

Proctored versus Un-proctored Testing: Does Cheating Really Help?

Mohammad Ashraf -- The University of North Carolina at Pembroke

Abstract

Significant resources are expended on proctoring tests in schools, colleges and universities alike. In this study we compare the averages of quiz scores of seventy-six proctored and un-proctored quizzes given to various sections of principles of economics classes from the Fall semester of 1999 through the Spring semester of 2003. The results indicate that the averages of quiz scores are not statistically significantly different whether quizzes are proctored or not.

Introduction and Overview

Usually professors prefer to proctor students while students are taking tests. The underlying belief is that students might cheat if they are not being watched. Several studies look at the probability of students cheating and assess the ways to prevent cheating (Bunn, Caudill, and Gropper, 1992; Kerkvliet, 1994; kerkvliet and Sigmund, 1999). Nowell and Laufer (1997) report that according to the American Council on Higher Education, cheating was escalating during 1990.

The difficulty in assessing the extent of cheating and prevention further fuels the interest in this topic. Researchers have devised various ways to gain insight into this behavior. These techniques range from self grading by students after issuing moral appeals and sanction threats (Tittle and Rowe, 1973), to analyzing survey data using random response questionnaires using an extension of the Becker's (1968) model which compares the expected benefits and costs of crime (Kerkvliet, 1994).

In this study we try to overcome the difficult task of assessing the extent of cheating by giving quizzes, which are similar in content and rigor, under two different environments; from the Fall semester of 1999 through the Spring semester of 2002 students were given in-class quizzes that were proctored by the instructor. In the second set of quizzes, which spanned over the Fall semester of 2002 through the Spring semester of 2003, students were given quizzes of similar content and rigor, but in this case the quizzes were not proctored. Students were allowed to use notes, books, and even external help, such as a friend. Since students were *allowed* to use notes, books, and even external help, this cannot be considered "cheating" in the traditional sense of the word. However the outcome of this experiment should replicate the cheating outcome absent the guilt or euphoria, depending upon the student's attitude. Since we expect the objective outcome of the later testing technique to replicate the outcome of cheating, in this study we refer to this testing technique as "cheating." We compared the scores of these two testing techniques. The results indicate that the quiz scores where students were proctored versus the quiz scores where students were not proctored are not statistically significantly different from each other.

Data

In this study we make use of data on quiz scores from principles microeconomics and macroeconomics courses at a regional Southern university in the United States.

Different sections of principles of microeconomics and principles of macroeconomics from Fall 1999 through Spring 2003, taught by the same instructor, were given a total of seventy six quizzes. All the sections were given similar multiple-choice questions for quizzes spread across the semester. That is, no distinction was made with respect to the content and rigor whether the quiz was a proctored quiz or an un-proctored quiz. From the Fall semester of 1999 through the Spring semester of 2002 students were given fifty eight quizzes containing multiple choice questions and were proctored by the instructor. Each quiz had ten questions and was worth twenty points each. That is, each question was worth two points. The students were given about twenty minutes to complete the quiz.

From the Fall semester of 2002 through the Spring semester 2003 in various of sections of principles of microeconomics and principles of macroeconomic students were given eighteen quizzes containing multiple choice questions using *Blackboard*[®]. However, during this time period quizzes were not proctored and students were allowed to use notes, textbooks, and even external help, such as a friend. In order to ensure that students took the “open notes, open book” quizzes given over the internet seriously we rearranged the incentives structure and added the following statement in the syllabus “You have to score, on average, 60 percent, or better, on your exams for your quizzes to be counted.” Please note that exams were in addition to quizzes and were in-class, proctored exams.

Students were allowed to take quizzes on or off campus. The quizzes were left online for about ten hours. During this time period, a student could logon and take the quiz anytime he/she desired. However once the student had logged on to take the quiz he/she had thirty minutes to complete the quiz, and was allowed only one attempt. Each quiz had ten multiple choice questions and was worth twenty points; two points for each question. Each student received the same quiz.

To ensure that students did not feel pressured since some students might have been new to online testing, students were given thirty minutes to take an un-proctored quiz (from Fall 2002 through Spring 2003), as opposed to twenty minutes when the quiz was proctored (Fall 1999-Spring 2002). Furthermore, in the beginning of each semester (Fall 2002-Spring 2003), students were given a “practice quiz” before the first quiz was given.

Although the student was informed about the quiz score after he/she had submitted the answers, the correct answers were not revealed. This means that students had the opportunity to discuss the questions with each other. Furthermore, quizzes were of similar difficulty level and the types of questions—critical thinking, recall and reproduce, mathematical calculations, etc. In other words no distinction was made in the selection of questions whether the quiz was proctored or un-proctored. The latest edition available of the same textbook was used and questions were selected from the test bank provided with the text.

It is important to note that in this experiment, the same instructor taught all the different sections from the Fall semester of 1999 through the Spring semester of 2003. This controls for the instructor variability that may otherwise become a confounding factor in the results. Also, since the latest editions of the same text were used, this controls for the possible text variability.

Results and Discussion

In this section we present and discuss the results of the study. In order to conduct the analysis, we calculated the average scores on each of the seventy six quizzes. We tested the hypothesis that:

$$H_0: \mu_{\text{Proctored}} = \mu_{\text{Un-proctored}}$$

$$H_1: \mu_{\text{Proctored}} \neq \mu_{\text{Un-proctored}}$$

Where $\mu_{\text{Proctored}}$ is the population mean score of the proctored quizzes and $\mu_{\text{Un-proctored}}$ is the population mean score of the un-proctored quizzes. Sample statistics along with t-stats are presented in Table 1.

The results point out that the averages of the quiz scores are not statistically significantly different whether the quiz is proctored or un-proctored. As we can see that we cannot reject the null hypothesis of the equality of mean quiz scores at the conventional levels.

This result is rather surprising considering that the online quizzes were not proctored and all the students received the same quiz. A student who had already taken the quiz had the opportunity to at least discuss the questions with other students. Even though the correct answers were not revealed, the student was notified about his/her score on the test. One would expect that under these circumstances the students taking the quiz online would score higher points. But this does not seem to be the case and the averages of quiz scores are not statistically significantly different.

This finding may be due to the fact that hard working students are usually friends with hard working students and less serious students socialize with less serious students. This means that even if a student cheats by asking advice from friends, he/she may not necessarily get the right answer since his/her friends do not know the right answer either.

The fact that proctored and un-proctored quizzes were similar with respect to rigor and content, and that students had fifty percent more time to take an un-proctored quiz, makes the results even more telling.

Conclusion

Does cheating on a test pay off in terms of higher points on a test? In this study we have tried to answer this important question. We used data for fifty eight proctored and eighteen un-proctored quizzes given to the various sections of principles level economics classes to compare the average quiz points. The results of the study point out that the average points on a multiple choice quiz are not affected whether a student is given the opportunity to cheat, as defined in the study, or the quiz is proctored.

References

- Becker, B. (1968) "Crime and Punishment: An Economic Approach" *Journal of Political Economy*, 76(2). pp.168-217.
- Bunn, D. N., Caudill, S. B. and Gropper, D. M. (1992) "Crime in the Classroom: An Economic Analysis of Undergraduate Student Cheating Behavior" *Journal of Economic Education*. Summer 1992. pp.197-207.
- Kerkvliet, J. (1994) "Cheating by Economics Students: A Comparison of Survey Results" *Journal of Economic Education*. Spring 1994. pp.121-133.
- Kerkvliet, J. and Sigmund, C. L. (1999) "Can We Control Cheating in the Classroom?" *Journal of Economic Education*. Fall 1999. pp.331-343.
- Nowell, C. and Laufer, D. (1997) "Undergraduate Student Cheating in the Fields of Business and Economics" *Journal of Economic Education*. Winter 1997. pp.3-12.
- Tittle, C. R. and Rowe, A. R. (1973) "Moral Appeals, Sanction Threat and Deterrence: An Experimental Test." *Social Problems*, 20 (4). pp.488-498.

Table 1

	<u>Proctored Quizzes</u>	<u>Un-proctored Quizzes</u>
Mean	15.12	15.38
Variance	3.63	7.67
Observations	58	18
t-stat	-0.373	
Critical Value (one-tail, 5%)	1.67	
Critical Value (two-tail, 5%)	2.07	