# Text Messaging for Parental Engagement in Student Learning Pre-Analysis Plan

Sarah Kopper and Anja Sautmann

November 2, 2020

#### 1 Overview

This document outlines the experimental design and pre-analysis plan for using exploration sampling to test the effects of weekly text messaging to parents' phones as a means to increase student use of a Whatsapp-based quiz platform and ultimately learning outcomes.

Our partner runs private schools in Lagos, Nigeria for students in early primary education and primaries 1-6 and is carrying out the implementation of the experiment.

Our study will proceed in two stages:

- 1. Use adaptive treatment assignment with exploration sampling (Kasy and Sautmann (2020)) and stratification to choose one out of four "types" of text messages. The text message type chosen is the one that promises to lead to the greatest increase in quiz use by students.
- 2. Use a standard RCT with a treatment group that receives the message type chosen above, and a control group that receives no messaging, to test whether quiz use as encouraged by the messages leads to sustained increases in exam scores.

#### Stage 1: adaptive treatment assignment to choose a message type

The goal is to choose one treatment (message type) for implementation in the RCT in Stage 2. In the five weeks prior to the winter break 2020 (Nov 2-Dec 7 weeks), we will conduct 5 rounds of adaptive testing.

**Treatments.** Each treatment consists of a set of text messages to be sent to parents weekly in order to encourage the use of the partner's proprietary Whatsapp-based quiz platform. Message templates for each treatment are in the attached table 1. The four treatments consist of four "types" of messages, sent each week to the randomly selected treatment groups.

- 1. Motivation framing a set of 10 messages focusing on different motivations for learning.
- 2. Testimonial framing a set of 10 messages focusing on the value of the quiz app.
- 3. One specific quiz proposing 1 quiz tailored to what the student covered in class that week/term.
- 4. Many quizzes proposing several quizzes tailored to what the student covered in class that week/term.

Interventions 1 and 2 consist of different messages. In the full implementation, in a given week, all students will receive the same message in that week. During stage 1, different messages in the same treatment arm will be randomly assigned in order to measure average quiz use across different messages.

**Outcome of interest.** The outcome is a binary variable indicating "success." It is equal to 1 if at least one quiz was completed within 30 hours from the text message being sent, equal to 0 if no quiz was completed from the phone number contacted.

**Covariates for targeting.** We form three strata: primary 1 and 2, primary 3-5, and primary 6. The objective is to pick the messaging type (treatment arm) with the highest average outcome (highest quiz use) for each stratum.

**Notation.** We use the following notation: treatment  $D \in \{1, ..., k\}$ , outcome  $Y \in \{0, 1\}$ , stratum  $X \in \{1, ..., n_x\}$ , student *i*. In our setting, we have k = 4 treatment levels, and the number of strata is equal to  $n_x = 3$ .

**Sample.** The phone on record for the parent is used to select children and identify siblings/members of the same family. All phone numbers are anonymized by the partner before sharing with the researchers. The only data shared is an identifier for the phone, an identifier for the child, and the primary grade level. 6066 children with associated phone numbers in the partner's schools are included in the data set shared.

To avoid spillovers and issues with interpreting results when parents may receive messages about several children (in treatment arms 3 and 4) vs. the quiz system overall (in treatment arm 1 and 2), only one child of a given family is included. For each phone number, one sibling is selected randomly, generating a sample of 4705 children.

We then stratify the sample and randomly select an equal share of each stratum for each of the 5 experimental waves. Finally, we randomly assign children in each stratum and wave to the four treatment arms. In the first wave, assignment to each treatment arm is equal (aside from "misfits" arising from indivisibilities in the total numbers). In all following waves, we use exploration sampling to choose the assignment share to each treatment arm within each stratum (see below).

Assignment according to these assignment shares by stratum is carried out using the *randtreat* command in Stata which carries out factorial assignment within stratum with tie-breaking (see Carril (2017)).

Following a common approach for settings such as ours (cf. Gelman et al. (2014), chapter 5), we specify a hierarchical Bayesian model for learning about the average outcome variable from wave to wave.

**Hierarchical priors.** We employ a prior with statistical dependence between the vectors of average potential outcomes  $\boldsymbol{\theta}^x = (\theta^{1x}, \dots, \theta^{kx})$  across different strata x.

The outcome within each treatment arm d and stratum x has a Bernoulli distribution unknown mean  $\theta^{dx}$ .

$$Y_{it}^d | (X_{it} = x, \theta^{dx}, (\alpha_0^d, \beta_0^d)) \sim Ber(\theta^{dx})$$

For a given treatment d, the parameters  $\theta^{dx}$  are drawn from a common prior distribution (Beta distribution) across strata x with unknown hyperparameters  $(\alpha_0^d, \beta_0^d)$ .

$$\theta^{dx}|(\alpha_0^d,\beta_0^d) \sim Beta(\alpha_0^d,\beta_0^d)$$

The hyper-parameters  $(\alpha^d, \beta^d)$  governing the common prior distribution are assumed to come from a diffuse hyper-prior distribution  $\pi$ 

$$(\alpha_0^d, \beta_0^d) \sim \pi$$

The hyper-prior  $\pi$  has a common density equal to  $(\alpha + \beta)^{-2.5}$ , up to a multiplicative constant, following the recommendation of Gelman et al. (2014) (p. 110).

We assume that parameters  $(\alpha^d, \beta^d, \theta^{d_{\cdot}})$  are independent across the treatment arms d. This means that there is learning about the treatment effect of a given treatment arm across strata (within the treatment arm), but not about average outcomes in a stratum across treatment arms.

In every period, we observe the success rates for all past experimental participants in a given treatment arm. This yields information about the mean and dispersion of the distribution of the  $\theta^{dx}$  for  $x = 1, ... n_x$  and thus a posterior distribution for the hyper-parameters of this treatment arm. This estimate combined with the observed success rates in a given stratum x allow to calculate the posterior distribution of the success parameter  $\theta^{dx}$  in that stratum.

Finally, these posterior distributions can be used to calculate the probabilities  $\hat{p}_t^{dx}$  that a given

treatment is optimal for a given stratum.

In this model there are  $k \cdot n_x = 12$  parameters  $\theta_x^d$  that we wish to estimate, and 2 hyper-parameters.

**Exploration Sampling.** After each wave, we obtain the posterior probability of a treatment d being optimal for stratum x given past trials and successes  $m_{t-1}$  and  $r_{t-1}$ ,

$$p_t^{dx} = P\left(d = \operatorname*{arg\,max}_{d'} \theta^{d'x} | \boldsymbol{m}_{t-1}, \boldsymbol{r}_{t-1}\right)$$

where  $\theta^{dx}$  is the probability of success for an observation in stratum x, when receiving treatment d.

We use a Markov Chain Monte Carlo algorithm described in Caria et al. (2020) Appendix A.2, programmed in an R-based app, to sample from the posterior distribution to obtain the probabilities  $\hat{p}_t^{dx}$ .

We use exploration sampling to transform these probabilities to obtain assignment shares  $q_{t,x}^d$  to treatment d in wave t within stratum x:

$$q_{t,x}^{d} = S_{t,x} \cdot p_{t,x}^{d} \cdot (1 - p_{t,x}^{d})$$

$$S_{t,x} = \frac{1}{\sum_{d} p_{t,x}^{d} \cdot (1 - p_{t,x}^{d})}.$$
(1)

#### Stage 2: Randomized Control Trial

**Sampling and treatment assignment.** Using the full sample of children, families will be stratified by the primary grade level of their child or children, using the 3 strata defined above, and randomly assigned in equal proportions to the treatment and control group.

If for any child in a given family, the most successful treatment type is treatment 3 or 4 (specific quizzes), the parent will receive one message for each child in that treatment type. The message will include the name of the child to attribute the correct quizzes to each child.

If any children in the same family are assigned messages of type 1 or 2 according to Stage 1 results, the family will receive one message for all those treatment arms. Messages will be randomized according to the proportion of children in each treatment arm (to avoid duplicate messaging with similar contents).

Example: suppose a family has five children, 2 in primary 1 and 2, 2 in primary 4 and 5, and 1 in primary 6. Primary 1-2 is assigned treatment arm 3, primaries 3-5 are assigned arm 2, and primary 6 is assigned arm 1. Then this family will receive 3 messages: one for each child in the quiz messaging arm (primary 1 and 2) and one randomly selected message that is with 2/3 probability

treatment arm 2, and 1/3 probability treatment arm 1.

Parents will receive weekly messages during the entire spring term starting in January 2021.

**Outcomes.** The outcome of interest are normalized test scores in midterm and endterm exams (normalization based on prior year grade distribution by primary grade level).

**Decision to conduct.** Based on the estimated quiz uptake rates in Stage 1, the quiz usage rates prior to the experiment, and grade distributions from prior years, we will calculate the minimal average increase in assignment scores we would need to observe for each child who has adopted the quizzes, in order to see a significant treatment effect using traditional frequentist inference for treatment and control group. If this increase is unrealistic, the partner may decide not to conduct the trial. We will update this pre-analysis plan with more detailed power calculations.

### 2 Analysis

#### 2.1 Stage 1

Average potential outcomes from Stage 1. We will report estimates of average potential outcomes, both conditional on covariates and unconditional. That is, we will report posterior estimates of

- $E[\theta_x^d]$  for all d and x,
- $\sum_{x} E\left[\theta_{x}^{d}\right] p^{x}$  for all d, where  $p^{x}$  is the sample share of stratum x.

**Optimal treatment.** The key outcome for our partner is which policy we recommend to implement to scale, either conditional on covariates, or unconditionally:

- $d_x^* = \arg \max_d E[\theta_x^d]$ , and
- $d^* = \arg \max_d \sum_x E\left[\theta_x^d\right] p^x$ .

Inference after Stage 1. Our primary form of inference will be Bayesian, based on the hierarchical default prior described above. We will report 0.025 and 0.975 quantiles for the parameters above, based on MCMC draws, yielding sets that have a posterior probability of 95% to contain the true parameters, conditional on the data of the experiment.

Additionally, we will provide randomization-based p-values that are valid under the sharp null hypothesis that all treatment arms have the same effect, i.e., under the null that  $\theta_x^d = \theta_x^{d'}$  for

all d, d', x. Under this null, we can generate counterfactual data by re-running our assignment algorithm repeatedly, leaving outcomes as they are in our data, but generating new treatment assignments. The distribution of test-statistics over this re-randomization distribution can be used to construct critical values and p-values that are exact in finite samples, under the sharp null.

Inference after Stage 2: quiz use. We will use the same success/failure definition as in Stage 1 to continue monitoring quiz uptake in the 30 hours after messages are sent. We will use Bayesian inference to update the estimate of average quiz use in each treatment arm according to the procedures above, and carry out the same tests as above relative to the control group. Moreover, we will examine if there are time trends in quiz uptake over the semester.

**Inference after Stage 2: exam scores.** We will estimate the ITT of messaging about the quiz platform and the LATE of quiz use relative to the control group on normalized exam scores (by primary grade) using standard frequentist inference.

Treatment	Message type	Template
1	Motivation framing	Dear parent,
		Mobile quizzes are 5-question quizzes on lots of topics in
		Maths, English, Science, or Social Studies. Quizzes provide
		instant feedback and a chance to practice important topics.
		Text 0 to [number] on Whatsapp or click here to take a
		mobile quiz now!
2	Testimonial framing	Dear parent,
		One Lagos parent described how quizzes helped during
		Covid closures: "I am very proud and appreciative of your
		brilliant and excellent initiative to keep our children study-
		ing during the Covid pandemic." Text 0 to [number] on
		Whatsapp or click here to take a mobile quiz now!
3	One specific quiz	Dear parent,
		This week, take the following quizzes with your pupil:
		Maths - Numbers, Words, and Counting - Quiz 3
		Text 0 to [number] on Whatsapp or click here to take a
		mobile quiz now!
4	Many quizzes	Dear parent,
		This week, take the following quizzes with your pupil:
		Maths - Numbers, Words, and Counting - Quiz 7
		Maths - Numbers, Words, and Counting - Quiz 8
		Maths - Numbers, Words, and Counting - Quiz 9
		Maths - Numbers, Words, and Counting - Quiz 10
		Text 0 to [number] on Whatsapp or click here to take a
		mobile quiz now!

Table 1: Template messages

## References

- Caria, A. S., Gordon, G., Kasy, M., Quinn, S., Shami, S., and Teytelboym, A. (2020). An adaptive targeted field experiment: Job search assistance for refugees in jordan. *working paper*.
- Carril, A. (2017). Dealing with misfits in random treatment assignment. Stata Journal, 17(3).
- Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (2014). *Bayesian data analysis*, volume 2. Taylor & Francis.
- Kasy, M. and Sautmann, A. (2020). Adaptive treatment assignment in experiments for policy choice. *Econometrica (forthcoming)*.