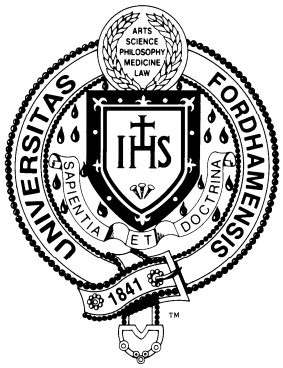


Fordham University
Department of Economics
Discussion Paper Series

Explaining Randomized Evaluation Techniques Using Classroom Games

Subha Mani
Fordham University, Department of Economics

Utteeyo Dasgupta
Franklin and Marshall College, Department of Economics

Discussion Paper No: 2010-06
September 2010

Department of Economics
Fordham University
441 E Fordham Rd, Dealy Hall
Bronx, NY 10458
(718) 817-4048

Explaining Randomized Evaluation Techniques Using Classroom Games

Subha Mani* and Utteeyo Dasgupta**

Abstract: Over the last decade, randomized evaluations have taken the field of development economics by storm. Despite the availability of strong review pieces in the topic, there is no pedagogical paper on randomized evaluation. This paper bridges the gap by introducing three interactive classroom games to communicate the concepts of Average Treatment Effect (ATE), Intent–to-Treat Effect (ITT), Sub-group Average Treatment Effect (SATE), and Externality Effect (EE). The classroom games are easy to implement and provide students an opportunity to participate in a simple randomized trial of their own.

Keywords: program evaluation, classroom experiment, pedagogy, economic development

JEL codes: A22, C70

Acknowledgements:

We thank Sarah Scarcelli for excellent research assistance on this project. We are grateful to students at Fordham University for their enthusiastic participation in the games.

Contact: *Subha Mani**,** Department of Economics, Fordham University, 441 East Fordham Road, Bronx, New York 10458. Email – smani@fordham.edu.

**Utteeyo Dasgupta, Department of Economics, Franklin and Marshall College, PO Box 3003, Lancaster, PA 17604. Email – utteeyo.dasgupta@fandm.edu.

1. **Introduction**

Classroom-experiments are an important pedagogic tool in economics teaching. Its popularity is evident in frequent citing under education journals as well as general economics journals.[1] There are online journals (classroom Experinomics), as well as economic pedagogy repositories (Econport.org) that exclusively promote and facilitate classroom experiments. Many popular textbooks at the introductory, and advanced level include experiments for classroom use (See Delemeester and Neral 1995, Ortman and Colander 1995, Bergstrom and Miller 1997, Stodder 1998, Holt 2006, Schotter 2008). Research suggests that classroom teaching-experiments not only enhance student involvement, but also improve learning outcomes and test scores (See Frank 1997, Gremmen and Potters 1997, Emerson and Taylor 2004, Dickie 2006).

However, in spite of the positive association between classroom games and student learning, classroom experiments mainly remain restricted to a few topics in introductory economics, microeconomics, industrial organization, and game theory. While there is scope for building imaginative and entertaining classroom games in other sub-fields of economics, surprisingly little works exists in these other areas. The popular tool of randomized evaluations used by development economists is one such area (see Duflo et. al 2008 for recent review). The increase in research publications, blogs and newspaper coverage, and overwhelming student and instructor interest in the area indicates its prominence in the discipline.[2] Even though the very

[1] Southern Economic Journal in the recent past came up with a symposium on economics experiments, and the Journal of Economic Perspectives has a regular feature on classroom experiments. Economic Inquiry as well as Experimental Economics often publishes such pieces.

[2] 1. http://www.povertyactionlab.org 2. http://www.3ieimpact.org 3.http://poverty-action.org

idea of randomized evaluation is amenable to classroom experiments, there is no pedagogical paper on the topic so far. To the knowledge of the authors this is the first paper to use classroom games to communicate some of the core concepts in the area. Three classroom games are used to discuss the concepts of Average Treatment Effect (ATE), Intent-to-Treat Effect (ITT), Externality Effect (EE), and Sub-group Average Treatment Effect (SATE). The classroom games described here are easy to implement and provide students an opportunity to participate in a simple randomized trial of their own. The classroom games can be used in both undergraduate as well graduates courses in development economics, applied microeconomics, and program evaluation.

The rest of the paper is organized as follows. Section 2 provides a brief introduction to concepts in program evaluation. Section 3 outlines the relevant classroom games. Section 4 provides a discussion, followed by further readings in section 5.

2. Definitions

The focus of this section is to introduce some of the key concepts used in the evaluation literature. Consider a pool of applicants (N) for a job training program. A randomly selected subset N_T gets assigned to the treatment group (T), and receives the treatment (the job training program). The remaining sample $N_C = N - N_T$ gets assigned to the control (C) group which does not receive the training. In our example we are interested in measuring the effect of the training program on some measurable outcome variable (Y) such as wage earnings.

---

4. http://cega.berkeley.edu 5. http://economix.blogs.nytimes.com 6.http://greeneconomics.blogspot.com
7. emlab.berkeley.edu/users/webfac/emiguel/e270c_s09/syllabus.pdf
8. http://www.isites.harvard.edu/icb/icb.dokeyword=k62906
9. http://catalog.middlebury.edu/offerings/view/catalog/catalog/MCUG/offering/section/200810/11093
10. http://karlan.yale.edu/courses.html

Average Treatment Effect (ATE)

The ATE measures the overall impact of a program on an observable outcome variable. Under perfect compliance,[3] it is defined to be the difference in the empirical means of the outcome variable (Y collected at the end of the training program) between the treatment and the control group. Thus, under perfect compliance,

$ATE = \overline{Y_T} - \overline{Y_c}$, where $\overline{Y_T}$ is the sample mean of the outcome variable for everyone in the treatment group and $\overline{Y_c}$ is the sample mean of the outcome variable for everyone in the control group.

In many social experiments, imperfect compliance is a source of concern as it affects the measurement of the impact of the program. It can come about in two ways - one, where some of the individuals originally assigned to receive the treatment do not receive the treatment. Two, when some of the individuals originally chosen not to receive the treatment (i.e., assigned to the control group) end up receiving the treatment. Consequently, under imperfect compliance, we are interested in measuring two related effects – (a) the impact of being treated and (b) the impact of offering the treatment.

Impact of the Treatment on the Treated (TOT)

Under imperfect compliance, the TOT captures the average gain of the program for those who actually get treated. TOT is defined as the difference in the sample means of the outcome variable between the treatment and the control group, divided by the difference in the probability

[3]Under perfect compliance, everyone in the treatment group gets treated and no one from the control group receives the treatment.

of being treated in the treatment group and the probability of being treated in the control group. So,

$$TOT = \frac{\overline{Y_T} - \overline{Y_c}}{P[treated \mid T] - P[treated \mid C]}$$

If no one in the control group gets treated,

$$TOT = \frac{\overline{Y_T} - \overline{Y_c}}{P[treated \mid T]}$$

Note, under perfect compliance, $P[treated \mid T] = 1$ and $P[treated \mid C] = 0$, thus the TOT collapses to the ATE.

Intent-to-treat effect (ITT)

Program participation is often voluntary in social programs, and as a result, randomization only affects the probability of being exposed to the treatment. Hence, under voluntary participation the researcher is interested in measuring the effect of being offered the program, rather than the actual treatment. ITT measures the average impact of offering a program using the initial random assignment as a way to avoid the re-introduction of selection bias. Under partial-compliance, the difference in the sample means of the outcome variable between the treatment group (those originally assigned to receive the treatment) and the control group (originally assigned not to receive the treatment) measures the ITT effects of the program. Notice, that under perfect compliance, the ITT and ATE will be identical. So,

$ITT = \overline{Y_T} - \overline{Y_c}$, where $\overline{Y_T}$ is sample mean of the outcome variable for those initially assigned to the treatment group (T) and $\overline{Y_c}$ is the sample mean of the outcome variable for those assigned to the control group (C); regardless of the treatment they actually receive. Under partial compliance as long as the treatment has non-negative effects, the ITT effects will normally be smaller than the ATE of the program.

Sub-group Average Treatment Effect (SATE)

SATE measures the impact of the treatment for exogenous sub-groups where the formation of the sub-group (X) is not affected by the treatment. SATE is defined to be the difference in the sample means of the outcome variable between the treatment and control group with a certain identical characteristic X (ex: Bobonis et. al (2006) measure program impacts in the following sub-categories – gender, age, mother's schooling and baseline anemia). So,

$$SATE = \overline{Y_{TX}} - \overline{Y_{CX}}$$

For example: when X= male, the SATE will be the difference in the sample means of the outcome variable between all males in the treatment group and all males in the control group.

Externality Effect (EE)

EE measures the impact of the treatment on individuals and groups who are not targeted to receive the treatment. Let us assume that we have information on the friends of the people who applied for this job training program and for simplicity, let's assume that the friends did not apply for this training program. Let us call the friends of our treatment group, FT and let us call

the friends of our control group, FC. The externality effect of the program/treatment is measured as the difference in the sample means of the outcome variable between the FT and FC group.

$EE = \overline{Y_{FT}} - \overline{Y_{FC}}$, where $\overline{Y_{FT}}$ is the sample mean of the outcome variable for the FT group and $\overline{Y_{FC}}$ is the sample mean of the outcome variable for the FC group.

3. Classroom Games

There are three games discussed here; the Average Treatment Game (ATG), to explain the concept of average treatment effect, the Intent-To-Treat Game (ITG) to explain the concept of intent-to-treat effects, and the Externality Game (EG) to explain the concept of externality effect. The ATG is later used to explain the concept of sub-group average treatment effect as well.

Preparation:

The instructor needs to prepare the following before running the experiments. First, a bag full of red and white poker chips in equal proportions, enough to distribute amongst students. Second, three separate sets of wordlists which can be constructed using standard GRE vocabulary lists. The wordlists should contain a list of words with their associated meanings [See appendix tables]. One needs to make enough copies of each wordlist to hand out to about half the students in class. Third, prepare three different quizzes, each containing some words from the earlier constructed wordlists. Each quiz presents words with three possible choices next to each of them for students to circle/mark the closest synonym for each word in the exercises to follow [see appendix tables]. We used lists containing 15 words, and tests containing 10 words.

The treatment in our classroom games is the exposure to the wordlists. Our outcome variable is the quiz scores for each student.

Description of the Classroom Activity

At the beginning of class, each student is asked to pick up a chip from the bag. Once each student has got a chip the instructor asks the students to re-seat themselves in one of the sides of the room based on the color of the chip they picked. Students who picked up red chips are asked to seat themselves on the right side, and students who picked up white chips are asked to sit on the left side. Note that there is nothing special about which color-group goes to which side as long as the two color-groups are seated separately. The students carrying red chips play the role of the Treatment (T) group and students carrying the white chips play the role of the Control(C) group.

Average-Treatment Game

Each member in the T-group receives a copy of wordlist 1 [see Appendix Table A1]. They are given 5 minutes to review the wordlist. Students in the C-group have no task at the time. After five minutes are over the instructor collects all distributed copies of wordlist 1 from members of the T-group.

Next the quiz for ATG (See Appendix, Table A2) is distributed to all students in class. The students are allowed ten minutes to complete the vocabulary test. At the completion of the test, the instructor reads out the correct answers and asks the students to score their tests – a point for each correct answer. The students are asked to write the total on the left hand corner of the test. The instructor then collects all quiz sheets and computes the average score for the T-group,

the average score for the C-group, and the difference in the averages for the T and C groups gives us the Average Treatment Effect of the program.[4]

Intent-to-Treat Game

The instructor goes through the same exercise again, i.e., distributes a new worldlist (wordlist 2; see Appendix, Table A3) except that now this new wordlist is not only given to all students in the T-group but also to a sub-set of students from the C-group. The idea is that a part of the C group is now exposed to the treatment as well. To implement this easily in class, the instructor can distribute wordlist 2 to all students sitting in the first two rows of the C group as well.[5] After 5 minutes, the instructor collects these wordlists and administers the quiz for the Intent to Treat Game [see Appendix table A4] to all students in class.

At the completion of the test, the instructor reads out the correct answers as before, and asks the students to score their tests – a point for each correct answer. The students are asked to write the total on the left hand corner of the test. The instructor then collects all quiz sheets and computes the average score for the T-group, the average score for the C-group, and the difference in the averages for the T and C groups gives us the Intent-to-treat effect of the program.

Externality Game

The instructor verbally assigns numbers to sitting positions in the T-group in a sequential manner. For explanatory purposes define T-odd group to be the students who are in the T group

[4] It might be useful to have an excel-sheet with the appropriate formula written on it already. This can simplify things for the instructor in class.

[5] Usually it would be enough to distribute this word list to roughly about 6 students from the C-group where the control group has about 20 students. The instructor can improvise here depending up the size of their C group.

and are in odd seating positions (i.e., positions 1,3, 5, 7…13,15); define T-even to be the rest of the students (i.e., students sitting in even sitting positions). The instructor intends to provide only T-odd group members with the wordlist (Wordlist 3). A similar ordering is enforced in the C group, and we define every student receiving an even number to be part of the C-even group, and every student receiving an odd number to be part of the C-odd group. Neither of the C-even and the C-odd groups receives the treatment.

The idea behind this procedure is the following: Although, it is not intended for members of the T-even group to receive the treatment, the fact that their adjacent neighbors (members of the T-odd group) have the worldlist could expose the former group members to the list (our treatment). The exact process by which such an event happens can vary. Since the instructor does not give any explicit rules regarding sharing of the lists, it is possible that the T-odd members end up discussing the lists with their neighbors. Alternatively, a simple peek from T-even positions can do the trick! Either way, this provides a situation where the impact of the program (i.e., exposure to the wordlist) has the potential to go beyond those who were intended to receive the treatment.

The rest of the procedure is as before, i.e., the instructor takes back the wordlists from T-odd group members, and then administers a vocabulary test (see Appendix table A6) to all students in the classroom. At the completion of the test, the instructor reads out the correct answers as before and asks the students to score their tests – a point for each correct answer. The students are asked to write the total on the left hand corner of the test. The instructor then collects all quiz sheets and computes the average score for the T-even group, the average score for the C-even group, and the difference in the averages for the T-even and C-even groups. This

difference in the averages of the T-even and C-even group captures the externality effect of the treatment.

4. Discussion

At the completion of the experiments, the instructor can discuss the results of the three games. It is useful to have excel graphs ready here. The results discussed below were obtained from running these games at Fordham University.

The difference in the average quiz scores of the T and C group in the Average Treatment Game captures the average treatment effect**.** In the experiments we ran in class, we found that the sample mean of the T group was 9.7 while the sample average of the C group is only 5.8. The average treatment effect turns out to be 3.8 (with a standard error = 0.41, and statistically significant at 1% significance level).

The difference in the average quiz scores of the T and C group in the Intent-To-Treat game captures the intent-to-treat effect**.** In the experiments we ran in class, we found that the sample mean of the T group was 9.63 while the sample average of the C had increased to 7.02. The intent-to-treat effect turns out to be 2.6 (with a standard error = 0.45, and statistically significantly at 1% significance level).

Figures 1 and 2 can be computed using excel and should be used to depict the decline in the average difference in the quiz scores (3.8 to 2.6) from the ATG to the ITG. This should be used to discuss how positive spill-over's/contamination/partial compliance can reduce the impact of the program (ITT<ATE).

The findings from these games can further be used to disseminate the concept of SATE. The test scores from average treatment game can be used to compute the average treatment effects separately for boys and girls. For instance, let's assume that the first twelve students in

the T group were male and the remaining students were female. Similarly assume that the first twelve students in the C group are male and the remaining students are female. We can now compute the difference in the quiz scores for the male treatment and male control and similarly for females. In our example, we find that the SATE for males is 3.25 (standard error = 0.62) and SATE for females is 4.7 (standard error = 0.42).

Finally, we are interested in capturing the externality effect of being exposed to the T group. To capture this, we now compute the difference in the average test scores for only T-even and C-even. In our example we find that the average score for the T-even group is 7.66 and the average score among C-even group is 7.09, a difference of 0.57 in the right direction although not statistically significant. Note, that one can possibly strengthen the externality effects if the instructor were to announce that students in group T-odd were free to decide whether to share the information (wordlist) or not with students in group T-even.

5. Further readings

At the end of the discussion comprising the experiment results and the appropriate concepts, one can follow up with some of the reading below depending on class interest. Duflo et. al (2008) provides a comprehensive review of randomized evaluation techniques. To discuss specific interventions in – (a) Health, see Miguel and Kremer (2004); Bobonis et. al (2006); Cohen and Dupas (2009); Thomas et. al (2003), (b) Education, refer to Banerjee et. al (2007); Glewwe and Kremer (2008); Duflo and Hanna (2006); Kremer (2003); Kremer et. al (2004); Parker et. al (2008), (c) Agriculture, refer to Duflo et. al, (2006), (d) Microfinance, see Field et. al 2010; Pande and Field 2008; Banerjee et. al 2009.

References:

Bobonis, G., E. Miguel, C. P-Sharma. 2006. Iron Deficiency Anemia and School Participation, Journal of Human Resources, vol 41 (4), 692-721.

Banerjee, A, S. Cole, E. Duflo and L. Linden. 2007. Remedying Education: Evidence from Two Randomized Experiments in India, The Quarterly Journal of Economics, MIT Press, vol. 122(3), 1235-1264

Banerjee, A., E. Duflo, R. Glennerster, C. Kinnan. 2009. The Miracle of Microfinance? Evidence from a Randomized Evaluation [available at: http://econ-www.mit.edu/files/4161]

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

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

Cohen, J. and P. Dupas. 2009. Free Distribution or Cost-Sharing? Evidence from a Randomized Malaria Prevention Experiment, Quarterly Journal of Economics, vol. 125(1), 1-45.

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

Duflo, E., R. Glennerster and M. Kremer. 2008. Using Randomization in Development Economics Research: A Toolkit, in T.P. Schultz and J. Strauss (eds.), Handbook of Development Economics, Volume 4, Amsterdam: North Holland Press (selected sections only).

Duflo, E., Kremer,M., Robinson, J. 2006. Understanding technology adoption: Fertilizer in western Kenya, preliminary results from field experiments. Mimeo.

Duflo, E., Hanna, R. 2006. Monitoring works: Getting teachers to come to school. Working paper No.11880. NBER.

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

Field, E., S. Jayachandran, and R. Pande. 2010. Do Traditional Institutions Constrain Female Entrepreneurship? A Field Experiment on Business Training in India. American Economic Review, Papers and Proceedings..

Frank, Bjorn. 1997. The Impact of Classroom Experiments on the Learning of Economics: An

Empirical Investigation, Economic Inquiry 35(4): 763-769.

Glewwe, P., Kremer, M. 2008. Teacher incentives. Working paper 9671. National Bureau of Economic Research.

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

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

Kremer, M. 2003. Randomized evaluations of educational programs in developing countries: Some lessons. American Economic Review 93 (2), 102–106.

Kremer, M., Miguel, E., Thornton, R. 2004. Incentives to learn. Mimeo, Harvard University.

Miguel E. and M. Kremer. 2004. Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities, Econometrica, 72.1:159-217

Pande, R. and E. Field. 2008. Repayment Frequency and Default in Micro-Finance: Evidence from India, Journal of European Economic Association Papers and Proceedings.

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

Schotter, Andrew. 2008. Microeconomics: A Modern Approach, South Western Publishing.

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

Thomas, D., Frankenberg, E., Friedman, J., Habicht, J.-P., Al E. 2003. Iron deficiency and the well being of older adults: Early results from a randomized nutrition intervention. Mimeo, UCLA.

Figure 1: Average Treatment Game

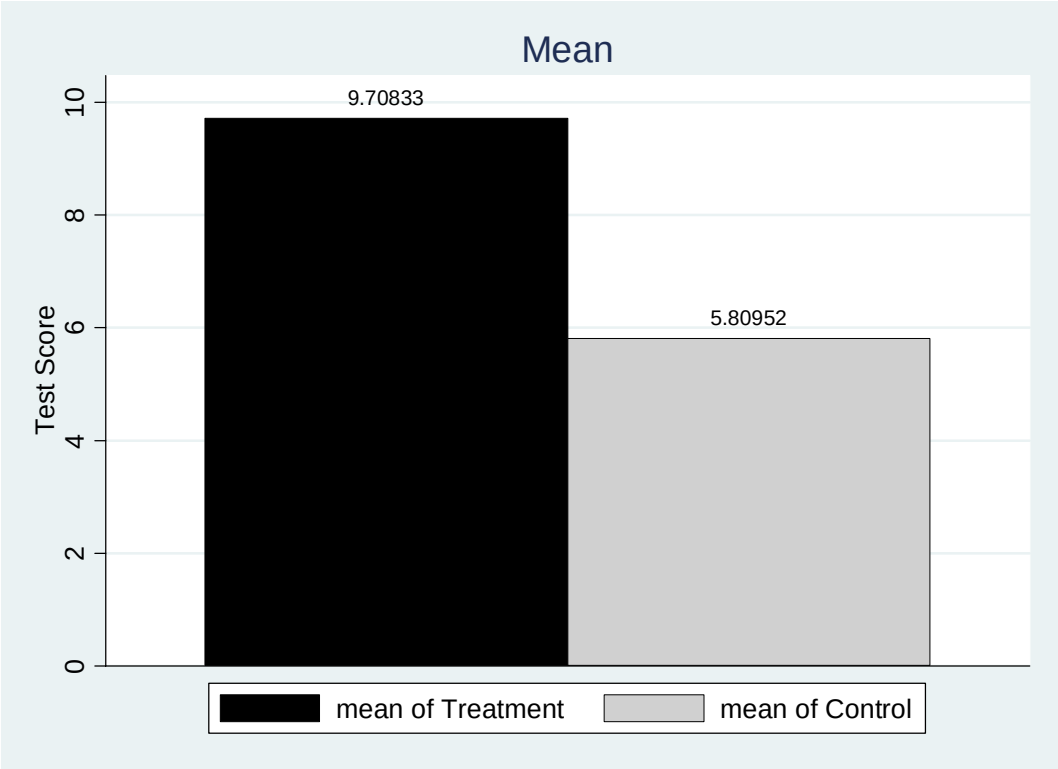


Figure2: Intent-to-treat Game

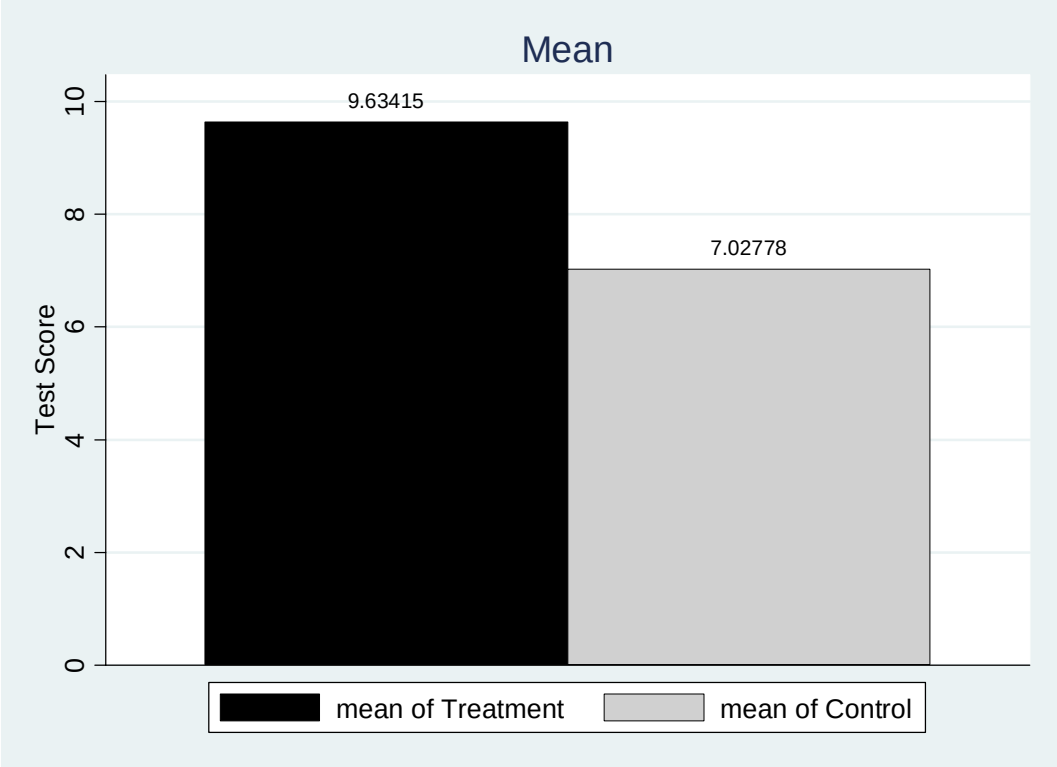
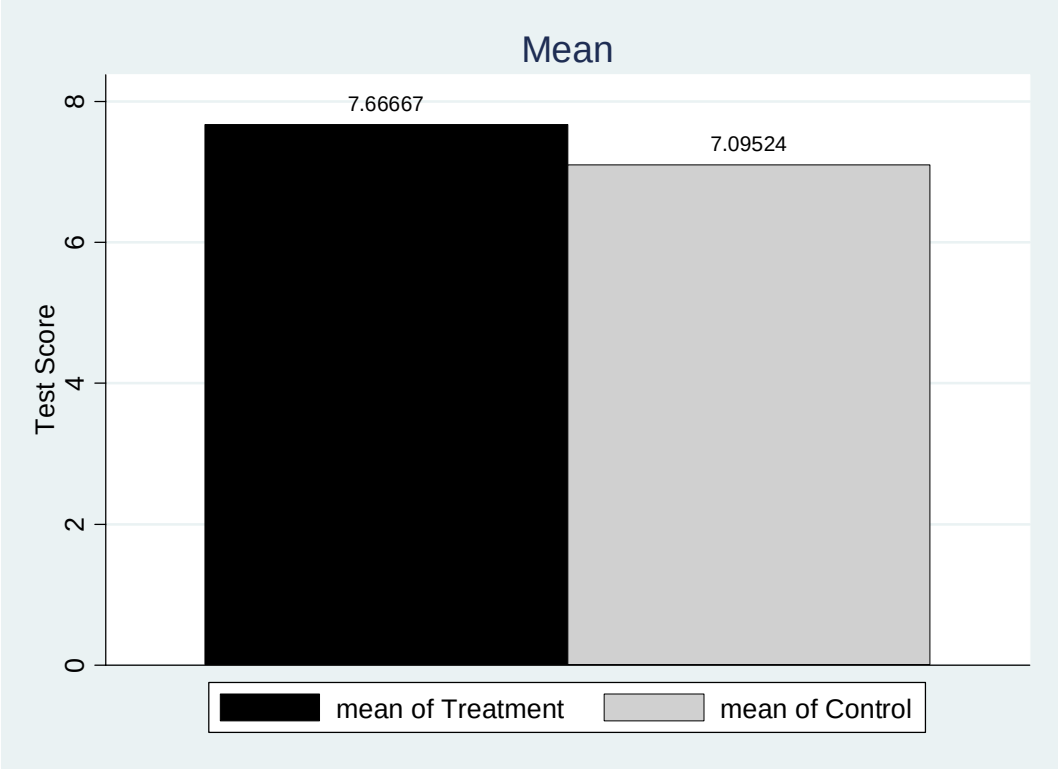


Figure 3: Externality Game



Appendix

Table A1: Wordlist 1

1) Antediluvian - Ancient
2) Anomalous - unique
3) Ambrosial - Delicious
4) Emollient - Softening
5) Inchoate – incomplete
6) Dearth - scarcity
7) Nefarious – Evil
8) Efficacy - effectiveness
9) Breach - gap
10) Ossified - Inflexible
11) Perfidious - Dishonest
12) Vex - Irritate
13) Quiescence – Inactivity
14) Erudite – scholarly
15) Sedulous - Diligent

Table A2: Quiz for ATG

Please circle the closest synonym/meaning for the following words:

1) Antediluvian – (a) Ancient (b) Aggravate (c) New
2) Ambrosial – (a) Ugly (b) Delicious (c) Similar
3) Emollient – (a) Softening (b) Loud (c) Unfortunate
4) Inchoate – (a) Incomplete (b) Complete (c) Difficult
5) Nefarious - (a) Delightful (b) Fun (c) Evil
6) Ossified – (a) Inflexible (b) Sanguine (c) Tractable
7) Perfidious – (a) Dishonest (b) Occlude (c) Honest
8) Vex – (a) Short (b) Poor (c) Irritate
9) Quiescence – (a) Inactivity (b) Silent (c) Sate
10) Sedulous – (a) Diligent (b) Careless (c) Impatient

Table A3: Wordlist-2

1) Abjure - promise
2) Admonitory – containing warning
3) Baneful – causing harm
4) Cadge – to beg
5) Commodious– plenty of space
6) Contrite-filled with deep sorrow
7) Eschew – avoid
8) Fecund - fertile
9) Garrulous - talkative
10) Halcyon – calm and peaceful
11) Esoteric – difficult to understand
12) Nadir – lowest point
13) Petulant – unreasonably impatient
14) Recant – take back
15) Sanguine - cheerful

Table A4: Quiz for ITG

Please circle the closest synonym/meaning for the following words:

(1) Abjure – (a) abhor (b) allude (c) promise
(2) Baneful – (a) promise (b) supportive (c) causing harm
(3) Cadge – (a) beg (b) candor (c) talkative
(4) Contrite- (a) short (b) filled with deep sorrow (c) sanguine
(5) Eschew – (a) avoid (b) difficult (c) painful
(6) Fecund – (a) barren (b) recant (c) fertile
(7) Halcyon – (a) Irritable (b) calm and peaceful (c) garrulous
(8) Nadir – (a) highest point (b) contrite (c) lowest point
(9) Petulant – (a) onerous (b) valiant (c) unreasonably impatient
(10) Sanguine – (a) cheerful (b) taciturn (c) mettlesome

Table A5: Wordlist 3

1) Aplomb – Self-confidence
2) disparate– different
3) Egress – exit
4) Immaculate – faultless
5) Indigenous – Native
6) Contrite-filled with deep sorrow
7) Maladroit – Tactless
8) Mettlesome – Courageous
9) Onerous - burdensome
10) Parley – negotiation
11) Irksome - tiresome
12) Pariah – outcast
13) Taciturn – silent
14) Unscathed– unharmed
15) Guile – cunning

Table A6: Quiz for Externality Game

Please circle the closest synonym/meaning for the following words:

(1) Aplomb – (a) self-confidence (b) brave (c) truth
(2) Immaculate –(a) trace (b)faultless (c) imperfect
(3) Indigenous – (a) native (b) volatile (c) taut
(4) Maladroit – (a) clever (b) versatile (c) tactless
(5) Onerous – (a) malign (b) burdensome (c) mendacity
(6) Parley – (a) miser (b) nexus (c) negotiation
(7) Irksome – (a) tiresome (b) petrify (c) sudden
(8) Pariah – (a) rebuff (b) rivet (c) outcast
(9) Taciturn – (a) talkative (b) silent (c) funny
(10) Unscathed– (a) unharmed (b) untoward (c) earth