Do Monetary Incentives Matter in Classroom Experiments?  
Effects on Course Performance

Running Head: Monetary Incentives in Classroom Experiments

Matthew C. Rousu*  
Professor of Economics, Susquehanna University, (570) 372-4186, rousu@susqu.edu

Jay R. Corrigan  
Professor of Economics, Kenyon College, (740) 427-5281, corriganj@kenyon.edu

David Harris  
Assistant Professor of Economics, Benedictine College, (913) 360-7456, dharris@benedictine.edu

Jill K. Hayter  
Assistant Professor of Economics, East Tennessee State University, (423) 439-5926, hayter@etsu.edu

Scott Houser  
Teaching Associate Professor of Economics, Colorado School of Mines, (303) 384-2045, shouser@mines.edu

Becky A. Lafrancois  
Teaching Associate Professor of Economics, Colorado School of Mines, (303) 384-2191, blafanc@mines.edu

Olugbenga Onafowora  
Professor of Economics, Susquehanna University, (570) 372-4187, onafowor@susqu.edu

Gregory Colson  
Assistant Professor of Agricultural and Applied Economics, University of Georgia,  
(706) 583-0616, gcolson@uga.edu

Adam Hoffer  
Assistant Professor of Economics, University of Wisconsin, La Crosse, (608) 785-5293, ahoffer@uwla.x.edu

The authors are grateful for helpful comments at the 2013 AEA Conference on Teaching and Research on Teaching, the 2013 Southern Economics Association Conference, the 2014 ASSA conference in Philadelphia, the 2014 Appalachian Research in Business Symposium, and from seminar participants at Susquehanna University. The authors also thank two anonymous reviewers for their helpful feedback.

* Corresponding author.
ABSTRACT

Using 641 principles of economics students across four universities, we examine whether providing monetary incentives in a prisoner’s dilemma game enhances student learning as measured by a set of common exam questions. Subjects either play a two-player prisoner’s dilemma game for real money, play the same game with no money at stake (i.e., play a hypothetical version), or are in a control group where no game is played. We find strong evidence that students who played the classroom game for real money earned higher test scores than students who played the hypothetical game or where no game was played. Our findings challenge the conventional wisdom that monetary incentives are unnecessary in classroom experiments.

Key words: Classroom experiments, monetary incentives, course performance (JEL: A20, C90)

Beginning with the first classroom experiments on market efficiency by Chamberlin (1948), experiments have become an increasingly popular component of economics pedagogy, particularly in principles of economics courses. Just as in the case of laboratory experiments in the hard sciences such as chemistry and biology, experiments in economics are seen by many instructors as a valuable tool for reinforcing student learning of theoretical economic concepts via a fun and tangible in-class experience. The rise in popularity of economic experiments for the purpose of instruction is reflected by the broad range of in-class, face-to-face, and online computerized experiments that have been developed covering topics such as market efficiency, public goods, externalities, and game theory (Emerson and Hazlett 2012). These experiments are published in many outlets, including peer-reviewed journals (Corrigan 2011; Delemester and
Brauer 2000; Neral and Ray 1995), books (Bergstrom and Miller 2000; Hazlett 1999; Ortman and Colander 1995; Vazquez-Cognet et al. 2010), and also through Internet sites (see Charles Holt’s VeconLab).

Perhaps the most compelling reason for using classroom experiments is the empirical evidence that they can improve student learning. Emerson and Taylor (2004) found higher student test scores from a class that employed economic experiments extensively compared to a class that did not use experiments. Durham et al. (2007) found that experiments helped some students learn better, but the benefits of experiments on learning were correlated with students’ preferences for learning (e.g., read-write vs. visual). Carter and Emerson (2012) examined whether conducting experiments in the classroom as opposed to an online setting affected student learning. They found no differences in learning between on-line and in-class experiments; however, students enjoyed the experiments more when they were held in the classroom. Cartwright and Stepanova (2012) found learning from an experiment was much stronger when students had to write a report on the experiment after it had been implemented. Dickie (2006) reported pre-test and post-test results from classes taught by a single professor to assess the effects of conventional lectures, experiments without grade incentives, and experiments with grade incentives. He found that experiments without grade incentives lead to an increase in student performance, while classroom experiments associated with grade incentives decreased effectiveness of learning.

While there is evidence that classroom experiments can improve learning outcomes, there is a remarkable irony in how economists have implemented them. While the power of incentives is a central theme in economics principles classes,1 most classroom experiments do not employ monetary incentives. In fact, in his overview article titled “Teaching Economics with Classroom
Experiments,” Holt (1999:605) said “such [monetary] incentives are often unnecessary in classroom exercises.”

While there is a growing literature about how various incentives affect student performance in general (e.g. see Angrist et al. 2009, Angrist and Lavy 2009, Levitt et al. 2012, and Gneezy et al. 2011), there is a surprising lack of research on whether monetary incentives in economic experiments influence student learning. This void in the literature is surprising for two primary reasons. First, the impact of monetary incentives on the internal- and external-validity of experiments for non-instructional purposes has been a tremendously active area of research in recent years. While there is conflicting evidence on the impact of monetary incentives in experiments (Camerer and Hogarth 1999, Gneezy and Rustichini 2000, Smith and Walker 1993), the presence of issues such as hypothetical bias (Carlsson and Martinsson 2001, Harrison and Rutström 2008, List and Gallet 2001), and social desirability bias (Norwood and Lusk 2011), economic theory and experimental evidence generally suggests that even small monetary incentives can affect individual behavior. Given this evidence, it is conceivable that money may also matter in the context of learning outcomes. Second, many instructors employ incentives (e.g., candy, doughnuts, grades, money) at their own expense when carrying out classroom experiments without concrete evidence that their personal expenditures have succeeded at improving student outcomes. Without clear evidence on the returns to providing incentives in classroom experiments, instructors cannot know whether these expenses are a prudent investment in enhancing student learning.

In this study we contribute to the broader experimental literature, and specifically the economic education literature, by assessing the impact of monetary incentives on student learning. This study utilized controlled experiments administered by four professors at four
different universities to ensure results are robust and not a mere reflection of the peculiarities of one professor, group of students, or university. Subjects either played a two-player prisoner’s dilemma game for real money, played the same game with no money at stake (i.e., played a hypothetical version), or were in a control group that did not play the game. We expected that students playing the game for real monetary stakes would be more engaged in the prisoner’s dilemma game. This increased engagement could be attributed to the power of incentives or “reciprocal obligation” (Corrigan and Rousu 2006) where students who receive something from the instructor feel compelled to reciprocate by devoting more thought to the exercise. Increased engagement should lead to improved student learning. More specifically, we hypothesize that students in the real-money treatment will score higher on corresponding test questions compared to students playing the hypothetical game or students in the control group.

EXPERIMENTAL DESIGN

We assigned students in principles of economics classes to one of three experimental groups: a control group where students did not play a prisoner’s dilemma game, a group where students played a hypothetical version of the prisoner’s dilemma game, and a group where students played a prisoner’s dilemma game for real money. Professors selected which experimental group (treatment) would be assigned to each section of their course.

We attempted to assign treatments to sections in a random manner, but some concessions were made for tractability. For example, in one instance an instructor randomized across just the hypothetical and real treatments in the fall of 2012 so that she could keep both sections offered that semester at the same pace regarding the material she covered in class. Her students in the fall of 2013 then took part in the control treatment, which occupied less class time since students
didn’t play the game. That said, students did not know which treatment they would be in when they signed up for a particular section of economics (and indeed had no idea they were in a treatment or that an experiment would even take place), and professors did not know which students had signed up in advance of deciding how treatments were assigned to sections.

Students in the hypothetical and real money groups played a simple prisoner’s dilemma game with an anonymous partner. If both players cooperated, both received $3. If both players defected, both received $1. If one player cooperated and the other defected, the cooperating player received nothing and the defecting player received $5. The game was played either immediately prior to learning game theory or in the middle of the game theory lesson. In both the real money and hypothetical games, students were given immediate feedback as to the results of the experiment. In sections where the game was played with hypothetical payoffs, the professor asked the students to pretend they were playing for real money, but no financial rewards were actually provided. In the sections where the game was played for real money, the professor paid the students in cash at the end of the class period.

Six hundred forty-one students in principles of economics courses at four universities (Benedictine College, Colorado School of Mines, East Tennessee State University, and Kenyon College) took part in this study. These schools provide a broad geographical representation of schools in the U.S., have diverse characteristics (e.g., public vs. private, focus on engineering vs. humanities), and provide variation in class size for introductory economics courses.

We assess the impact of monetary incentives on student learning by examining how monetary incentives affect students’ scores on a series of assessment questions dealing specifically with game theory. The assessment presented students with a simple prisoner’s dilemma game and asked them to identify the dominant strategies, the Nash equilibrium
outcome, and to explain whether the Nash equilibrium outcome was Pareto efficient. All professors in this study used the same assessment questions and then examined the percentage of points each student earned on these questions. While each professor used the same questions, the specific way each professor chose to score each question varied based on the grading preferences of the individual professors. One professor used multiple choice questions, while the others put the question in the form of short answer questions.\(^3\)

Our experimental design has one major advantage over previous studies used to assess the impact of classroom experiments. By utilizing four professors at four universities to conduct this experiment, we ensure that results are not an artifact of a single professor’s teaching style or specific to a single university’s student composition. While many studies have used a single professor to run both the control and treatment groups, employing many professors across several institutions gives us greater confidence that our results are generalizable beyond our own classrooms or universities.

**ECONOMETRIC MODEL AND RESULTS**

Results presented in Table 1 show students in the real money group scored seven percentage points higher on the game theory assessment compared to those in the hypothetical group, and this result was statistically significant at the 5% level (\(t = 2.29\)). Those who played the game for real money also scored seven percentage points higher on the game theory assessment compared to those in the control group where no game was played, and this result was statistically significant at the 5% level (\(t = 2.27\)).\(^4\) There was no difference in scores between the hypothetical and control groups (\(t = 0.24\)). As a validity check, we also ran the same analysis for the students’ percentage score on the rest of the exam (i.e., exam score omitting the game
theory questions). Test scores on the rest of the exam were not statistically significantly different across any of the three groups, as the largest difference in the rest of exam scores was between the hypothetical and real money treatments (difference of 2 percentage points with $t = 1.30$).

[Insert Table 1 about here]

As a further check of the effectiveness of monetary incentives we compare the differences between the percentage of points the students earned on the game theory question versus the points they earned on the rest of the exam. The difference in the percentage of points earned for those in the control group and the hypothetical group was about one percentage point and those differences were not statistically significant at any conventional level. Those in the real money group scored five percentage points higher on the game theory question than on the rest of the exam and that difference was statistically significant at the 1% level ($t = 2.82$).

Tests examining the differences in game theory assessment scores do not show statistically significant differences in test scores between three treatments within any of our sample universities. However, there appears to be variation across universities, which we examine in later analysis. Specifically, students from all three experimental groups at the Colorado School of Mines (CSM) earned almost identical test scores (86%-89%), indicating that there was no observable return from in-class experiments or monetary incentives. Given that CSM is an engineering school where students have significantly more math training than at the other universities in the study, the uniformly high test scores come as no surprise. Although the study was not designed to test this hypothesis, we find evidence that the returns to an additional in-class game (real or hypothetical) are likely minimal for students who would otherwise perform well on an exam about game theory.
To further examine the impact of monetary incentives on learning while controlling for university and student effects we estimate four linear regressions. The dependent variable is students’ score on the assessment questions, and dummy variables are used to indicate experimental conditions. We assess whether participation in the hypothetical or real treatments affect student performance controlling for university effects by estimating:

\[
Score_i = \beta_0 + \beta_1 H_i + \beta_2 R_i + \beta_3 F_i + \sum_{j=1}^{3} \theta_j U_i^j + \gamma U_i^{CSM} \ast C_i + \omega U_i^{CSM} \ast H_i + \epsilon_i,
\]

where \( H_i \) and \( R_i \) are dummy variables indicating that student \( i \) participated in a hypothetical prisoner’s dilemma treatment or the real money group (with the control group excluded as a reference), \( F_i \) is equal to one if the student is a female, \( U_i^j \) are dummy variables for the four universities at which the experiments were conducted (with ETSU excluded as a reference university), and \( \beta \) and \( \theta \) are the parameters to be estimated. We include interaction terms for students at CSM with the hypothetical treatment and the real money group in order to control for further potential differences between their high mathematical focus compared to other universities in this study, and \( \gamma \) and \( \omega \) are the parameters to be estimated. We also estimate equation (1) with only the treatment variables, with only the treatment variables and dummy variables for the universities but not the interaction variables, and with all variables plus gender.

Table 2 presents the results from equation (1).\(^6\) Students in the control group are the excluded group so the coefficients on the hypothetical-game and real-game group variables represent the change in score from playing the hypothetical game or playing the game for real money. Students in the real money group earned test scores that were four to eight percentage points higher than students in the control group. This result is statistically significant in three of the four models \((p < 0.05)\) while marginally insignificant in one model \((p = 0.16)\). Students in the hypothetical group earned test scores that were statistically indistinguishable from those in the
control group. Post-estimation tests comparing scores between the hypothetical and real groups show statistically significant differences (i.e., the students in the real group earned higher scores.) We found statistically significant differences in scores across institutions, with students at Benedictine College, the Colorado School of Mines, and Kenyon College all earning higher scores than students at East Tennessee State University (the excluded group). We did not find any statistically significant effects of student gender on test scores.

[Insert Table 2 about here]

Originally, we had 100 additional observations from three separate professors at Susquehanna University each of whom conducted only two of the three treatments. We do not include these observations in our current analysis because we worry that unobservable professor or student-body characteristics may not be appropriately controlled for when only two of the three treatments are included for any given professor. This is especially true since all three professors were from the same university. When we include these 100 observations, we find that students in the real group scored four to eight percentage points higher than students in the control group and that this difference is statistically significant (p < 0.05) in two of the four models (p = 0.11 and 0.16 in other models). These results also indicate that students in the hypothetical group scored two to four percentage points lower than students in the control group, but those results are only statistically significant in one of the four models.

DISCUSSION AND CONCLUSION

We find significant differences in test scores between students who played a prisoner’s dilemma game for real money and students who either played a hypothetical version of the game or played no game at all. Students who played the game for real money earned test scores four to
eight percentage points higher than other students. Our findings challenge the conventional wisdom that monetary incentives are unnecessary in classroom experiments.

While we find that monetary incentives matter overall, this finding was not true for students at all universities. More specifically, students at the Colorado School of Mines earned test scores that were remarkably similar across all three treatments. This school is unique to our sample because of its focus on engineering. CSM students likely have a more substantial mathematics background, diminishing the effectiveness of the experiment in improving test scores. CSM students in the control group scored almost 90% on the assessment questions. For students who do this well, it is unlikely that a classroom game would have a large impact on learning. This does not necessarily mean the experiments had no value for these students. If students maintain the same exam scores but found economics more enjoyable because of the experiments, one could argue that the experiments were worthwhile.

Our research provides evidence that monetary incentives in experiments can boost student performance. Doing some simple back-of-envelope calculations gives an initial sense of the cost-effectiveness of the incentives on learning. We found an average payout of $2 per student increased scores by about seven percentage points (relative to the hypothetical or control groups) on an exam question worth between 1% and 4% of the overall course grade. This works out to be between a 0.07% and 0.28% increase in students’ overall course grade. Therefore, a ten percentage point increase in overall grades would require $72 to $286 per student. While cost-prohibitive for an instructor to implement, relative to the cost of higher education this is an inexpensive intervention to increase student performance.

While our results show monetary incentives can improve test scores, some caveats are in order. First, because most professors in our sample did not teach three sections of the principles
course in the same semester, they collected observations from students over the course of two or three semesters. A replication of this research with the control and treatment groups all occurring in the same semester could further ensure results weren’t driven by differences in student body composition across semesters.

In addition, future research could examine how large the payouts must be to achieve results. Our games resulted in an average payout of about $2 per student. Our results may have been driven by the fact that providing monetary incentives is so unusual that it signaled the importance of this topic, causing students to work hard. However, it is possible that much smaller monetary payouts or even non-monetary payouts like candy or donuts would yield similar increases in assessment scores. There are also other avenues for future research. We only conducted one experiment per class. It is not clear whether the positive impact of monetary incentives would diminish or grow stronger if an instructor conducted a series of games with monetary incentives. We found differences across institutions, but more research would also be valuable for better understanding why different institutions saw different gains in student learning.

The differences in results we found across universities highlights the importance of our research design. If data had only been collected from one university, the theme of this paper could have been dramatically different. Given different student body compositions and professor instruction styles, we recommend that researchers who study the effectiveness of various classroom pedagogies collect data from students at multiple universities and various professors.
The concept of individual response to incentives, both positive and negative, is a cornerstone of many widely adopted principles of economics textbooks. Mankiw (2014) highlights the power incentives as one of his core Ten Principles of Economics and builds upon this principle throughout the textbook.

Note that in the game, we called the choices “share” or “hoard” as we felt those terms were easier for students to understand.

While the specific method of grading the questions varied across professors, that should not affect the variation in student scores for any given professor.

We also ran non-parametric Wilcoxon rank-sum tests, which showed results that were virtually identical to the t-tests. Results are available from the authors upon request.

The share of the total possible points of the game theory assessment questions on the exam varied across professors, but was approximately 15% for all but the Kenyon students, who took a separate quiz solely on the game theory assessment questions. For Kenyon students, when we did comparisons of the score on the game theory questions to other exam questions, we used the scores students’ earned on the previous exam.

We also estimated the same four regressions shown in Table 2 using the students’ score on the non-game theory portion of the exam as the dependent variable. We found no statistically significant effect of being in the real money group versus other groups in three of the four models, but found a marginal effect (p < 0.10) in the remaining model. This provides modest evidence that while the game for real money stakes improved learning on the game theory
questions, it appears that there were not significant spillover effects to other topics within the course. Results are available from the author upon request.

7 The t-values for post-estimation tests to examine differences between the hypothetical and real groups for the three models shown were 2.31, 2.30, and 1.84, meaning that in models 1 and 2 the difference is statistically significant at the 5% level, while in model 3 the difference is statistically significant at the 10% level.
REFERENCES


### TABLE 1: Mean Scores on the Game Theory Assessment

<table>
<thead>
<tr>
<th></th>
<th>No Game</th>
<th>Hypothetical Game</th>
<th>Real Money Game</th>
</tr>
</thead>
<tbody>
<tr>
<td>Professor 1: ETSU</td>
<td>0.59</td>
<td>0.61</td>
<td>0.71</td>
</tr>
<tr>
<td>(N=148)</td>
<td>(N=68)</td>
<td>(N=59)</td>
<td>(N=21)</td>
</tr>
<tr>
<td>Professor 2: Kenyon</td>
<td>0.80(^b)</td>
<td>0.92</td>
<td>0.91</td>
</tr>
<tr>
<td>(N=142)</td>
<td>(N=91)</td>
<td>(N=27)</td>
<td>(N=24)</td>
</tr>
<tr>
<td>Professor 3: Ben</td>
<td>0.81(^c)</td>
<td>0.74</td>
<td>0.82</td>
</tr>
<tr>
<td>(N=157)</td>
<td>(N=59)</td>
<td>(N=74)</td>
<td>(N=24)</td>
</tr>
<tr>
<td>Professor 4: CSM</td>
<td>0.89(^d)</td>
<td>0.87</td>
<td>0.86</td>
</tr>
<tr>
<td>(N=194)</td>
<td>(N=65)</td>
<td>(N=70)</td>
<td>(N=59)</td>
</tr>
<tr>
<td>Overall</td>
<td>0.77</td>
<td>0.77</td>
<td>0.84</td>
</tr>
<tr>
<td>(N=641)</td>
<td>(N=283)</td>
<td>(N=230)</td>
<td>(N=128)</td>
</tr>
<tr>
<td>Overall score on rest of exam</td>
<td>0.77</td>
<td>0.76</td>
<td>0.78</td>
</tr>
<tr>
<td>(N=641)(^e)</td>
<td>(N=283)</td>
<td>(N=230)</td>
<td>(N=128)</td>
</tr>
</tbody>
</table>

\(^a\) For all professors, the observations for each treatment are all from one section unless noted.

\(^b\) Three sections were in the no-game treatment for professor 2.

\(^c\) Two sections were in the no-game treatment and in the hypothetical game treatment for professor 3.

\(^d\) For Professor 4, all students were in a large class, and each treatment was given to two different recitation sections.

\(^e\) Across all professors, we see no statistically significant differences in overall scores on the rest of the exam.
TABLE 2: Linear Regression Estimates of Student Assessment Scores

<table>
<thead>
<tr>
<th></th>
<th>Model 1 (N = 641)</th>
<th>Model 2 (N = 641)</th>
<th>Model 3 (N = 641)</th>
<th>Model 4 (N = 641)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.77***</td>
<td>0.65***</td>
<td>0.69***</td>
<td>0.60***</td>
</tr>
<tr>
<td></td>
<td>(0.02)</td>
<td>(0.03)</td>
<td>(0.04)</td>
<td>(0.03)</td>
</tr>
<tr>
<td>Hypothetical game</td>
<td>-0.01</td>
<td>0.00</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>(0.03)</td>
<td>-0.02</td>
<td>(0.03)</td>
<td>(0.04)</td>
</tr>
<tr>
<td>Real game</td>
<td>0.07**</td>
<td>0.04</td>
<td>0.08**</td>
<td>0.08**</td>
</tr>
<tr>
<td></td>
<td>(0.03)</td>
<td>(0.03)</td>
<td>(0.04)</td>
<td>(0.04)</td>
</tr>
<tr>
<td>Female</td>
<td></td>
<td></td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.02)</td>
<td></td>
</tr>
<tr>
<td>Kenyon College</td>
<td>0.22***</td>
<td>0.22***</td>
<td>0.22***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.03)</td>
<td>(0.03)</td>
<td>(0.03)</td>
<td></td>
</tr>
<tr>
<td>Benedictine College</td>
<td>0.16***</td>
<td>0.15***</td>
<td>0.15***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.03)</td>
<td>(0.03)</td>
<td>(0.03)</td>
<td></td>
</tr>
<tr>
<td>Colorado School of Mines</td>
<td>0.25***</td>
<td>0.29***</td>
<td>0.29***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.03)</td>
<td>(0.04)</td>
<td>(0.04)</td>
<td></td>
</tr>
<tr>
<td>Interaction of Colorado School of Mines and hypothetical game</td>
<td>-0.04</td>
<td>-0.04</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.05)</td>
<td>(0.05)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interaction of Colorado School of Mines and real game</td>
<td>-0.11*</td>
<td>-0.11*</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.06)</td>
<td>(0.06)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R-squared</td>
<td>0.01</td>
<td>0.12</td>
<td>0.13</td>
<td>0.13</td>
</tr>
<tr>
<td>F-value</td>
<td>3.03**</td>
<td>17.39***</td>
<td>12.93***</td>
<td>11.30***</td>
</tr>
</tbody>
</table>

Note: Dependent variable is the percentage score earned on the game theory assessment.

* p < 0.10; ** p < 0.05; *** p < 0.01.