6%**	**75%**	**57.1%**	**60.8%**	**33%**

Table 10 *Proportions of correct answers for quantitative questions*

Gender	*Question 1*		*Question 2*		*Question 3*		*Total*	
	MCQ	*Written*	*MCQ*	*Written*	*MCQ*	*Written*	*MCQ*	*Written*
Male	**67.5%**	**52.9%**	**58.8%**	36.8%	**63.1%**	34.9%	**63.1%**	41.6%
Female	61.3%	49.6%	58.5%	**48.3%**	61.8%	**41.7%**	60.5%	**46.6%**

Table 11 *Proportions of correct answers for all questions*

Gender	Aggregate of all questions	
	MCQ	*Written*
Male	**63.8%**	42.7%
Female	63.1%	**47.7%**

Table 12 *Proportions of students guessing in MCQs*

Gender	*Proportion guessing*
Male	84.2%
Female	83.8%

Table 13 *Proportions of students guessing correctly in MCQs*

Gender	*Proportion guessing*
Male	65.4%
Female	62.8%

Figure 1

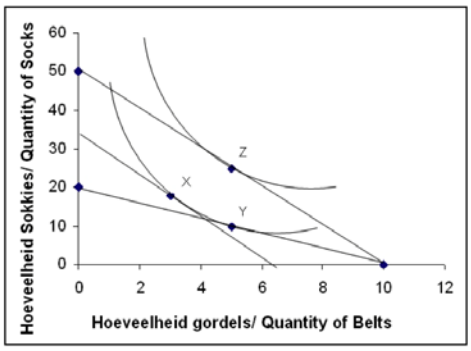

Figure 2

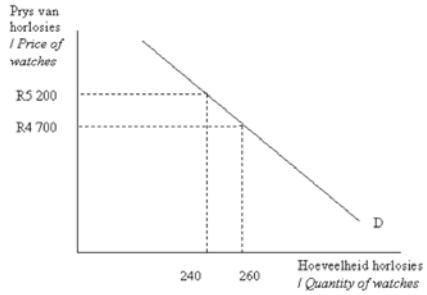

Figure 3

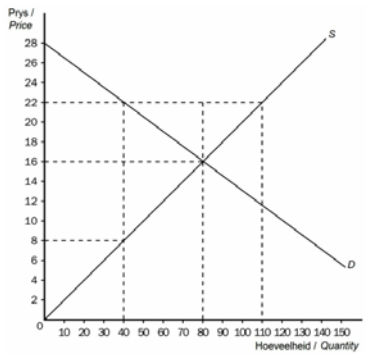

Figure 4