EXPERIMENTING ON CLASSROOM EXPERIMENTS: DO THEY INCREASE LEARNING IN INTRODUCTORY MICROECONOMICS?*

by

Mark Dickie
Department of Economics
University of Southern Mississippi
Hattiesburg, MS 39406
M.Dickie@usm.edu

March 13, 2000

* Preliminary draft. Thanks to Rajshree Agarwal, A. Edward Day, Franklin G. Mixon, Jr. and W. Charles Sawyer for helpful comments on a previous version.

EXPERIMENTING ON CLASSROOM EXPERIMENTS: DO THEY INCREASE LEARNING IN INTRODUCTORY MICROECONOMICS?

Abstract

A controlled experiment is used to test effects of classroom experiments on learning of microeconomic principles. Seven experiments were conducted in each of two class sections, with one section provided a grade incentive for profits earned in experimental markets. No experiments were conducted in a third (control) section. The Test of Understanding in College Economics (TUCE) was used as a pre- and post-test. Compared to students in the control section, students in experimental sections achieved significantly greater improvements in TUCE scores. Experiments seem to offer larger benefits to better students, and may reduce achievement among the very worst students. Grade incentives for experimental profits do not appear to increase learning. Results are consistent with the idea that active learning improves achievement.

1. Introduction

A number of indicators suggest that economists are increasing their use of classroom experiments and games for teaching. Supplements accompanying some introductory textbooks now include experiments for classroom use (Delemeester and Neral 1995, Ortman and Colander 1995), while the book of experiments by Bergstrom and Miller (1997) may be used as a supplement or as a stand-alone text. Textbooks for more advanced classes also sometimes incorporate games or experiments (Stodder 1998), and the *Journal of Economic Perspectives* has a regular feature on classroom experiments (Holt 1996).

While interest in classroom experiments appears to be increasing, quantitative evidence on whether experiments and games improve student learning is mixed. Some authors have found that student participation in one game or experiment improves understanding of the specific economic topic addressed in the game. Frank (1997) found that students exposed to a short classroom experiment about use of common-property resources performed better on a test about the "tragedy of the commons" than students in control groups who did not witness the experiment. Gremmen and Potters (1997) tested the effectiveness of a classroom game relative to lectures on the same topic. Students who participated in a computerized game about international economic interdependence performed better on a related test than did students who received lectures instead of the game. Walker et al. (1998) report that students exposed to a classroom demonstration of a pit auction scored 3.5 percent higher on a post-test of the supply and demand model than students exposed to lectures instead. On the other hand, Cardell et al. (1996) found that integrating four classroom experiments into the introductory micro- and macroeconomics curriculum, along with other innovations, yielded no statistically significant gains in achievement relative to students receiving conventional lectures and only one classroom experiment.

These results are difficult to reconcile with one another. If a single experiment is beneficial, one would think that integrating several experiments into the curriculum would yield an even larger cumulative impact on learning (Frank 1997). One explanation consistent with each of the results above is that substituting experiments for other classroom activities is subject to sharply declining marginal benefits and/or sharply increasing marginal opportunity costs, so that the optimal number of classroom experiments is one.[1] Another possibility is that the classroom experiments used by Cardell et al. (1996), which were focused primarily on microeconomic topics, improved understanding of micro- but not of macroeconomics, while other innovations had no effect on learning in either subject.[2]

This paper extends earlier work in the following ways. First, the study tests whether integrating several experiments into the curriculum during a semester improves learning of microeconomics. This focus is somewhat broader than previous tests of whether conducting a single experiment increases understanding of one topic, but narrower than the Cardell et al. (1996) test of whether introducing experiments along with other innovations improves learning of both micro- and macroeconomics. The focus adopted in this paper may reflect the objectives of many instructors, as well as the way many might use experiments in class. For example, instructors may be more concerned with students' overall learning of a subject than with understanding of any single topic, but may want to consider effects of classroom experiments in isolation from other innovations. Instructors might conduct several experiments during a semester, and use results of experiments as starting points for subsequent class discussions, lectures or assignments.

This paper also extends previous work by testing whether effects of experiments on learning depend on student aptitude. Differences in mean outcomes are important in considering whether to use classroom games or experiments, but instructors also may be concerned with effects of experiments on the dispersion of learning outcomes. Do experiments mainly benefit better students, widening existing differentials in learning? Or do experiments narrow dispersion by improving the relative performance of weaker students?

A third new feature of the present paper is a test of whether provision of a grade incentive for profits earned in experimental markets affects student achievement. The issue of incentives is raised frequently in discussions of classroom experiments. Many authors suggest using grade incentives in classroom experiments (Delemeester and Neral 1995), while others question the efficacy of at least some applications of grade incentives (Stodder 1998). Finally, as in Cardell, et al. (1996), the study employs experimental and control groups together with a pre-test/post-test design to focus on differences in learning, rather than differences in understanding, associated with experiments. Controls for student aptitude and other characteristics are used in a multiple-regression framework to isolate the partial effect of classroom experiments on achievement.

The paper reports results of an experiment involving over 100 students in three sections of microeconomic principles taught by the same instructor during the same semester. In one section, seven classroom experiments were conducted, and students were given a grade incentive for participation and for profits earned in experimental markets. The same experiments were conducted in the second section, but with no grade incentive. No experiments were conducted in the third section, with lectures used instead. The Test of Understanding in College Economics (TUCE, Saunders 1991) was used as a pre-test and post-test in each section. Use of classroom experiments is associated with a 1.7-question greater mean improvement in test scores on the 33-question TUCE (Microeconomics Section), an effect that is significant at the five percent level. Controlling for student aptitude and other characteristics in a multiple-regression framework, or adjusting for potential sample selection bias associated with incomplete observation of test scores, does not substantively alter the conclusion that integrating classroom experiments into the introductory microeconomics curriculum increases learning. Experiments are found to offer the greatest benefit to better students, and may actually reduce learning among the very worst students, with a resulting increase in the dispersion in achievement. Grade incentives to reward performance in experimental markets appear to have no impact on learning beyond that attributable to the experiments themselves.

2. Methods

The study was conducted at a comprehensive public university in the South with an enrollment of about 13,000 students, approximately 85 percent of whom are undergraduates. About three-fourths of applicants are admitted, and the university is designated a Doctoral I institution in the Carnegie classification. Subjects in the study were students in three sections of microeconomic principles, each taught by the author during Spring semester 1999. Students enrolled in introductory microeconomics at this university typically are business, but not economics, majors who have taken one semester of macroeconomic principles.

In two of the class sections, referred to as the "experimental" sections, seven classroom experiments were conducted over the course of the 15-week semester. No experiments were conducted in the third, or "control," section, with conventional lectures used instead. Students in one of the experimental sections were offered a grade incentive for participation and profits earned in experimental markets – up to five points on their final semester average (on a 100-

point scale).[3] No grade incentive was offered to students in the second experimental section. Table 1 lists the experiments. The TUCE, Microeconomics section, was given as a pre-test on the first day of class and as a post-test in conjunction with the final exam. To motivate students to take the tests seriously, test results were used in computing students' semester grades.[4]

Additional data were collected from the University Registrar on each student's composite score on the American College Test (ACT), cumulative grade-point average (GPA), number of semester hours passed, age, race and gender. These data, along with TUCE scores and indicators of participation in experiments with and without incentives, are used to estimate a "production function" for understanding of microeconomic principles (Becker 1997). Output is defined in value-added terms as the difference between post-test and pre-test scores.

Several features of the study and methods should be noted. First, the study employs experimental and control groups together with a pre-test/post-test design to focus on differences in learning, rather than differences in understanding, of microeconomics. Controls for student aptitude and other characteristics are used in a multiple-regression framework to isolate the partial effect of experiments on learning. Second, student achievement is measured using the TUCE, a standardized, well documented and widely used instrument for measuring understanding of economic principles. Recent uses of the TUCE include assessments of teachers' economic knowledge (Allgood and Walstad 1999) and effects of the Internet on economic education (Agarwal and Day 1998).

Third, the instructor's prior experience in experimental economics was limited to conducting a few classroom experiments during the preceding semester. Thus, results may be of interest to instructors who are not experimental economists but who may consider using experiments in class. Fourth, each of the seven experiments provided the basis for at least one subsequent lecture, and four of the seven provided the basis for required homework assignments. To insure that effects of homework were not inadvertently attributed to classroom experiments, students in the control section were required to complete an equal number of homework assignments of approximately the same length, taken from the textbook and study guide.[5] Midterm and final examinations reflected course content as taught in each section. Exams in the experimental sections included questions pertaining to the experiments and related lectures and assignments, while exams in the control section emphasized material from that section's lectures and homework assignments. Thus, the study tests effects on overall learning from integrating experiments into the course throughout the semester, rather than the effect of a single experiment on understanding of one topic.

Fifth, some students (28 of 142) missed the pre-test because they were not present on the first day of class, while others (six of 142) missed the post-test because they did not take the final exam. Completion of both tests reflects student choices that may be affected by unobserved factors such as motivation that are related to learning. Conventional sample-selection methods are used to investigate sensitivity of results to students' endogenous choices affecting test completion.

3. Data and Initial Tests

Table 2 provides summary measures of variables used in the study. As shown, the average subject is a 23 year-old junior with better than a "C" grade-point average, who earned a composite score of about 21 on the ACT (equal to the national mean). Four-fifths of the students were business majors, three-fourths were Caucasian, and half were female. In the experimental sections, the average student participated in five or six of the seven experiments conducted.

(Although not shown in Table 2, about 40 percent of the students participated in all, and about 30 percent in six, of the seven experiments.)

Table 2 provides information about performance on the TUCE. On average, students correctly answered nine of the 33 questions on the pre-test and 12 on the post-test. Relative to the TUCE "norming sample," pre-test scores are at the 25th percentile while post-test scores are at the 27th percentile. Students in the control section scored slightly higher on the pre-test but were out-performed on the post-test by their counterparts in the experimental sections.

In this study, the difference between post-test and pre-test scores is taken as the measure of achievement or learning. Students in the experimental sections improved their pre-test score by an average of 3.6 correctly answered questions, 1.7 more than the 1.9 question mean improvement achieved in the control section. Use of experiments also appears to increase somewhat the dispersion of test scores, as variances of post-test scores and of the change in test-scores are larger in the experimental than in the control section. Within the experimental group, students in the section receiving a grade incentive for profits improved their TUCE scores by slightly more than students given no grade incentive.

Table 3 presents results of testing for differences between sub-sample means of variables used in the study. Results in the middle column of Table 3 pertain to differences between the control section and the two experimental sections combined, while results in the right-hand column pertain to differences between the two experimental sections, according to whether or not a grade incentive was offered for profits earned in experimental markets. The *t*-statistics presented in Table 3 can be computed from information in Table 2 assuming unknown but equal variances between groups, and *p*-values presented correspond to two-tail tests.

As shown in Table 3, the difference between mean changes in TUCE scores in experimental and control groups is statistically significant at the five percent level in a two-tail test. The null hypothesis that variances of the change in test score are equal between experimental and control sections is not rejected at the five percent level, although it would be rejected at the ten percent level. (The test statistic, $F(79,27) = 1.837,$ is not presented in Table 3 but can be computed from information in Table 2.)[6] Within the experimental group, there is no significant difference between variances (F = 1.119) or means (t = 0.843) of the change in test scores between the section with a grade incentive and the section without.

The change in test score is the only measured variable showing a statistically significant (at ten percent or less) difference between experimental and control groups, suggesting that the greater achievement of students in the experimental sections is not attributable to differences in student aptitude or other characteristics.[7] Within the experimental group, however, there are some significant differences in student characteristics between the two sections. Student aptitude (GPA) and academic experience (credit hours earned) are greater in the section given a grade incentive for profits, although these advantages did not lead to significantly greater achievement gains than students in the section without a grade incentive. Students in the section with an incentive also were more likely to complete both pre- and post-tests.

In summary, use of classroom experiments is associated with a statistically significant increase in mean student achievement in microeconomic principles. The effect of experiments on achievement seems not to depend on provision of a grade incentive for experimental profits. Somewhat weaker evidence suggests that use of experiments may increase the dispersion of student achievement. The next section investigates the sensitivity of these conclusions to regression-based controls for student characteristics, and to adjustments for potential sample selection bias.

4. Estimates of Educational Production Functions

Estimated educational production functions presented in this section support inferences about partial effects of classroom experiments on student achievement, net of influences of other determinants of learning, such as aptitude (as measured by GPA or by ACT composite score). Results are presented with and without adjustment for potential sample selection bias associated with incomplete observation of test scores.

Estimates of four specifications of the educational production function are presented in Table 4. The dependent variable is the change in test score, and there is no correction for sample selection. In each specification, a likelihood-ratio test fails to reject the hypothesis of groupwise homoskedasticity (between experimental and control groups). Regressions reported in columns labeled (1) and (2) include only a constant term, the indicator for membership in the experimental group, and a measure of aptitude (cumulative GPA or ACT composite score). Column (3) builds on column (1) by adding several demographic measures and the indicator for whether a grade incentive for profits was offered. Coefficients of the additional variables are jointly insignificant at the ten percent level ($p = .177$). (The test statistic, $F(6,99) = 1.529,$ can be computed from information presented in columns (1) and (3)). Finally, the column (4) regression retains only the explanatory variables with statistically significant coefficients from column (3).

As shown in Table 4, students in the experimental sections improve their TUCE scores by 1.3 to 1.8 more correct answers than their counterparts in the non-experimental group, net of effects of GPA and other variables. This effect is several times larger than the effect of teaching innovations, including classroom experiments, on micro- and macroeconomics students reported by Cardell et al. (1996). The partial effect of being in the experimental group is statistically significant at the ten percent level in a two-tail test in column (4), and at five percent in each of the other three regressions. Apart from membership in the experimental group and GPA, only age has a statistically significant effect on the change in test scores. In particular, a grade incentive for profits earned in experimental sessions appears to add no value to student achievement, beyond the effect of the experiments.

Table 5 presents results of re-estimating regressions in Table 4 using Heckman's (1979) two-stage correction for potential sample selection bias associated with the fact that not all students took both pre- and post-tests. Over 80 percent of missing test scores are pre-tests (see Table 2), suggesting that students' characteristics at the beginning of the semester, rather than events during the semester, are likely to be the main determinants of missing data. In the first-stage probit regression, the propensity to take both tests is specified as a function of GPA, hours passed, business major, gender, race and age. Results, presented in a supplemental appendix, indicate that students with higher GPAs, and Caucasian students, were significantly more likely to complete both tests, while coefficients of all other explanatory variables were insignificant at the ten percent level in two-tail tests.

Four points should be noted about results in Table 5. First, the coefficient of the inverse Mills' ratio is not statistically significant at conventional levels in any specification, providing little support for the importance of self-selection. Second, if the need to correct for selectivity bias is granted, conclusions about beneficial effects of classroom experiments are weakened slightly. Estimated coefficients of the experimental section indicator are slightly smaller, with smaller *t*-ratios, in Table 4 than in Table 5. The coefficients are statistically significant at the ten percent level in two-tail tests in each of the four specifications. Third, the major apparent difference between Table 4 and Table 5 regressions is the coefficient of cumulative GPA, which is about two times larger in selectivity-corrected regressions. The marginal effect of GPA on the change in test scores, however, ranges from 0.88 to 0.94 in the selectivity-corrected regressions, a magnitude similar to coefficients of cumulative GPA in Table 4. Fourth, each of the four regressions in Table 5 was re-estimated twice based on two alternative specifications of the probit selection equation. The alternative specifications added the experimental section indicator and/or the grade incentive indicator to the explanatory variables already included in the probit equation. In the resulting estimates of the educational production functions, which are presented in a supplemental appendix available on request, the experimental section indicator retained a statistically significant coefficient of about 1.7.[8] In summary, beneficial effects of classroom experiments on student learning of microeconomic principles do not appear attributable to sample selection bias associated with incomplete observation of test scores.[9]

Two final questions are addressed in the remainder of this section. The first concerns the relationship between student aptitude and achievement effects of experiments. Do classroom experiments equally benefit students of all aptitudes? If not, do benefits accrue primarily to better or to worse students? Results in Table 2 provide at least some evidence that experiments increase dispersion in learning, which would occur if beneficial effects of experiments increased with student aptitude. To investigate this possibility, Table 4 regressions were re-estimated by ordinary least squares while allowing slope coefficients, as well as intercepts, to vary between groups. Results presented in columns (1), (2) and (4) of Table 6 indicate that beneficial effects of experiments on learning increase with student aptitude, as measured by cumulative GPA or

ACT composite score. The column (3) regression, on the other hand, suggests that this effect may arise from the grade incentive rather than from the experiments themselves.[10]

Results in columns (1) and (2) are particularly interesting because they imply that use of classroom experiments actually reduces learning among the worst students. In column (1), experiments increase the change in TUCE scores only for students whose GPA exceeds 2.04 (roughly one standard deviation below the mean among all students taking both pre-test and post-test). In column (2), experiments benefit only students whose ACT composite score exceeds 16.6 (again about one standard deviation below the mean). Thus, the very weakest students may learn less when classroom experiments are used, while the majority of students learn more.[11]

The final issue considered here is whether achievement increases with the number of experiments in which a student participates, and if so, whether achievement increases at a decreasing rate. To investigate these questions, the change in test scores for students in the experimental sections was regressed on a constant, cumulative GPA, and the number of experiments and its square. The predicted change in tests scores is given by:

$$\underset{(-2.151)}{-9.18} + \underset{(1.541)}{1.05(GPA)} + \underset{(2.133)}{3.88(Numex)} - \underset{(-1.861)}{0.347(Numex)^2}, \qquad N = 80, \quad SER = 3.708, \quad R^2 = 0.14, \quad F = 4.27,$$

where t-statistics are shown in parentheses beneath estimated regression coefficients and $Numex$ denotes the number of experiments in which a student participated. According to this regression, achievement increases at a decreasing rate with the number of experiments, reaching a maximum at 5.59 experiments. This result probably overstates the partial effect of experiments on learning, however, because regular class attendance was not separately measured. Students who attend class regularly are more likely to participate in a larger number of experiments, and also are likely to learn more than their class-cutting counterparts regardless of experiments (Durden and Ellis 1995).

5. Discussion

The paper reports results of a controlled experiment in which use of classroom experiments is associated with a 1.7-question greater mean improvement in test scores on the 33-question TUCE (Microeconomics Section), an effect that is significant at the five percent level. Controlling for student aptitude and other characteristics in a multiple-regression framework, or adjusting for potential sample selection bias associated with incomplete observation of test scores, does not substantively alter the conclusion that integrating classroom experiments into the introductory microeconomics curriculum increases learning. Experiments are found to offer the greatest benefit to better students, and may actually reduce learning among the very worst students and thus may increase somewhat the dispersion in achievement. Grade incentives to reward performance in experimental markets do not appear to have a major impact on the average student's learning beyond that attributable to the experiments themselves.

Results confirm the conjecture of Frank (1997) that tests of single experiments conducted in isolation understate beneficial effects of making more extensive use of classroom games or experiments. But results run counter to findings of Cardell et al. (1996) that integrating four classroom experiments into the introductory economics curriculum, along with other innovations, did not significantly increase achievement among micro- and macroeconomic students, relative to students receiving conventional lectures and only one classroom experiment. A possible explanation for this disparity is that many classroom experiments, with their focus on interactions in individual markets and other microeconomic topics, improve understanding of introductory microeconomics but provide few benefits for understanding macroeconomics.

More generally, benefits may vary markedly over different experiments, with some experiments adding little value and others increasing learning substantially. Future research could examine this question by administering a separate test after each of several experiments, as well as testing the aggregate effect.

The finding that grade incentives do not enhance the average student's benefit from classroom experiments of course pertains only to the incentive used in this study. Perhaps a larger, or differently designed, incentive would affect learning. Additionally, use of incentives may improve the correspondence between experimental results and theoretical predictions.

Results presented do not explain why classroom benefits are found to increase learning. To the extent that students enjoy participating in experiments or games, the activities may encourage attendance or effort. Resulting changes in student behavior may cause the increase in learning. Testing this explanation would require measurement of attendance and study effort, and estimation of educational production functions accounting for endogeneity of effort (Becker 1997). Alternatively, the benefit may be a return to additional effort from the instructor, rather than from the students, in view of the time required to prepare experiments and related materials. Another possible explanation for benefits of experiments is the hands-on experience of seeing theory work. It may be easier to understand how and why tax incidence is shared between buyers and sellers after seeing this result occur in a classroom market.

On the other hand, the achievement gains measured here may not arise specifically from experiments. Results are consistent with a narrow interpretation that experiments improve achievement, as well as with a broader view that the real benefit arises from engaging students in active learning. These two interpretations cannot be distinguished based on the data used in this study alone, but results of other research support the broader, "active learning" explanation. For example, Agarwal and Day (1998) report an increase in student achievement from use of the Internet in economics education. Classroom experiments and the Internet are dissimilar innovations, but both encourage interactive participation by students. Since experiments and the Internet both seem to increase achievement, the real cause may be the active learning that each innovation encourages.

In any event, conducting experiments provides some side benefits apart from the gains in achievement measured by the TUCE. First, learning may not be confined to the topics of the experiments (Fels 1993). Students may learn something about conducting experiments, or about testing theories. Second, students seem to enjoy participating in experiments. In the semester preceding the study reported in this paper, students in introductory microeconomics classes where four experiments had been conducted completed a brief survey. Seventy-seven percent of the 69 respondents reported that experiments were more interesting than lectures, while only 51 percent reported that experiments were better at helping them learn.[12]

Conducting experiments also involves costs. The benefits of experiments presented here are measured net of the opportunity cost of students' time, in terms of lectures and other assignments foregone. But a significant unmeasured cost is the instructor's time to prepare materials, compile results, and prepare for any subsequent lectures, discussions or assignments incorporating the experimental results.[13] There is also the possibility that a classroom experiment may not work as intended. These costs are reduced substantially but not eliminated through use of a guide such as Bergstrom and Miller (1997). Also, results suggest that one cost of experiments may be reduced achievement among the very worst students. Instructors have to judge for themselves whether benefits of conducting classroom experiments justify all opportunity costs.

References

Allgood, Sam and William B. Walstad (1999). "The Longitudinal Effects of Economic Education on Teachers and Their Students," Journal of Economic Education 30(2): 99-111 (Spring).

Agarwal, Rajshree and A. Edward Day (1998). "The Impact of the Internet on Economic Education," *Journal of Economic Education* 29(2): 99-110 (Spring).

Becker, William E. (1997). "Teaching Economics to Undergraduates," *Journal of Economic Literature* 35(3): 1347- 1373 (September).

Bergstrom, Theodore C. and John H. Miller (1997). *Experiments with Economic Principles.* New York: The McGraw-Hill Companies, Inc.

Cardell, N. Scott, Rodney Fort, Wayne Joerding, Fred Inaba, David Lamoreaux, Robert Rosenman, Ernst Stromsdorfer, and Robin Barlett (1996). "Laboratory-Based Experimental and Demonstration Initiatives in Teaching Undergraduate Economics," *American Economic Review* 86(2): 454-459 (May).

Delemeester, Greg and John Neral (1995). *Classroom Experiments To Accompany Taylor's Economics: A User's Guide.* Boston: Houghton Mifflin Company.

Durden, Garey C. and Larry V. Ellis (1995). "The Effects of Attendance on Student Learning in Principles of Economics," *American Economic Review* 85(2): 343-346 (May).

Eaton, B. Curtis and Diane F. Eaton (1988). *Microeconomics.* W.H. Freeman and Company.

Fels, Rendig (1993). "This is What I Do, and I Like It," *Journal of Economic Education* 24(4): 365-70 (Fall).

Fraas, John W. (1982). "The Influence of Student Characteristics on the Effectiveness of Simulations in the Principles Course," *Journal of Economic Education* 13(1): 56-61 (Winter).

Frank, Bjorn (1997). "The Impact of Classroom Experiments on the Learning of Economics: An Empirical Investigation," *Economic Inquiry* 35(4): 763-769 (October).

Greene, William H. (1997). *Econometric Analysis.* Third Edition. Prentice Hall.

Gremmen, Hans and Jan Potters (1997). "Assessing the Efficacy of Gaming in Economic Education," *Journal of Economic Education* 28(4): 291-303 (Fall).

Heckman, James (1979). "Sample Selection Bias as a Specification Error," *Econometrica* 47: 153-161.

Holt, Charles A. (1996). "Classroom Games: Trading in a Pit Market," *Journal of Economic Perspectives* 10(1): 193-203 (Winter).

Ortman, A. and D.C. Colander (1995). *Experiments in Teaching and Understanding Economics, to accompany* Economics *(2nd. Edition) by D.C. Colander.* Chicago: Irwin.

Parkin, Michael (1997). *Economics*. Fourth Edition. Addison-Wesley.

Rush, Mark (1997). *Study Guide: Parkin, Microeconomics, Fourth Edition.* Addison-Wesley.

Saunders, Phillip (1991). *Test of Understanding in College Economics Third Edition: Examiner's Manual.* New York: Joint Council on Economic Education.

Stodder, James (1998). "Experimental Moralities: Ethics in Classroom Experiments," *Journal of Economic Education* 29(2): 127-138 (Spring).

Endnotes

[1] Or the hands-on experience in experiments and games may enhance immediate understanding while leading to no lasting gains in learning.

[2] Regressions reported by Cardell et al. (1996) show a significantly larger achievement gain among students in microeconomics, but no interaction between microeconomics course material and membership in the experimental group is reported.

[3] Points were assigned based on a student's actual profits relative to profits he would be expected to earn in equilibrium. Each student in the section with incentives was "endowed" with two of the five points at the beginning of the semester, so that losses would have an explicit cost. In the prior semester, when an incentive was offered entirely as extra credit, some students took losses to increase profits of a popular but weak student.

[4] The pre-test score was converted to bonus points on the first regular exam. Then, the difference between post-test and pre-test scores was converted to a bonus on the final exam with weights chosen so the pre-test score ultimately had no effect on a student's grade. At the time they took the pre-test, however, students were not informed that a post-test would be given, or how it would be graded.

[5] Parkin (1997) was the main textbook in all sections. The associated study guide (Rush 1997) was required in the control section, while the book of experiments by Bergstrom and Miller (1997) was required in the experimental sections.

[6]The difference between means remains significant at five percent when equal variances are not assumed (t = 2.480).

[7] If the same midterm or final exams had been administered in each section, students' scores on these tests would provide another basis for testing achievement effects of experiments. As mentioned previously, however, exams in each section covered course material as presented in that section. Although the information is not presented in Table 3, there were no significant differences in final exam scores between any of the sections. Yet another test might be based on student evaluations of instructor effectiveness, but these were not conducted in Spring 1999 as part of a transition to a new administration.

[8] The first alternative probit equation added the indicator for a grade incentive to the probit equation, which Table 2 suggests may be related to the propensity to take both pre- and post-tests. The second alternative probit equation included both the grade incentive and the experimental section indicator variables. None of the additional coefficients in these probit equations is significant at the five percent level in a two-tail test. When the four specifications of the educational production function in Table 5 were re-estimated using these specifications of the

probit selection equation, the estimated error correlation (between the regression and the selection equation) exceeded unity in three of eight cases. In the other five regressions, the coefficient of the experimental section indicator fell between 1.40 and 1.92, and was statistically significant at the ten percent level or less.

[9] A related selection bias may arise because students were not randomly assigned to experimental and control groups, but indirectly chose their group by choosing their class section. Students in the morning sections are in the experimental group, while students in the afternoon section are in the control group. (An alternative design with the control section in the morning did not seem feasible, because the two morning sections met back-to-back in the same classroom, with lots of interaction between the students.) If students taking morning classes are "better" than students taking afternoon sections, in the sense that they would learn more economics with or without classroom experiments, than ordinary least squares estimators overstate the effect of experiments. The fact that the afternoon (control section) scored higher on the pre-test, and that there are no significant differences in GPA or ACT composite between experimental and non-experimental sections, would lend little credence to this possibility. But the potential bias hinges on correlations between unobservable factors and membership in the experimental group. Thus, a two-stage, "treatment effects" estimator was used to correct for possible correlation between the error in the education production function and membership in the experimental group (see Greene 1997, pp. 981-82). This methodology yields substantially larger estimates of effects of classroom experiments, but estimated coefficients are statistically very imprecise. In any event, beneficial effects of experiments do not appear attributable to self-selection of class sections. Unfortunately, it does not appear possible to test the related possibility that the instructor performs better in the morning than in the afternoon.

[10] The null hypothesis that the coefficients of interaction terms (those variables included in Table 6 regressions but excluded from Table 4 regressions) are zero was tested for each of the four specifications. This hypothesis would be rejected at the ten percent level or less for the column (1) and (2) regressions but not for the column (3) or (4) regressions. The test statistics for the columns (3) and (4) regressions, which can be computed from information in Tables 4 and 6, are respectively $F(12,87) = 1.054$ and $F(2,102) = 2.22$. Corresponding p-values are .408 and .114. In columns (1) and (2) of Table 6, only one interaction term is added to the corresponding Table 4 regressions and the relevant test statistic is the t-ratio presented in Table 6.

[11] This result runs somewhat counter to Fraas (1982), who reports that better students learn more from lectures, while worse students learn more from simulation-gaming.

[12] As one student put it, "Anything is better than a lecture." A copy of the eight-question survey, which was not pre-tested and was completed anonymously by students in attendance the day it was administered, and a summary of survey results are presented in a supplemental appendix available on request.

[13] The largest cost may be preparing subsequent lectures, because experimental results vary somewhat from section to section, and students in one section may not find results obtained in another section particularly interesting.

Table 1. Classroom Experiments Used.	
Topic	Source
Spatial Competition	Adapted from Eaton and Eaton (1988)
Comparative Advantage and Trade	Adapted from Bergstrom and Miller (1997)
Demand-Revealing Auction	Adapted from Delemeester and Neral (1995)
Supply, Demand and Market Equilibrium	Bergstrom and Miller (1997)
Sales (Excise) Tax	Bergstrom and Miller (1997)
Prohibition	Bergstrom and Miller (1997)
Labor Market and Minimum Wage	Bergstrom and Miller (1997)

Table 2. Descriptive Statistics by Sub-sample.				
	Means (Standard Deviations) and Numbers of Valid Observations.			
	Experimental Sections			Control Section
Variable	No Grade Incentive for Profits (*N*=55 except where noted)	Grade Incentive for Profits (*N*=50 except where noted)	Combined Experimental Sections (*N*=105 except where noted)	No Classroom Experiments (*N*=37 except where noted)
Pre-Test (Number Correct)	8.9 (2.34) *N*=44	9.1 (3.06) *N*=40	9.0 (2.69) *N*=84	9.3 (3.33) *N*=30
Post-Test (Number Correct)	12.2 (3.50) *N*=49	13.1 (3.67) *N*=52	12.7 (3.59) *N*=101	11.6 (2.98) *N*=35
Change = Posttest – Pretest	3.3 (4.04) *N*=43	4.0 (3.82) *N*=37	3.6 (3.93) *N*=80	1.9 (2.90) *N*=28
Cumulative GPA (4-point Scale)	2.33 (0.78)	2.72 (0.66) *N*=49	2.51 (0.747) *N*=104	2.37 (0.694)
ACT Composite	20.35 (3.31) *N*=51	21.04 (3.58) *N*=45	20.68 (3.44) *N*=96	20.69 (3.00) *N*=35
Semester Credit Hours Passed	75.24 (22.91)	85.45 (25.36)	80.10 (24.53)	79.75 (32.30)
Business Major (0=No, 1=Yes)	0.80 (0.40)	0.84 (0.37)	0.82 (0.39)	0.78 (0.42)
Female (0=No, 1=Yes)	0.51 (0.50)	0.60 (0.49)	0.55 (0.50)	0.41 (0.50)
Caucasian (0=No, 1=Yes)	0.73 (0.45)	0.74 (0.44)	0.73 (0.44)	0.78 (0.42)
Number of Experiments Participated In (0 to 7)	5.49 (1.81)	5.90 (1.42)	5.69 (1.64)	0.0 (0.0)

Table 3. Tests of Differences Between Means[a]: Absolute values of t-statistics and p-values for Two-Tail Tests.		
Variable	Experimental *vs.* Control	Grade Incentive *vs.* None
Pre-Test	$\|t\| = 0.511$ $p = .611$	$\|t\| = 0.200$ $p = .842$
Post-Test	$\|t\| = 1.56$ $p = .122$	$\|t\| = 1.167$ $p = .246$
Change = Posttest – Pretest	$\|t\| = 2.147$ $p = .034$	$\|t\| = 0.843$ $p = .402$
Took Both Post-test and Pre-test	$\|t\| = 0.063$ $p = .950$	$\|t\| = 2.284$ $p = .024$
Cumulative GPA	$\|t\| = 1.018$ $p = .310$	$\|t\| = 2.750$ $p = .007$
ACT Composite	$\|t\| = 0.013$ $p = .990$	$\|t\| = 0.983$ $p = .328$
Semester Credit Hours Passed	$\|t\| = 0.068$ $p = .946$	$\|t\| = 2.167$ $p = .033$
Business Major	$\|t\| = 0.467$ $p = .641$	$\|t\| = 0.527$ $p = .599$
Female	$\|t\| = 1.540$ $p = .126$	$\|t\| = 0.931$ $p = .354$
Caucasian	$\|t\| = 0.603$ $p = .547$	$\|t\| = 0.146$ $p = .884$
Number of Experiments Participated In	---[b]	$\|t\| = 1.278$ $p = .204$

[a] Independent samples *t*-tests of differences between means, assuming unknown but equal variances; *t*-statistics can be computed from information in Table 2.
[b] No experiments conducted in control section.

Table 4. Estimates of Education Production Functions: Coefficients and (*t*-ratios).[a]				
Explanatory Variable	(1)	(2)	(3)	(4)
Experimental Group	1.62 (2.247)	1.717 (2.335)	1.762 (2.039)	1.369 (1.923)
Grade Incentive for Profits			-0.799 (-1.021)	
Cumulative GPA	0.930 (1.630)		0.939 (1.773)	0.773 (1.411)
ACT Composite		0.00203 (0.017)		
Hours Passed			-0.00827 (-0.478)	
Business Major			-0.835 (-1.020)	
Female			-0.0338 (-0.049)	
Caucasian			0.273 (0.326)	
Age			-0.260 (-3.547)	-0.284 (-4.651)
Constant	-0.505 (-0.320)	1.774 (0.721)	6.775 (2.687)	6.54 (2.855)
Number of Observations	108	102[b]	108	108
Standard Error of Regression	3.665	6.780	3.611	3.562
F-statistic	3.76	2.05	2.12	5.05
R^2	0.07	0.04	0.15	0.13

[a] The dependent variable is the difference between post-test and pre-test scores. Heteroskedasticity-consistent estimator of covariance matrix.
[b] ACT composite scores were unavailable for six students.

Table 5. Selectivity-Corrected Estimates of Education Production Functions: Coefficients and (asymptotic *t*-ratios).[a]

Explanatory Variable	(1)	(2)	(3)	(4)
Experimental Group	1.509 (1.919)	1.716 (2.041)	1.660 (1.893)	1.35 (1.745)
Grade Incentive for Profits			-0.723 (-0.890)	
Cumulative GPA	2.226 (2.033)		2.138 (1.260)	1.38 (1.344)
ACT Composite		-0.00650 (-0.0540)		
Hours Passed			-0.0143 (-0.674)	
Business Major			-0.0859 (-0.058)	
Female			0.164 (0.188)	
Caucasian			1.127 (0.748)	
Age			-0.162 (-0.824)	-0.251 (-2.114)
Constant	-5.314 (-1.421)	2.058 (0.737)	-0.817 (-0.077)	3.59 (0.684)
Inverse Mills' Ratio	4.345 (1.469)	-0.339 (-0.204)	4.107 (0.770)	1.92 (0.703)
Number of Observations	108	102[b]	108	108
Standard Error of Regression	3.565	3.723	3.447	3.487
Asymptotic *F*-statistic	3.52	1.37	1.94	3.89
R^{2}[c]	0.09	0.04	0.15	0.13

[a] The dependent variable is the difference between post-test and pre-test scores.
[b] ACT Composite score unavailable for six students.
[c] Not bounded between zero and one.

Table 6. Estimates of Education Production Functions, Allow Slope and Intercept Shifts by Experiments and Grade Incentives: Coefficients and (*t*-ratios).[a]				
Explanatory Variable	(1)	(2)	(3)	(4)
Experimental Group	-6.30 (-2.138)	-7.53 (-1.480)	17.99 (1.200)	1.55 (0.229)
Grade Incentive for Profits			-22.28 (-1.591)	
Cumulative GPA	-1.59 (-1.806)		-1.789 (-1.225)	-1.60 (-1.244)
ACT Composite		-0.370 (-1.742)		
Hours Passed			-.000851 (-0.033)	
Business Major			.538 (0.253)	
Female			-.0128 (-0.009)	
Caucasian			-0.529 (-0.287)	
Age			0.0409 (.124)	0.00955 (.038)
Constant	5.89 (2.533)	9.31 (2.247)	5.52 (0.755)	5.70 (0.970)
Cumulative GPA x Experimental Group	3.09 (2.837)		1.61 (0.892)	2.79 (1.966)
ACT Composite x Experimental Group		0.453 (1.783)		
Hours Passed x Experimental Group			0.0508 (0.972)	
Business Major x Experimental Group			-1.73 (-0.634)	
Female x Experimental Group			-0.553 (-0.276)	
Caucasian x Experimental Group			1.73 (0.675)	
Age x Experimental Group			-1.05 (-1.482)	-0.315 (-1.135)

[a] The dependent variable is the difference between post-test and pre-test scores. Heteroskedasticity-consistent estimator of covariance matrix.

Table 6 (Continued).				
	(1)	(2)	(3)	(4)
Cumulative GPA x Grade Incentive for Profits			2.76 (1.920)	
Hours Passed x Grade Incentive for Profits			-0.096 (-1.755)	
Business Major x Grade Incentive for Profits			1.92 (0.831)	
Female x Grade Incentive for Profits			0.541 (0.297)	
Caucasian x Grade Incentive for Profits			-0.559 (-0.242)	
Age x Grade Incentive for Profits			0.901 (1.386)	
Number of Observations	108	102[a]	108	108
Standard Error of Regression	3.599	3.756	3.599	3.521
F-statistic	4.23	2.15	1.62	3.99
R^2	0.11	0.06	0.18	0.16

[a] ACT composite scores were unavailable for six students.

Appendix: Supplemental Results

Table A-1. Descriptive Statistics of Explanatory Variables		
	Sample Mean (Standard Deviation)	
Explanatory Variable	Students Taking both Pre- and Post-Tests	Students Not Taking Both Tests
Experimental Group (0=No, 1=Yes)	0.74 (0.440)	0.74 (0.448)
Grade Incentive for Profits in Experiments[a] (0=No, 1=Yes)	0.40 (0.492)	0.21 (0.410)
Cumulative GPA[a] (4-point scale)	2.64 (0.645)	1.94 (0.758)
ACT Composite	20.84 (3.269)	20.10 (3.468)
Semester Credit Hours Passed	80.36 (24.859)	78.90 (32.081)
Business Major (0=No, 1=Yes)	0.83 (0.374)	0.74 (0.448)
Female (0=No, 1=Yes)	0.54 (0.501)	0.44 (0.504)
Caucasian[a] (0=No, 1=Yes)	0.80 (0.405)	0.59 (0.500)
Number of Experiments Participated (experimental group only)	5.80 (1.594)	5.32 (1.773)
Number of Students	108	34

[a] Difference between means is statistically significant at the five percent level or less in a two-tail test, assuming unknown but equal variances.

Table A-2. Probit Regression for Probability of Taking Both Pre-test and Post-test: Coefficients and (asymptotic *t*-ratios).			
Explanatory Variable	(1)	(2)	(3)
Cumulative GPA	0.902 (4.181)	0.862 (3.890)	0.856 (3.851)
Hours Passed	-0.00602 (-0.988)	-0.00810 (-1.233)	-0.00807 (-1.222)
Business Major	0.431 (1.373)	0.397 (1.241)	0.406 (1.267)
Female	0.153 (0.562)	0.172 (0.622)	0.179 (0.644)
Caucasian	0.577 (1.985)	0.630 (2.106)	0.622 (2.070)
Age	0.0883 (1.105)	0.100 (1.091)	0.0975 (1.053)
Constant	-3.691 (-2.124)	-3.863 (-1.973)	-3.667 (-1.841)
Grade Incentive for Profits		0.518 (1.611)	0.594 (1.735)
Experimental Group			-0.201 (-0.641)
Number of Observations[a]	141	141	141
Chi-Square	32.19	34.93	35.35
Proportion Correct Predictions	0.82	0.80	0.80

[a] Cumulative GPA was unavailable for one student.

Appendix: Supplemental Results (Continued)

Table A-3. Selectivity-Corrected Estimates of Education Production Functions: Coefficients and (asymptotic *t*-ratios).[a] Using Probit Selection Equation (2) from Table A-2.				
Explanatory Variable	(1)	(2)	(3)[b]	(4)
Experimental Group	1.920 (2.217)	1.773 (2.087)		1.64 (2.106)
Grade Incentive for Profits				
Cumulative GPA	2.542 (2.290)			1.94 (2.019)
ACT Composite		0.0175 (0.147)		
Hours Passed				
Business Major				
Female				
Caucasian				
Age				-0.221 (-1.750)
Constant	-6.600 (-1.762)	1.202 (0.4383)		0.683 (0.134)
Inverse Mills' Ratio	5.249 (1.856)	0.674 (0.441)		3.69 (1.565)
Number of Observations	108	102[c]		108
Standard Error of Regression	3.516	3.721		3.452
Asymptotic *F*-statistic	4.59	1.42		4.50
R^2[d]	0.12	0.04		0.15

[a] The dependent variable is the difference between post-test and pre-test scores.
[b] Estimated correlation between disturbance in regression and selection equation exceeds unity.
[c] ACT Composite score unavailable for six students.
[d] Not bounded between zero and one.

Appendix: Supplemental Results (Continued)

Table A-4. Selectivity-Corrected Estimates of Education Production Functions: Coefficients and (asymptotic *t*-ratios).[a] Using Probit Selection Equation (3) from Table A-2.

Explanatory Variable	(1)[b]	(2)	(3)[c]	(4)
Experimental Group		1.731 (2.050)		1.40 (1.660)
Grade Incentive for Profits				
Cumulative GPA				1.90 (2.033)
ACT Composite		0.0175 (0.147)		
Hours Passed				
Business Major				
Female				
Caucasian				
Age				-0.226 (-1.824)
Constant		1.239 (0.459)		1.13 (0.234)
Inverse Mills' Ratio		0.660 (0.445)		3.51 (1.561)
Number of Observations		102		108
Standard Error of Regression		3.721		3.452
Asymptotic *F*-statistic		1.42		4.50
R^2[d]		0.04		0.15

[a] The dependent variable is the difference between post-test and pre-test scores.
[b] Estimated correlation between disturbance in regression and selection equation greater than unity.
[c] Estimated correlation between disturbance in regression and selection equation greater than unity.
[d] Not bounded between zero and one.

Appendix: Supplemental Results (Continued)

Student Survey Results, Fall 1998 (N = 69)

Note: The survey was conducted in the semester preceding the experimental test reported in the paper.

1. Which of the following classroom games have you participated in during this semester?

Richland / Poorland	Yes	**90%**	No	**10%**
Doughnut Auction	Yes	**67%**	No	**33%**
Apple Market #1	Yes	**77%**	No	**23%**
Apple Market / Price Controls	Yes	**78%**	No	**22%**

2. Which are more interesting, lectures in this class, or the classroom games?
   a. lectures **4%**   b. games **77%**   c. they are about the same. **19%**
3. Which are better at helping you learn?
   a. lectures **13%**   b. games **51%**   c. they are about the same. **36%**

4. Comparing "ordinary" lectures in this class (that do not discuss the games) and the lectures that discuss results of the games, which are more interesting?
   a. ordinary lectures **13%**   b. lectures about games **45%**   c. they are about the same. **42%**

5. Which are better at helping you learn?
   a. ordinary lectures **14%**   b. lectures about games **45%**   c. they are about the same. **41%**

6. Of the games you've participated in, which was most interesting? _____________________

Richland / Poorland	**38%**
Doughnut Auction	**26%**
Apple Market #1	**26%**
Apple Market / Price Controls	**6%**

Least interesting? _________________________

Richland / Poorland	**12%**
Doughnut Auction	**16%**
Apple Market #1	**32%**
Apple Market / Price Controls	**9%**

7. Which was best at helping you learn? __________________________

Richland / Poorland	**38%**
Doughnut Auction	**13%**
Apple Market #1	**23%**
Apple Market / Price Controls	**4%**

Worst at helping you learn? __________________________________

Richland / Poorland	**4%**
Doughnut Auction	**9%**
Apple Market #1	**17%**
Apple Market / Price Controls	**6%**

8. Do you have any other comments about the classroom games?