

Incentives and Student Learning: A Natural Experiment with Economics Problem Sets

By WAYNE A. GROVE AND TIM WASSERMAN*

While carefully explaining the importance and role of optimal decision-making to students, ironically economics faculty lack the most basic information regarding their own instructional production decisions. Despite three and a half decades of economic education scholarship, little is known about what a faculty member can do differently to increase college student learning (Becker, 1997; Siegfried and Walstad, 1998).¹ Faculty assigned problem sets, one of the most common pedagogical practices in college economics courses, are a case in point.² No study assesses of the effectiveness of regular student practice of economic analysis, relative to a control group, in spite of the considerable time and effort required by professors, and perhaps their assistants, preparing and grading and by students in completing these assignments. The limited use of randomized experiments in higher education principally explains the paucity of assessment studies of pedagogical choices. Ethical concerns and institutional restrictions upon human subject's research have discouraged experimental research design, making it difficult to confidently assess the efficacy of a particular instructional choice since sample selection problems always arise in the absence of a control group (see Becker, 2002).³ Thus, Fels (1993, p. 365) observed "that those who use controlled experiments in their research on economics do not use controlled experiments to evaluate their teaching".

Fortuitously, then, in the fall of 1998 a natural experiment occurred at Syracuse University, a large, private residential university in the northeast (Carnegie Classification: Doctoral Research Universities II-Extensive). Inadvertently, 239 students were divided into two groups, differentiated by whether or not assignments counted toward their course grade. In this

paper, then, we evaluate whether students' mastery of course material changes if they have a grade incentive to complete economics homework assignments.

A large empirical literature largely finds a positive relationship between achievement in elementary, middle, and high school and either student reported hours of homework completed or the amount of homework assigned, but not the amount graded, by teachers (Betts, 1997). Kennedy and Siegfried (1997, 393), in a multi-school investigation, report that college economics instructors "cannot influence learning through . . . assigning more homework . . .". Regarding what types of college students might benefit most from practice assignments, existing studies emphasize students' relative level of academic aptitude but provide contradictory evidence. Marlin and Niss (1982), for example, report that weekly computer tests increase the course grades of students in the three lowest GPA quartiles (perhaps, they speculate, by increasing study discipline). Salemi and Tauchen (1982) and Borg, Mason and Shapiro (1989) found that an increase in study time only had a large effect on higher aptitude students (perhaps because below average students had much lower study-time productivity). Kirby, Winston and Santiesteban (2002) found a negative relationship between discount rates and college GPA, suggesting that shortening the time between study and reward/punishment may elicit more effort from scholastically-below average students.

Among faculty who assign problem sets, some base a portion of the course grade upon them and others do not. Why might graded assignments increase learning? Since ungraded assignments offer no direct grade benefit (or penalty), the indirect benefit of better exam or course performance must motivate students to complete them. The crux of students' cost-benefit calculus regarding problem set assignments is that the costs of foregoing some academic, athletic, social, or employment interest is certain and immediate whereas the payoff of

completing the assignments is uncertain and long run (e.g., end of the semester or of college). A grade incentive might induce students to complete problem sets under three circumstances. First, if uncertainty is the obstacle (will completing the assignments really increase my exam and course performance?), graded assignments provide a direct reward/penalty (the magnitude of which depends upon its share of the course grade—for this study, 15 percent). Second, for those who expect an indirect payoff from completing the homework (i.e., better course grade) but deem it too low to bother, graded assignments provide an additional reason. Finally, for students with time-management problems who intend to complete the ungraded assignments, a grade incentive/penalty may operate as a prioritizing mechanism so that intentions are actualized. According to this line of reasoning, then, two sets of students would not be induced by the grade incentive to change their behavior: some would have completed the problem sets without any incentives and for others the expected payoff (direct and indirect) remained too low relative to the value of their time.

In this paper we estimate whether students with graded problem sets learn more than do students without a grade incentive to do so, all else equal. If grade incentives do increase student learning, what is the magnitude of that treatment effect, which subgroups of students, if any, benefited most, and why? As suggested above, without an experimental design the *ceteris paribus* qualification is difficult to confidently establish.

Our results indicate that the grade incentive to practice economics throughout the semester boosted the exam performance of freshmen and did so by at least one-third of a letter grade (e.g., from a C+ to a B-) in Economics 101 compared with control-group students who received identical assignments but no direct grade incentive to complete them. Although our data does not allow us to directly test the cause of the learning gain, limited evidence suggests

that the grade incentive boosted the productivity of first year students' study time rather than the amount of it. These findings about actions any professor can take to increase student learning should interest higher education stakeholders, especially those who wish to reverse high and rising college dropout rates.⁴

I. The Natural Experiment and the Data

During the fall of 1998, 239 undergraduate students enrolled in and completed four sections of introductory microeconomics taught by one of the authors at Syracuse University. The average student was a 19-year old white male. When a colleague fell seriously ill as classes began, a "natural experiment" occurred that separated students into one group (of 143) whose course grades were based on problem set performance and another (of 96) whose course grades were not.⁵ All lectures, handouts, exams, and review sessions were as identical as possible. Every student received the problem sets at the same time and encouragement to practice economics by solving the assignments as important preparation for the exams. Students in both groups received the answer keys at the same time (when the graded-group students handed in their problem sets). Five problem sets were assigned: two prior to the first exam, two before the second exam, and one preceding the final exam. Students with graded problem sets received a course grade that included the average of the best four of five possible problem set scores (comprising 15 percent of the final course grade). Two-thirds of the graded-group students completed all 5 problem sets and only 10 percent handed in fewer than four assignments.

Exam scores and problem set data come from the author's own records, while all other information used in this study come from university records (not from student surveys which have been shown to overstate actual performance). Mastery of the course material was measured

by performance on three required exams of equal value, each of which was assessed with a 0 to 100 point scale. The exams contained a combination of problem solving, short answer questions, and multiple-choice questions. In response to student complaints of different grading standards by each of the three teaching assistants on the first problem set, the same graduate student graded all students' answer to each (non-multiple choice) question in order to make the grading as uniform as possible. The order in which individual student exams and groups of exams were graded varied, TAs either graded all the questions they were responsible for on exams before passing them on or marked one batch of exams at a time before swapping with the other graders. Thus, the only discernible difference between the four sections of introductory microeconomics was that students in three sections had a direct grade incentive to practice economics problems throughout the semester (the treatment group), whereas those in the control group received neither a direct reward for completing problem sets nor a penalty for failing to do so.

The data for this study come from a natural, rather than a perfectly randomized, experiment since students enrolled in individual sections of Economics 101 but did not know *ex ante* (neither did the professor) whether they had enrolled in an experimental group. We find no evidence of sample selection bias regarding either class size or the sixteen students (out of the seventy-five) in the non-graded group who had taken Principles of Macroeconomics the previous semester with the professor who fell seriously ill (see Grove and Wasserman, 2005). Twenty-seven students in each group dropped the course. The average attriter was academically weaker, measured by that semester's GPA. With fewer non-graded group students, a higher fraction of them attrited (22 percent compared with 16 percent in the treatment group); thus, if anything, attrition biases down the treatment effect. With differences in right hand side variables

controlled for, drop outs may have caused sample selection bias if some other unmeasured component of ability caused attrition.

No SAT scores are available for forty-one students, reducing the sample with complete data from 239 to 198 students (i.e., 20 from the graded and 21 from the non-graded group). To correct for these missing observations, a dummy variable was created and set equal to zero for missing SAT scores and equal to one for students with SAT scores. The estimated coefficient for each dummy variable will represent the product of the average imputed value of the missing observation and its estimated coefficient. By retaining in the sample those records which have missing observations, this procedure makes it possible to use the information provided by the observations on the other variables in those records. Furthermore, the regression results are similar to and lead to identical conclusions as obtained when records with missing SAT scores were omitted.

II. Empirical Methodology

Motivating our work with an educational production function approach that is standard in the literature, we specify the following reduced-form model:

- (1) Student learning = $f(\text{academic ability; demographic characteristics; membership in a fraternity or sorority; membership in the experimental group})$.

Our academic ability measures take two forms: collegiate GPA and math SAT scores (having shown elsewhere that they are the best control variables for performance in economics courses, Grove et al., 2006). Math SAT scores provide a standardized measure of individual knowledge and cognitive aptitude at a point in time in the past, while collegiate grade point average (GPA) reflects both cognitive ability and noncognitive skills, such as motivation,

organization, perseverance, industriousness, consistency and study skills. Since GPA may poorly measure comparative academic performance due to the variability of grade distributions by faculty members, courses, and departments, we standardized semester GPA data (creating a z-score *for every class* students in our sample took that semester by subtracting raw grades in each course from the sample mean course grade and dividing that by the standard deviation of the course grades). Our measure of collegiate academic performance, zSemGPA-ECN, is students' standardized GPA data minus their Economics 101 grade the semester the experiment occurred.

Although the evidence is mixed, most studies of college economics find that first year students perform worse than upperclassmen (Siegfried and Walstad, 1998, p. 151) and male students outperformed females on exams (ibid, 150). Durden and Ellis (1995) report that membership in a sorority or fraternity negatively affects academic achievement.

Thus, hypothesizing that the incentive to practice economics throughout the semester improved college students' exam performance, we write our empirical model as:

$$(2) \quad \text{MeanExamScore}_i = \alpha + \beta(\text{GradedGroup})_i + \gamma \mathbf{X}_i + \epsilon_i$$

where \mathbf{X}_i includes controls for student ability and demographics (standardized current semester college GPA without the economics grade, SAT math scores, whether the student's record contained an SAT math score, race, gender, year in college, and membership in a Greek organization).

III. Results

We provide the results of five specifications of our estimated graded group treatment effect, obtained using ordinary least squares, in Table 1. Conditioning on Greek participation, academic aptitude measures, and demographic characteristics, students with a grade incentive to study (the treatment effect) increased their average exam performance by an estimated 2.27 points on a 100 point scale, statistically significant at the 1-percent level (Table 1, column 1). Although 2.27 points constitute only a quarter of a 10-point letter grade, this treatment effect boosted the mean students' course grade from a C+ to a B-, by a third of a letter grade. Both measures of aptitude are strongly, positively associated with student learning. In addition, freshmen, sophomores and members of the Greek system scored significantly below others in Economics 101 (at the 1-percent level), other things equal (Table 1, column 1). Using the sample of students with SAT scores (N=198) (rather than the full sample of student data with imputed SAT math scores, N=239), we obtain a slightly larger treatment effect of 2.84 points with otherwise essentially identical coefficient estimates (not shown in Table 1; see Grove and Wasserman, 2005).

To test whether students of different academic aptitude benefit differently from graded assignments, as suggested by existing studies, in Specification 2 we include a term interacting membership in the treatment group with students' GPA. The coefficient on that interaction term is insignificant, quite small and, most importantly, does not alter the magnitude or significance of the graded group variable; thus, we find no evidence that either academically-above or -below average students benefited more from graded problem sets, controlling for aptitude and demographic variables. Nonetheless, we probe whether students with different levels of academic success responded differently to this treatment by separately conducting the sam

analysis but of the treatment sample and of the control sample separately and including a below-average GPA variable. Doing so, we find that students with below average grades may benefit more: the coefficient on the term indicating whether a student had below average GPA is twice as large and strongly significant for students without a grade incentive to study (control group), whereas treatment group students with the grade incentive had a smaller and much less significant coefficient (-5.03, t statistic = -2.55; -2.74, t statistic = -1.68, respectively).

To evaluate whether and how students with different demographic characteristics responded to the grade incentive to practice, we interact, in succession, membership in the treatment group with each of the demographic variables, Greek system membership, and year in college. Only freshmen, according to our results, responded significantly differently than others to the grade incentive to practice. Our estimates indicate that first year students in the experimental group performed 4.58 points higher on exams (Table 1, column 3) relative to freshmen in the control group, *ceteris paribus*; thus, while the average freshman scored 6.7 points worse than others on exams, first year students with a grade incentive to complete the assignments, on average, made up two-thirds of their disadvantage of being new to college. Inclusion of the freshman-graded group interaction term caused the general treatment group effect to lose its significance which did not occur when adding any other variables or interaction terms.

As a check on the freshmen treatment effect results, Specification 4 replicates Specification 3 using only the sample of students with SAT scores (N=198). The coefficient on the freshman-graded group interaction term is 3.74 points, 18 percent less than 4.58 points reported above for the full sample, and the other estimated coefficients results are essentially the same in Specifications 3 and 4. Finally, including both the freshmen and the GPA interaction

terms in Specification 5, we obtain estimates for each term similar to those in Specifications 2 and 3: no evidence that the treatment effect varies according to students' academic performance that semester but strong evidence that freshmen improved their cognitive achievement due to the grade incentive to study.

These impressive and meaningfully large results represent lower bound estimates for several reasons. We only measure group membership rather than changes in individual behavior (since no information was collected about the quantity or quality of non-graded group students' assignments). Many students would have completed all assignments without the grade incentive and that reward failed to induce 10 percent of the treatment students to hand in the problem sets. Although the same number of students in both groups dropped the course, a higher fraction of students without graded homework did so.

Because we do not know which control group members completed problem sets, we can not use our natural experiment data to determine whether the grade incentive to study induced first year students' learning gains via increased productivity or harder work. Focusing exclusively on the graded group sample, we can investigate, though, whether freshmen were more likely than others to complete all five assignments, even though only the best four affected their grade. We find that first year students were not more likely than treatment group students in other years to hand in all five problem sets either unconditionally or when regressing the independent variables in Table 1 and each year in college on whether or not a student handed in all five assignments (using both OLS and logistic analysis). Since freshmen appear not to have responded to the grade incentive by working harder than other students, we conjecture, based on this limited evidence, that their learning gains may have resulted from more productive studying.

IV. Conclusions

Academics know surprisingly little about how to increase college student learning, largely due to the difficulty of establishing a control group, free of selection problems, against which one can assess a particular instructional choice. A natural experiment permits us to analyze the effect of graded versus non-graded problem set assignments on cognitive achievement in a college introductory microeconomics course. Using a rich set of institutional and instructor data, we estimate that the grade incentive to practice throughout the semester boosted the average freshman student's exam performance, not that of academically below average or any other category of student. The freshmen treatment effect learning gain of 3.7-4.6 points increased their mean grades from a C+ to a B-, *ceteris paribus*, relative to those in the control group. With less certainty, we conclude that the grade incentive increased the productivity of freshman studying, rather than inducing them to devote more time to economics. These statistically significant and meaningful results offer an exception to Eric Hanusheck's (1991) suggestion that instructional variables are unimportant in explaining student learning but, most importantly, identify one that can be implemented by the typical professor.

Although scholars often have found freshmen to perform worse in introductory economics courses, like so many of the important student learning results, that insight lead to no satisfactory policy solutions. An administrative ban on freshman in economics courses, for example, would cause difficulties for sequentially structured majors. By contrast, encouraging professors to assign problem sets (with, for example, institutional funding for graders) offers a viable response to concerns about freshman retention.

Since our data come from a single professor's classes in one semester it will be important for other scholars to determine the generalizability of these findings. Given the obstacles and

reluctance of academics to conduct randomized experiments, we hope faculty will recognize the importance of classroom natural experiment data sets for testing the efficacy of teaching methods and assignments.

REFERENCES

- Becker, William E. "Teaching Economics to Undergraduates." *Journal of Economic Literature*, September 1997, 35(3), pp. 1347-73.
- _____. "Obstacles to the Scholarship of Teaching and Learning Economics." 2002 Annual Report of the *Journal of Economic Education* (<http://www.indiana.edu/~econed/anrpts/anrpt02/rep02.htm>).
- Betts, Julian. "The Role of Homework in Improving School Quality." Discussion Paper 96-16, Department of Economics, UCSD, 1996.
- Borg, Mary O., Paul M. Mason and Stephen L. Shapiro. "The Case of Effort Variables in Student Performance." *Journal of Economic Education*, Summer 1989, 20(3), pp. 308-313.
- Chizmar, John F. and Anthony L. Ostrosky. "The One-Minute Paper: Some Empirical Findings." *Journal of Economic Education*, Winter 1998, 29(1), pp. 3-10.
- Durden, G. C., and L. V. Ellis. "The Effects of Attendance on Student Learning in Principles of Economics." *American Economic Review*, May 1995 (Papers and Proceedings), 85(2), pp. 343-46.
- Fels, Rendigs. "This is what I do, and I like it." *Journal of Economic Education*, Fall 1993, 24(4), pp. 365-70.
- Emerson, Trisha L. N. and Beck A. Taylor. "Comparing Student Achievement across Experimental and Lecture-Oriented Sections of a Principles of Microeconomics Course." *Southern Economic Journal*, January 2004, 70(3), pp. 672-93
- Grove, Wayne and Tim Wasserman. "Incentives and Student Learning." Working paper, Le Moyne College, Management Division, 2005.
- Grove, Wayne, Tim Wasserman and Andrew Grodner. "Choosing a Proxy Academic Aptitude." *Journal of Economic Education*, Spring 2006 (forthcoming).
- Hanushek, Eric. "When school finance 'reform' may not be a good policy." *Harvard Journal of Legislation*, 1991, 28(2), pp. 423-56.
- Kennedy, Peter and John J. Siegfried. "Class size and achievement in introductory economics: Evidence from the TUCE III data." *Economic of Education Review*, October 1997, 16(4), pp. 385-94.

- Kirby, Kris N., Gordon C. Winston, and Mariana Santiesteban. "Impatience and grades: Delay-discount rates correlate negatively with college GPA." Williams Project on the Economics of Higher Education Discussion Paper, May 2002.
- Marlin, James W. and James F. Niss. "The advanced learning system, a computer-managed, self-paced system of instruction: An application in principles of economics." *Journal of Economic Education*, Summer 1982, 13(2), pp. 26-39.
- Pozo, Susan and Charles Stull. "Randomized Trials in Economics Education: Do Graded Math Units Improve Student Performance in Introductory Economics?" Mimeo, Western Michigan University, 2005.
- Salemi, Michael K. and George E. Tauchen. "Estimation of Nonlinear Learning Models." *Journal of American Statistical Association*, December 1982, 77(380), pp. 725-31.
- Siegfried, John J. and William B. Walstad. "Research on Teaching College Economics," in William B. Walstad and Philip Saunders, eds, *The Principles of Economics Course: A Handbook for Instructors*. NY: McGrawHill, 1998, pp. 141-66.

Table 1					
Estimated Graded Group Treatment Effects					
Dependent Variable: Mean Exam Score (in points)					
Independent Variables	Specifications				
	(1)	(2)	(3)	(4)	(5)
Intercept	61.78** (17.55)	61.61** (17.38)	63.03** (17.95)	62.72** (17.99)	62.83** (17.79)
zSemGPA-Econ	7.37** (13.43)	7.67** (9.03)	7.30** (13.46)	7.10** (11.80)	7.67** (9.14)
SAT-math	0.04** (5.89)	0.04** (5.88)	0.04** (5.98)	0.04** (6.04)	0.04** (5.98)
Without Sat-math	16.83** (4.60)	17.04** (4.61)	16.52** (4.57)		16.76** (4.59)
Male	0.86 (1.03)	0.85 (1.01)	0.68 (0.82)	0.68 (0.76)	0.66 (0.79)
White	-1.08 (-1.20)	-1.13 (-1.25)	-1.11 (-1.25)	-0.89 (-0.92)	-1.17 (-1.31)
Freshman	-3.36** (-2.73)	-3.39** (-2.74)	-6.68* (-3.77)	-6.02** (-3.15)	-6.74** (-3.79)
Sophomore	-3.47** (-3.04)	-3.50** (-3.06)	-3.86** (-3.39)	-3.98** (-3.03)	-3.89** (-3.41)
Fraternity/Sorority Member	-1.97* (-2.11)	-1.97* (-2.10)	-2.10* (-2.26)	-2.60** (-2.79)	-2.09* (-2.26)
Graded Group	2.27** (2.77)	2.25** (2.74)	0.88 (0.90)	1.54 (1.43)	0.84 (0.86)
Graded Group X zSemGPA-Econ		-0.51 (-0.47)			-0.61 (-0.57)
Graded Group X Freshman			4.58* (2.58)	3.74* (2.05)	4.62* (2.59)
Observations	239	239	239	198	239
Adjusted R ²	0.59	0.59	0.60	0.60	0.60

Note: t-statistics are in parentheses.
* Statistically significant at the 5-percent level.
** Statistically significant at the 1-percent level.

* Grove: Associate Professor, Department of Economics, Le Moyne College, Syracuse, NY, 13214 (grovewa@lemoyne.edu); Wasserman, Assistant Director for Institutional Research and Data Management, Center for Support of Teaching & Learning, Syracuse University, Syracuse, NY, 13244 (thwasser@syr.edu). We wish to acknowledge the helpful comments offered by Paul Romer, Dan Hamermesh, Jerry Evensky, Doug Holtz-Eakin, Dixie Blackley, Andrew Grodner, Paul Blackley, Kevin Rask, and by participants at the University of Illinois-Chicago and Colgate University/Hamilton College workshops. Wayne Grove thanks Le Moyne College for a Research and Development grant.

¹ For surveys of the research on teaching college economics, see Becker (1997) and Siegfried and Walstad (1998). Economics education research has been fostered since the 1950s by an annual session(s) of the American Economics Association meetings and since 1969 with a specialty journal, *Journal of Economics Education*.

² The mean number of courses with assigned problem sets varied from 48 percent for research institutions, 37 for doctoral granting programs, 38 percent for masters programs, 34 percent for liberal arts schools, and 29 for associate institutions (see Becker, 1997, Table 3, 1352).

³ Important exceptions include three experimental studies that suggest that student learning may improve with use of the one-minute paper (see Chizmar and Ostrosky, 1998), better math preparation (Poza and Stull, 2005) and classroom experiments (see Emerson and Taylor, 2004).

⁴ As part of the reauthorization of the Higher Education Act, both the House of Representatives and the President have threatened to reduce funding for higher education without progress on retention and graduation rate (correspondence with June Kronholz, reporter for the *Wall Street Journal*, on August 20, 2003).

⁵ The department chair asked one of the authors to teach the additional section of Economics 101 for a few class periods. At the end of the second week of classes it became apparent that the hospitalized professor could not return to the classroom soon. At that point in the semester after the drop period had passed, the newly-assigned professor thought it inappropriate to impose new requirements.