

Play to Learn? An Experiment

Martin Dufwenberg^a & J. Todd Swarthout^b

April 29, 2009

Abstract: Does playing a game in class improve students' ability to analyze the game using game theory? We report results from an experimental design which allows us to test a series of related hypotheses. We fail to find support for the conjectured learning-enhancing effects and discuss what lessons can be learned substantially and methodologically.

JEL codes: A22, C70

1. Introduction

What classroom activities enhance learning? We report experimental evidence addressing a special case of this question: Does playing a game in class, before attending a lecture on how to conduct a game theoretic analysis, improve students' performance solving the game?

For an intermediate microeconomics course that one of us taught, we picked *two* interesting 2×2 games to lecture about in class. Before the lecturing we ran a classroom experiment on *one* of the games in class, randomizing the choice of that game across the two sections. Come the exam, students were asked to: 1. solve both games for their Nash equilibria; and 2. choose one of the two games to count for triple credit. This design allows us to test a series of hypotheses about how game play fosters actual learning of theory, as well as students' perception of which problem is easiest to solve.

Section 2 describes our procedures. Section 3 presents the research questions the design is meant to explore, and the data analysis. Section 4 provides a concluding discussion. The rest of this introduction connects our exercise to previous literature.

^a Department of Economics, University of Arizona, martind@eller.arizona.edu

^b Department of Economics, Georgia State University, swarthout@gsu.edu

Economics instructors have been conducting classroom experiments for decades. One of the first reports of this practice was Chamberlin (1948), which discusses the pedagogic benefits of conducting classroom market experiments. In the years since, instructors have increasingly adopted class experiments as an active learning activity for their students. In particular, the 1990s saw significant growth in the use of classroom experiments: Wells (1991) discusses the use of computerized instructional market experiments, and in 1993 a special issue of the *Journal of Economic Education* was devoted entirely to classroom experimental economics. Publication of classroom instructional experiments subsequently became common in journals such as the *Journal of Economic Education*, *Economic Inquiry*, *Southern Economic Journal*, *Journal of Economic Perspectives*, and *Experimental Economics*. Further, in recent years textbooks are emerging which use class experiments as a central device for teaching economics concepts (Bergstrom and Miller, 1999; Holt, 2006; Schotter, 2008).

Despite the increasing use of economics experiments in the classroom, relatively little empirical evidence has been presented to support the pedagogic benefit of the approach. We speculate that many instructors elect to use class experiments simply because they personally view the activity as beneficial, and not because of evidence supporting the pedagogic value of the approach. But how might classroom experiments benefit student learning? One way is that the activity may generally improve student aptitude in the area of study, as measured by standardized tests or overall course performance. This is the evaluation approach generally taken by previous efficacy studies. Another way in which class experiments may improve student learning is by exposing students to specific scenarios that subsequently provide a benefit when solving a problem directly related to the scenario. This can be measured by designing test questions specifically tailored to the details of the class experiment. This is the approach we

adopt in our current study. But, before discussing our current study, we will first provide a brief survey of the prior classroom experiment efficacy studies.

Frank (1997) reports a study in which seven sections of environmental economics and public finance courses participated in a “tragedy of the commons” classroom experiment. Some students participated in the experiment before taking a test on the topic, while other students took the test without experiment participation. Limited support is provided that experiment participation improved test scores.

Gremmen and Potters (1997) report a study which implements a multi-day international economic simulation game. Students were randomly split into two groups, and either played the game during the semester or instead followed traditional lectures. Pretest results show both groups scoring about the same, while the group playing the game scored significantly higher on the posttest than the lecture group.

The two prior studies discussed above tested only a single class activity. Cardell, et al. (1996) and Dickie (2000), on the other hand, examine the use of multiple experiments in the principles classroom. In a sample of about 1,800 students, Cardell et al. find no statistically significant impact of the experimental approach on student achievement (as measured by differences in post- and pre-course scores on the Test of Understanding in College Economics, or TUCE). Using a much smaller sample of 142 students, Dickie does find statistically significant positive effects on student learning generally, with some evidence that higher-achieving students (i.e., those with higher GPAs) experience the largest benefits from the experimental approach.

One of the more comprehensive studies of the efficacy of economics classroom experiments is reported by Emerson and Taylor (2004). They describe a comparison between experiment- and lecture-oriented sections of a microeconomics principles course. The authors

administered the microeconomics portion of the TUCE at the start and end of the semester to assess student learning during the semester. The study compares two experiment sections (59 students), each of which used eleven pencil and paper experiments over the course of the semester that were taken from Bergstrom and Miller (2000), with seven non-experiment sections (241 students) that served as controls. The main finding is a significant improvement in TUCE scores for the experiment group: improvements of post-course over pre-course TUCE scores for the control group are similar to the national average, whereas improvements for the experiment group were 77 percent higher than improvements for the control group and 60 percent higher than the national average improvement.

Another comprehensive study of the efficacy of economics experiments in the classroom is reported by Durham, McKinnon, and Schulman (2007). The authors include a total of 1585 students from sixteen class sections in their study: two introductory microeconomics sections and two introductory macroeconomics sections per semester for four semesters. Many potential confounds were controlled, including instructor, time-of-day, and time-on-task effects. The authors initially developed both course materials and an assessment instrument to assess the efficacy of the course materials. Elements of the assessment instrument were administered to all students as part of both mid-term exams and a final exam. Relative performance between the control and treatment groups as measured by the instrument provides a test of the effect of classroom economics experiments on student learning. Additional student measurements were administered, including the VARK learning style questionnaire at the beginning of the semester, and an attitude survey both at the beginning and end of the semester.

Durham et al. conclude that classroom experiments improve student performance on test questions over the topics that the experiments were designed to cover, both for microeconomics

and macroeconomics. Additionally, the authors find that exposure to economics experiments positively influences student attitudes towards the study of economics – a finding also supported by Emerson and Taylor. Finally, the data suggest that the benefit of classroom experiments varies depending on a student’s preferred learning style: kinesthetic and multimodal learners (which make up the “vast majority” of students in the study) are significantly positively affected by the use of class experiments in both the microeconomic and macroeconomic sections.

Although prior efficacy studies have already been conducted to evaluate the benefit of classroom economics experiments, our new study reported herein adds to the existing literature in several ways. First, the preceding efficacy studies discussed above were conducted in introductory principles courses—except for Frank (1997), which used public finance and environmental economics courses.¹ We conduct our study with intermediate microeconomics students. Second, unlike prior studies which more broadly measure students’ economics aptitude, we instead more narrowly measure student performance on test questions directly related to the details of the experiments. Third, we present a novel protocol which differs from prior studies.

2. Procedures

One of us was teaching two morning sections of an Intermediate Microeconomics course at a large southwestern university. Our design revolves around the following two games:

<i>Game 1:</i>	c	d	<i>Game 2:</i>	g	h
a	7, 7	0, 6	e	0, 0	1, -x
b	6, 0	5, 5	f	1, 0	0, 1

¹ Cardell et al. did not explicitly state the type of economics class.

Our games are selected on the grounds that they are easy but not very easy to solve for Nash equilibria, and in addition that they offer some scope for discussion of exciting aspects beyond that. Game 1 is a so-called stag-hunt game, often used to discuss equilibrium selection and the role of cheap talk in this connection (see e.g. Aumann 1990; Farrell & Rabin 1996; Charness 2000). Game 2 is a kind of inspection game (cf Avenhaus, von Stengel & Zamir, 2002): the row player is a police who can inspect (e) or do paperwork (f); the column player is a thief who can stay at home (g) or rob (h); $x > 0$ is number of years in jail; etc.

Upon entering the classroom, students were invited to play a game. In Section 1 (treatment 1) they first played Game 1 once. Then they were asked to play that game once more, this time with a 3-minute pre-play communication phase. The overall purpose was to provide a fun and not too short episode of interaction, and to give the students time to reflect on the role of communication in ways which we figured could instill a *déjà-vu* experience during the upcoming lecture about Game 1. The students in Section 1 never played Game 2.

In Section 2 (treatment 2) they first played Game 2 once, with $x=5$; then they were asked to play the game once more, this time with $x=.2$.² The overall purpose was to provide a fun and not too short episode of interaction, and reasons to reflect on how incentives may change with x in ways which may instill a *déjà-vu* experience during the upcoming lecture about Game 2. The students in Section 2 never played Game 1.

In each section game play was followed by identical lectures covering theory for *both* games. For Game 1 we discussed equilibrium multiplicity (there are three Nash equilibria, one in mixed strategies) and the impact of pre-play communication in this connection. For Game 2 we derived the unique Nash equilibrium; it is in mixed strategies, it has player 1's strategy depend on

² The payoff numbers in the experiment actually added a constant of 5 to each player's payoff in each cell. When we later on lectured, we talked about how the game-theoretic solution remains the same if a constant of 5 is subtracted from each player's payoff in each cell, to get Game 2 above.

x , and the police-and-thief angle admits the intriguing criminological interpretation that harsher punishment does not reduce crime (in equilibrium). Talking about Games 1 and 2, with these spins, wraps up a fun class!

A few weeks later it was time for the midterm exam. This is where we got our data, via Question 3 which read as follows:

Question 3. (5+15=20 points) *This question asks you to analyze two games. You may influence your score as follows: One of the games is worth 5 points; the other game is worth 15 points. It is up to you to indicate which game is which on the next page (not on this page)!*

Now we move the question itself. For each of the following two games: (i) Find all Nash equilibria in pure strategies. (ii) Find all Nash equilibria in mixed strategies. Motivate your answers.

<u>Game 1:</u>	c	d	<u>Game 2:</u>	g	h
a	7, 7	0, 6	e	0, 0	1, - x
b	6, 0	5, 5	f	1, 0	0, 1

[In Game 2, x is a positive number (so that $-x$ is a negative number). You should provide an answer for each possible value of x .]

Please choose which of the games [...] should be worth 15 and which should be worth 5 points. Check *one* box below:

- | | |
|---|---|
| <input type="checkbox"/> <i>Game 1 worth 15 points</i>
<input type="checkbox"/> <i>Game 2 worth 5 points</i> | <input type="checkbox"/> <i>Game 2 worth 15 points</i>
<input type="checkbox"/> <i>Game 1 worth 5 points</i> |
|---|---|

Note: You should check *exactly one* box. (If you check no box, or if you check both boxes, the game for which your answer was the worst will be worth 15 points, and the other game will be worth 5 points.)

By asking the students to solve both games we can evaluate their game-theoretic performance on each. By asking them to select one of the games for triple credit we can evaluate

their perception of which problem is easiest to solve. Through our randomized assignment of treatment to section we can evaluate the impact of game-play exposure to performance and perception. We elaborate on a series of associated hypotheses in the next section.

3. Results

In this section we present our research questions in sequence, along with supporting data analysis and interpretation. Only students who participated in one of the experiments and also took the midterm are included in the analysis. The sample size of section 1 is 38 students, and the sample size of section 2 is 37 students.

Does participation in a class experiment lead to higher scores on the associated test question?

Combining traditional lecture with participation in a classroom experiment to highlight a topic may provide students with greater insight into the material than relying on only a lecture approach. Table 1 reports the results of three tests to evaluate whether test scores on question 3.1 are higher in the section which participated in the associated classroom experiment.³ First, we note that the mean of the question 3.1 score is actually lower in the section which played Game 1. All three statistical tests support the conclusion that participation in Game 1 does not lead to higher scores on question 3.1: a t-test of difference in means results in a p-value of 0.7619, a Wilcoxon rank sum test of a difference in medians results in a p-value of 0.8296, and a Kolmogorov-Smirnov test of a difference in distributions results in a p-value of 0.9989.

³ We denote the portion of question 3 devoted to solving game 1 as question 3.1. Similarly, we also refer to question 3.2. Additionally, all comparisons of questions 3.1 and 3.2 make use of only raw scores (out of 5 points), and do not include any tripled scores.

TABLE 1 ABOUT HERE

We now assess if participation in game 2 is related to higher scores on question 3.2. Table 2 reports the results of three one-sided tests. Again, we see no support that participation in the classroom game relates with higher score on the associated test question. As before, the section not participating in the experiment had a higher mean score. Further, the t-test yields a p-value of 0.5490, the Wilcoxon test provides a p-value of 0.5791, and the Kolmogorov-Smirnow test show a p-value of 0.8145.

TABLE 2 ABOUT HERE

Based on these two-sample tests, we are initially led to conclude that participation in the classroom experiments does not result in higher scores on test questions associated with the experiments.

Is participation in a particular game related with selecting the associated question for higher weight?

Perhaps participation in a classroom game may influence a student's perception of her ability to answer the related test question. So, even if performance is not ultimately influenced by the experiment participation, we might expect to see a student's subjective assessment of ability to be altered by exposure to the experiment. We will test for this by examining student choices on whether question 3.1 or 3.2 is chosen to receive triple weight.

Recall that initially both question 3.1 and 3.2 were worth up to 5 points each, and each student was required to specify which of these two questions would count for triple weight. In section 1, we see 8 of 38 students (21.1 percent) selecting question 3.2 for additional weight. And in section 2, we see 3 of 37 (8.1 percent) selecting question 3.2. A Chi-squared test for a treatment difference results in a test statistic of 1.5821 with an associated p-value of 0.2085. We conclude that there is no difference in this choice across treatments, and that exposure to playing in an experiment did not increase the likelihood of selecting the associated question for higher weight.

Does general student aptitude influence the impact of experiment participation on test scores?

Perhaps general aptitude has an interaction with the influence of experiment participation on the test scores. To address this question, we use an adjusted midterm test score (total test score minus question 3 points) as a proxy for student aptitude in the course, and conduct regression analysis. Table 3 reports the results of OLS estimation of question 3.1 raw score as a function of a section indicator, an adjusted midterm score, and an interaction term between the section indicator and adjusted midterm score. We see that the interaction term is not statistically significant in either model. In fact, the only term which is statistically significant is the adjusted test score coefficient, and this holds across both models.

TABLE 3 ABOUT HERE

4. Discussion

We have tested whether playing a game in class enhances students' ability to analyze that game game-theoretically, and not found support. Coming into this research we had a strong prior that we would find support, so we are surprised. Should we conclude that playing a game in class does not enhance students' game-theoretic ability? We feel that such a conclusion would be premature. We have several reasons. First, given our strong prior it will take more evidence to sway us. Perhaps our finding is but a statistical fluke?

Second, perhaps our ideas of how game play influences game-theoretic ability have been too blunt? Maybe details about a game's structure sometimes matter so that while on balance playing a game improves analytic ability this need not apply for every game? Game 2 can illustrate. It may seem intuitive that increasing x makes choosing h less attractive to player 2; indeed, effects of that sort have been observed in experiments (e.g. Ochs, 1995). This intuition is, however, not captured via the Nash equilibrium where player 2's probability of choosing h equals $1/2$ independent of x . Perhaps playing the game makes the "lower- x -boosts- h " intuition come alive at the expense of the logic of Nash equilibrium? Such an effect could conceivably counteract the positive effects of game play on game-theoretic understanding for Game 2.

Third, a confound for our tests may come into play if the two games we study are not equally difficult for students to analyze. Say, for example, that analyzing Game 2 is harder than analyzing Game 1. Perhaps playing Game 2 helps student realize this. This could explain why playing Game 2 does not make students more likely to select Game 2 for triple credit. In light of this reasoning we ran a test to examine whether one or the other question was unconditionally more difficult. Scores on question 3b are generally lower than scores on question 3a. We conducted a Wilcoxon signed rank test to assess how systematic this is at the individual student

level. We test the difference of question 3b minus question 3a for each student; the test statistic is 312, with an associated p-value of 0.0012. This strongly suggests that question 3b is the unconditionally more difficult question. Indeed, 64 of the 75 students in the study scored higher on question 3a than 3b.

However, we must also acknowledge the possibility that participation in a classroom experiment does not improve performance on a test question directly related to the game-theoretic details of the experiment. Perhaps the way by which some classroom experiments improve student performance is indirect: the experiments serve to better engage students in the material and make the students generally more interested in the course. This, in turn, leads to better overall performance in the class, as measured by relatively general aptitude measures. Further, if this logic holds, then this indirect benefit from class experiments would not be necessarily expected to directly impact student performance on a question related specifically to the mechanics of the experiment. This conjecture could help to explain why many of the prior efficacy studies found a benefit to student learning from class experiments, yet we did not: both Emerson and Taylor (2004) and Durham et al. (2007) exposed students to many class experiments during the semester and then measured general aptitude, while we conducted a single class experiment and then measured performance on a task concerned with the details of the experiment.

Working on this project we have come to think that the question of how game play enhancing learning may be multi-faceted and perhaps our study generates more questions than it answers. We nevertheless hope it will inspire future work in the area.

REFERENCES

Avenhaus, R., B. von Stengel & S. Zamir (2002), "Inspection games". Chapter 51, *Handbook of Game Theory*, Vol. 3 (eds. R. J. Aumann & S. Hart), North-Holland, Amsterdam, 1947-1987.

Aumann, R. (1990), "Nash Equilibria are not Self-Enforcing", in *Economic Decision-Making: Games, Econometrics and Optimisation*, Gabszewicz et al (eds.), Elsevier.

Bergstrom, T.C. & J. H. Miller. 1999. *Experiments with Economic Principles: Microeconomics*. New York: McGraw Hill.

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

Chamberlin, Edward H. (1948) "An Experimental Imperfect Market," *Journal of Political Economy*, 56: 95-108.

Charness, G. (2000) "Self-serving Cheap Talk and Credibility: A Test of Aumann's Conjecture," *Games & Economic Behavior* 33, 177-194.

Dickie, Mark (2006) "Do Classroom Experiments Increase Learning in Introductory Microeconomics?" *Journal of Economic Education*, 37: 267 - 288.

Durham, McKinnon & Schulman (2007) "Classroom Experiments: Not Just Fun and Games," *Economic Inquiry*, 45: 162-178.

Emerson, Tisha L.N. & Beck A. Taylor (2004) "Comparing Student Achievement across Experimental and Lecture-Oriented Sections of a Principles of Microeconomics Course," *Southern Economic Journal*, 70: 672-93.

Farrell, J. & M. Rabin (1996) "Cheap Talk", *Journal of Economic Perspectives* 10, 103-18.

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

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

Holt, Charles A. (2006) *Markets, Games, and Strategic Behavior*. Addison Wesley.

Ochs, Jack (1995) "Games with Unique, Mixed Strategy Equilibria: An Experimental Study," *Games and Economic Behavior* 10, 202-217.

Schotter, Andrew (2008) *Microeconomics: A Modern Approach* (new 1st ed.), South-Western Publishing.

Wells, Donald A. (1991) "Laboratory Experiments for Undergraduate Instruction in Economics," *Journal of Economic Education*, v. 22, iss. 3, pp. 293-300.

Table 1. Tests of whether section 1 scores higher on question 3.1 than section B (out of five points)

Section	Mean of question 3.1	t-test (p-value)	Wilcoxon (p-value)	Kolmogorov-Smirnov (p-value)
1	2.92	-0.7162	617	0.0156
2	3.19	(0.7619)	(0.8296)	(0.9989)

Each test was one-sided.

Table 2. Tests of whether section 2 scores higher on question 3.2 than section 1 (out of five points)

Section	Mean of question 3.2	t-test (p-value)	Wilcoxon (p-value)	Kolmogorov-Smirnov (p-value)
1	2.54	-0.1235	685	0.074
2	2.49	(0.5490)	(0.5791)	(0.8145)

Each test was one-sided.

Table 3. OLS models of question 3 performance. P-values in parentheses.

Coefficient	Question 3.1	Question 3.2
Section 2	-0.50206 (0.65700)	-1.53997 (0.26819)
Adjusted test score	0.05139 (0.00001) ***	0.04525 (0.00118) **
Interaction: Section 2 \times Adjusted test score	0.01060 (0.56500)	0.02281 (0.31360)
Constant	-0.02955 (0.96500)	-0.05909 (0.94219)
Observations	75	75

* Statistical significance at the five percent level

** Statistical significance at the one percent level

*** Statistical significance at the one-tenth of one percent level