Pre-analysis Plan for "Stimulating Curiosity to Enhance Learning: Results from a Randomized Intervention"

Sule Alan[*], Mert Gumren[†] and Ipek Mumcu[‡]

March 2019

1 Introduction

The study evaluates a large-scale randomized educational intervention that aims to stimulate curiosity in the classroom environment. It is motivated by the evidence on the importance of curiosity as a fundamental driver of academic achievement. By primarily emphasizing curiosity in science and technology, the intervention aims to increase curiosity and performance in STEM subjects.

The intervention involves an intensive teacher training and has two components:

1. Curricular component: An interdisciplinary team of teachers, education consultants and pedagogy specialists designed a curriculum to be delivered by students' own trained teachers. The curriculum contains visual and reading materials that emphasize the importance of having an inquisitive and open mind, and the benefits of asking questions.

2. Pedagogical Component: Teachers received various pedagogical (and practical) tips to practice inquiry-based learning and to create an inquisitive classroom culture.

Teachers are given 5 months to implement the proposed pedagogy supported by the activities prescribed in the curriculum.

[*]University of Essex, Department of Economics and J-PAL.
[†]Department of Economics, Koc University.
[‡]Institute for Social and Economic Research, University of Essex.

The trial involves 77 schools, 5892 officially registered 4th and 5th grade students in the province of Mersin, Turkey. The 4th grade students (the last year of primary school) are on average 9, the 5th grade students (the first year of post-primary school) are on average 10 years old. The unit of randomization is school where we assigned 40 schools to treatment and 37 to control after collecting rich baseline information. The randomization was done using stata, stratifying on grade level (post-primary vs primary) and district.

We plan to evaluate the intervention with respect to three groups of primary outcomes:

1. Experimental outcomes
2. Achievement outcomes
3. Educational Aspirations.

We will collect secondary outcomes to be used in mediation analyses to provide suggestive evidence on potential mechanisms.

Below, we explain our main empirical models and detail the construction of our outcomes.

2 Empirical Models

2.1 Benchmark Model

To test the null hypothesis that the program had no impact on the outcome y, we estimate the average treatment effect conditioning on baseline covariates that are predictive of the outcome of interest:

$$y_{ij} = \alpha_0 + \alpha_1 T_j + X_{ij}^{'}\gamma + \varepsilon_{ij}$$

where T_j is a dummy variable which equals 1 if school j is in the treatment group and zero otherwise, and X_{ij} is a vector of observables for student i in school j that are potentially predictive of the outcome y. The estimated $\hat{\alpha}_1$ is the average treatment effect on the treated. When estimating the treatment effects, in addition to variables used in stratification, we will use a number of baseline covariates. These are gender, student IQ (measured by Raven's progressive matrices), test scores, experimentally elicited risk and ambiguity aversion. We will also present our results without these covariates. Because most our measures are taken by physically visiting classrooms, and at any given day some pupils are absent from school, we will inevitably have some missing data. In the baseline, this absenteeism is balanced across treatment status. In estimating treatment effects using baseline covariates, we may utilize various imputation techniques to replace missing baseline covariates. Because the unit of randomization is school, we will cluster standard errors at school level. We will adjust our p-values for multiple testing.

2.2 Estimating LATE

Even though the program participation is voluntary on the part of the teacher, we expect heterogeneity in implementation intensity. To gauge this heterogeneity, we will ask treatment teachers to report their implementation intensity on a percentage scale. The model we will estimate then is :

$$y_{ij} = \beta_0 + \beta_1 TIntensity_{cj} + X_{ij}^{'}\theta + \epsilon_{ij}$$

where $TIntensity_{cj}$ is the intensity of the implementation reported by the teacher of class c in school j. This variable takes the value zero for control teachers by design. $TIntensity_{cj}$ variable will be instrumented with the binary treatment assignment. The estimated $\hat{\beta}_1$ will then give us LATE.

3 Primary Outcomes

In this section, we describe our primary outcomes in detail.

3.1 Experimental Outcomes: A Novel Behavioral Measure of Curiosity

We design a temporal behavioral task that elicits the key features of curiosity. Decisions and performances in this task constitute our "experimental outcomes". We did not implement this task in the baseline as we did not want students to be familiar with the task in the endline.

The implementation of the task consists of the following steps:

- Students will be introduced various booklets. They will be told that each booklet contains some rare information related to the topic written on it. Topics are animals, space, history, human body, chemistry/physics, technology (cars, aircrafts, helicopters), cartoon characters and sports. We prepared 10 pieces of rare information (unlikely to be known by most students) for each topic. For example, "The sunset on Mars is blue" is a piece of information we provide in the "Space" booklet.

- Students will then be asked to rank (in discretion) these booklets from most to least interesting based on their own preferences.
- They will then be offered 10 attractive stationary items worth of stamps (they will be given time to examine the stationary items we offer).
- They will be asked to allocate their 10 stamps between stationary items and one of these booklets.
- Students will be informed that there will be two payment regimes that are equally likely to occur. In the first regime (50% chance), we will randomly draw a price from an empirical distribution (obtained from pilot runs) and impose that to the classroom. This means that under this regime, those whose willingness to pay (WTP) for a booklet is lower than the randomly chosen price will not be able to receive their desired booklet. The rest will. In the second regime (50% chance), we will give random half of the class their desired booklet at their willingness to pay and the other half will not receive any booklet regardless of their willingness to pay (they will get 10 stationary items).
- We will then collect students' decisions (WTPs).
- We will visit the same classroom exactly 1 week later. In the second visit, we will conduct a test where we will have 5 questions from each booklet.

We will obtain three outcomes by implementing this task:

1. Ranking of science related topics (interest in science)
2. The number of stamps allocated to the desired booklet, WTP (desire for knowledge).
3. The number of correct answers given in the test in the second visit (knowledge retention).

We will test the null hypothesis there is no difference between the treatment and control groups with respect to these three outcomes.

We are also interested in how treatment affects the ways in which knowledge is transmitted within a classroom. In particular, we are interested in the effect of the treatment on the knowledge of students who did not receive any booklet. For this, we will estimate the effect of the treatment by conditioning on "not receiving a booklet". Because this status will emerge endogenously under Regime 1, we will restrict our sample to the classrooms where we imposed Regime 2. Recall that in this regime, "no booklet" status is randomly assigned by the experimenter.

We will test the hypothesis that the treatment will increase knowledge among students who did not receive a booklet.

We hypothesize that the treatment will increase knowledge on topics that are not chosen.

3.2 Achievement outcomes

We will also conduct science, math and verbal (Turkish) tests based on the national curriculum for the relevant grade level. The standardized performance of these tests will constitute our achievement outcomes. We will also collect official end-of-year grades from the school administration.

Because of its heavy science emphasis, *we will test the null hypothesis there is no difference between the treatment and control groups with respect to achievement in science topics (objective test scores and teacher given grades).* We do not apriori expect any effect on math and verbal test performances and end-of-year grades.

3.3 Educational Aspirations for STEM

We will collect self-reports on whether the student aspires to go to university and, what he/she wants to choose as a study major.

We will test the null hypothesis there is no difference between the treatment and control groups with respect to aspirations for studying STEM majors.

4 Heterogeneous Treatment Effects

We are interested in heterogeneous effects in 3 dimensions: Gender, IQ and baseline classroom network structure. For gender, we will test whether the treatment has a differential impact across gender. For IQ, we will test whether the treatment has differential impact on students with different IQ levels.

We hypothesize that classrooms with well-connected networks transmit knowledge better. To test this hypothesis, we will estimate treatment effect on knowledge in unchosen booklet topics interacted with network connectedness.We will also test this hypothesis by restricting the sample to students who did not receive a booklet in Regime 2, and estimate the effect of the treatment on booklet knowledge interacted with network connectedness. Recall that not receiving a booklet in Regime 2 is an experimentally (exogenously) imposed condition.

5 Secondary Outcomes for Exploring Mechanisms

We will collect a number of secondary outcomes in the endline. These outcomes will help us understand potential channels through which treatment affects our primary outcomes. These secondary outcomes are

1. Survey reports of perseverance. For this, we will use a factor extracted from a number of item set questions adapted from Duckworth 8-item Grit Scale (Duckworth and Quinn 2009).

2. Self-confidence. For this, we will utilize our math, verbal and science tests. Specifically, we will ask students the number of questions they think they answered correctly for each test after they have completed the tests. The difference between the student's response and his/her actual performance constitute our self-confidence measure.

3. Risk aversion and ambiguity aversion. These will be measured via incentivized tasks (in the exact same way as they were elicited in the baseline). Risk Elicitation Task: Risk preferences will be elicited via a mechanism that asks the student to allocate a number of "tokens" between a risky and safe option. Each student will be given 5 tokens. All tokens placed in the safe option will be the student's to keep, whereas all tokens that

are put into the risky option will have a 50% chance of being tripled, and 50% chance of being lost. The amount placed in the safe option is a measure of the individual's risk aversion. It is a commonly used task for child participants (see Gneezy and Potters 1997). Ambiguity Aversion Elicitation Task: This will be elicited in exactly the same way except, the students will not be be informed about the odds of winning and losing.

4. Survey reports of curiosity. For this, we will use a factor extracted from a number of item set questions adapted from Jordan and Jimerson (2004).

References

Gneezy, U., & Potters, J. (1997). An experiment on risk taking and evaluation periods. The Quarterly Journal of Economics, 112(2), 631-645.

Jordan A. Litman & Tiffany L. Jimerson (2004) The Measurement of Curiosity As a Feeling of Deprivation, Journal of Personality Assessment, 82:2, 147-157

Duckworth A. and Quinn, P. D. (2009). Development and Validation of the Short Grit Scale (GritS). Journal of Personality Assessment, 91 (2), 166–174.