Do Men and Women Perform Differently on Different Types of Test Questions?

Alexi S. Thompson
Abigail L. Jager
Robert O. Burton, Jr.
Senior authorship is share equally by all three authors.

Department of Ag. Econ.
Waters Hall
Kansas State University
Manhattan, KS 66506
E-mail: alexit@ksu.edu
Telephone: 334-707-0730

Department of Statistics
Dickens Hall
Kansas State University
Manhattan, KS 66506
E-mail: jager@ksu.edu
Telephone: 785-532-0514

Department of Ag. Econ.
Waters Hall
Kansas State University
Manhattan, KS 66506
E-mail: bburton@agecon.ksu.edu
Telephone: 785-532-4436

Selected paper prepared for presentation at the Southern Agricultural Economics Association Annual Meeting, Birmingham, AL, February 4-7, 2012.

Contribution No. 12-285-A from the Kansas Agricultural Experiment Station.

Do Men and Women Perform Differently on Different Types of Test Questions?

Introduction

The purpose of this study is to determine whether men and women students perform differently on different types of test questions. Data were collected from the four tests that were administered to Kansas State University students enrolled in AGEC 500 during Spring semester 2011. AGEC 500 is the first course in a two-course sequence that covers intermediate level microeconomic theory. It is a required course for all Agricultural Economics and Agribusiness majors and minors. AGEC 500 is focused on production, consumption, and markets in a perfectly competitive environment. If men and women perform differently on different types of tests questions, then tests may need to be changed to be fair to both genders.

Literature Review

Consequences of the current economic crisis is tighter budgets for universities potentially resulting in larger classes. With a large number of students enrolled in introductory economics courses, instructors may be inclined to give students multiple choice exams to save time grading exams. However, some studies find that multiple choice exams are not as effective in testing economic knowledge as open ended or discussion questions (Becker and Johnston 1999). In addition to effectiveness, many studies indicate a gender gap with respect to type of exam question. Siegfried (1979) surveyed the literature and generally found that men outperform women on multiple choice exam questions and women scored higher on essay questions. One common explanation for this gender gap is that women possess greater verbal skills than men and men have better quantitative skills than women (Lumsden and Scott, 1987).

Lumsden and Scott (1987) sampled over 3,000 students from 17 universities in the United Kingdom. Exams consisted of 20 multiple choice and one micro essay and macro essay question. They controlled for several student characteristics including high school qualifications and study habits. The

authors find that men outperform women on multiple choice questions and women outperform men on essay questions.

Different hypotheses for this gender gap have been proposed. Horvath, Beaudin, and Wright (1992) find that when controlling for grades in an introductory economics course, women need more encouragement in continuing their economics courses than men. This conclusion is supported by a number of other studies (Wright 1989). Heath (1989) suggests that the economic gender gap may be understated because of self-selection biases.

Ferber, Birnhaum, and Green (1983) administered tests to 589 students in an introductory economics course at the University of Illinois-Champagne. For each multiple choice question on the exam, the authors prepared a comparable essay question. Results indicate that men perform better on multiple choice questions and there is no statistical significance between genders on essay questions.

Williams, Waldauer, and Duggal (1992) collected data from introductory and upper level economics courses at Widener University. They categorize test questions based on type (multiple choice, essay, etc.) and student skill set (graphical ability, verbal ability, etc.) Results indicate that, contrary to previous studies, men outperformed women on essay questions in principles courses but women outscored men on essay questions in statistics courses.

Chan and Kennedy (2002) administered exams composed of multiple choice and essay questions to an introductory macroeconomics course. Multiple choice and essay questions are paired so that they cover the same basic topic. The authors do not find a gender gap.

The current paper goes further than previous studies by including several test question categories and by focusing on an upper level course. In addition to multiple choice and essay questions, short answer, graphical, and true/false questions were included. Questions were also categorized by content, thereby controlling for differences between men and women in their comprehension of economic concepts. Tests were administered to students at Kansas State University in the class, AGEC 500, Production Economics. Material covered in class is similar to material one may expect in an

Intermediate Microeconomics course in a typical university so that instructors may use results from this paper to construct a fairer exam.

Procedures

Data for the four tests were combined, using each test question for each student as an observation. To control for differences in students, students completed a background questionnaire that included demographics, farm experience, parental income, number of college level writing courses, and hours of reading per week. Then data were collected from four tests with test questions categorized according to type of question and category of information. Fifty-four students were enrolled; makeup tests were not included.

The dependent variable is the proportion of points received for each question; i.e., the student's score on each question ranging from zero to 100%. There were 6,858 total observations, of which 5,299 were used because 1,559 observations had missing data (mostly on parental income levels). A mathematical summary of the model follows.

(i = 1, . . . , 54 students)

(j = 1, . . . , 4 tests)

($k = 1, \ldots, K_j$)

where:

K_j = the number of questions on test j.

$$Y_{ijk} = \beta_0 + \beta_1^T z_{ijk} + \tau_{ijk} + \gamma_i + (\tau\gamma)_{ijk} + u_i + w_j + e_{ijk}$$

where:

- Y_{ijk} = proportion of points received on each test question
- z_{ijk} = vector of covariates (category of information, major, class, parental income, number of college level writing courses, hours of reading per week, and whether the student grew up with or without farm experience)
- τ_{ijk} = main effect for question type (true/false, multiple choice, short answer, essay and graphical)

γ_i = main effect for gender

$(\tau\gamma)_{ijk}$ = interaction between gender and question type

u_i = random effect for student

w_j = random effect for test

e_{ijk} = random error

Proc mixed in SAS was used to estimate the above mixed model that includes fixed effects for gender, type of question (including the five types of questions most commonly found on the four tests, i.e., true/false, multiple choice, short answer, essay, and graphical) and the interaction between gender and type of question. For an explanation of mixed models see Littell, et al., 2006. Fixed effects also included other factors that may have affected student scores on the tests such as major, class, parental income, number of writing courses taken at the college level, hours of reading per week, category of information for each test question, (i.e., understanding economic concepts, memory and calculations for linear functions, precision in language, and formulas and calculations), and situation in which the student grew up (i.e., farm experience or no farm experience). Random effects that allow for correlation of responses within students and tests were also included.

Results

The interaction term between gender and type of question is significant (*p*-value = 0.0328). This indicates that the effect of gender on the average proportion of points received on test questions varies among question type. Table 1 shows the least square means (with standard errors in parentheses) for each question type and gender. Because the interaction term between gender and type of question is significant, at least one of the differences between men and women in Table 1 is significant. The short answer type of question is most likely to be significant because, of the five types of questions, it has the largest difference between the least square means for men and women. In fact, the difference between men and women responses for short answer questions was significant (p-value = 0.0091).

Table 1. Least Square Means and Standard Errors for Interaction Between Gender and Type of Question.

Question Type	Men		Women	
	Least Square Mean	Standard Error[a]	Least Square Mean	Standard Error[a]
True/False	0.9429	(0.02423)	0.9631	(0.03402)
Multiple Choice	0.9579	(0.02564)	0.9633	(0.03587)
Short Answer	0.8540	(0.0310)	0.9763	(0.04279)
Essay	0.8339	(0.03037)	0.8563	(0.04233)
Graphical	0.9723	(0.02722)	0.9654	(0.03777)

[a]Standard errors in parenthesis.

There are three caveats concerning the analysis. First, the model is based on the assumption that the students are independent. We suspect that students are not independent, because they are all in the same class. We plan to add two classes to the analysis and we plan to adjust for student dependence when we add the two additional classes. Second, the model is based on the assumption that student responses to the test questions have a normal distribution. We suspect that student responses are not normally distributed because most students did well on the tests. For example, responses had large clusters around 1 (for receiving full credit for a test question), Small clusters around 0 (for receiving no credit for a test question), and scattered observations with responses between 0 and 1 (for partial credit). Third, the graphical questions were not evenly distributed among categories of information. Two of the four categories were not represented. So, the graphical part of the analysis is potentially unstable.

Conclusions

The data and model for this study suggest that women perform better than men on four of the five types of questions modeled. However, the differences between men and women responses are small. The largest difference between men and women responses is for short answer questions; and the differences were significant.

Implications

Selection bias is not addressed. However, selection bias as addressed by Heath (1989) has to do with applicability of results to the whole population of students. Whereas, the current study is focused on fairness in testing for an upper level economics course that is required for all majors and minors in the Department of Agricultural Economics at Kansas State University.

Because the differences between men and women responses on the five types of questions modeled are small, this study does not suggest that changes are needed in testing, for AGEC 500 at Kansas State University, as taught and tested by the instructor of the course that was modeled, to achieve fairness in testing of men and women. A possible weakness of this study is that only the five most common types of test questions were modeled and, therefore, several additional types of questions were not modeled. Also, robustness of results would likely be enhanced by adding courses with different instructors. In a future analysis, the authors of this study plan to add two courses with different instructors.

References

Becker, W. E., and C. Johnston. 1999. "The Relationship between Multiple Choice and Essay Response Questions in Assessing Economics Understanding," *Economic Record* 75(Winter): 348-57.

Chan, N. and P. Kennedy. 2002. "Are Multiple-Choice Exams Easier for Economics Students? A Comparison of Multiple-Choice and 'Equivalent' Constructed-Response Exam Questions." *Southern Economic Journal,* 68(Spring): 957-971.

Ferber, M.A., B.G. Birnbaum, and C.A. Green. 1983. "Gender Differences in Economic Knowledge: A Re-evaluation of the Evidence." *The Journal of Economic Education, 14(Spring): 24-37.*

Heath, J.A. 1989. "An Econometric Model of the Role of Gender in Economic Education." *The American Economic Review*," 79(Spring): 226-230.

Horvath, J., B.Q. Beaudin, and S.P. Wright. 1992. "Persisting in the Introductory Economics Course: An Exploration of Gender Differences." *The Journal of Economic Education*. 23(Spring): 101-108.

Littell, R.C., G.A. Milliken, W.W. Stroup, R.D. Wolfinger, and O. Schabenberger. 2006. SAS for Mixed Models. Second Edition. SAS Publishing.

Lumsden K.G. and A. Scott. 1987. "The Economics Student Reexamined: Male-Female Differences in Comprehension." *Journal of Economic Education*, 19(Fall): 365-75.

Siegfried, J. 1979. "Male-Female Differences in Economic Education: A Survey," *Journal of Economic Education*. 10(Spring): 1-11.

Williams, M.L., C. Waldauer, and V.G. Duggal. 1992. "Gender Differences in Economic Knowledge: An Extension of the Analysis," *Journal of Economic Education*, 23(Summer): 219-231.

Wright, S.P. 1989. "Selected Correlates of Intellectual Development of College Freshmen. Unpublished Doctorate Thesis. University of Connecticut, Storrs.