

Analysis Plan for EdX Research

Sample

- Of all students enrolled in “The Challenges of Global Poverty,” we include active students only. Active is defined as having attempted at least one finger exercise, homework, or final exam question or having attempted the final project. We only include those students that were active during the period that the course was in session.
- We exclude all course staff, beta testers, tutors, teaching assistants, and MIT students that were enrolled in the course but are not considered to be conventional online students.
- We exclude students who took the entrance survey after the midway point of the course (March 25, 2014) since these students joined after randomization. This applies to 34 students.

Data

- We received a bundle of data from edX that tracks student-level information on enrollment, performance, and course activity. These include:
 - **Registration/enrollment form** with student demographics and date of enrollment
 - **Student module** with attempts and scores for each student question-by-question
 - **Tracking logs** with the date and time of activities such as submitting answers to questions, pausing videos, showing or hiding video transcripts, etc.
 - **Certificates file** with overall course grade and certificate status
- We also collected student-specific information from the following sources:
 - **Entrance survey** with demographics, motivation, and random assignment to self-efficacy (SE) and regular study time (RST) treatment groups, etc.
 - **Exit survey** with reflections on course, level of difficulty of course, performance relative to expectations, etc.
 - **Tutoring** feedback forms, administered to tutors and students assigned to tutoring groups

Primary Outcomes

- **Staying until the end:** Defined as participating in the homework (HW) or one of the finger exercises (FE) for the last unit
- **Earned certificate:** Defined as average course grade of 50% or greater
- **Fraction of course completed:** Fraction of all FE and HW questions attempted
- **Time spent on the course:** Number of hours spent on the course
- **Final grade on the course:** Calculated as 25% for FE (dropping lowest 3 out of 22), 45% for HW (dropping lowest 2 out of 9), 20% for final exam, and 10% for final project (substituting in course team scores instead of self-graded scores)
- **Final exam grade:** Full sample (with zero if did not attempt)
- **Attempted final exam**
- **Final exam grade:** Conditional on attempting

Controls

- **English fluency:** Average z-scores from abilities in reading, writing, speaking, and listening
- **Primary language non- English:** Based primarily on language as reported in the entrance survey, supplementing from the registration form when not available
- **Age**
- **Gender**
- **Poor country:** Based primarily on country as reported in entrance survey, supplementing from registration form and/or geolocation from tutoring signup form when not available
- **Level of education:** Dummy for college degree or higher
- **Parental education:** Dummy for mother or father with college degree or higher
- **Level of economics/policy education:** Scale of 1 (none at all) to 4 (more than one year in college) or 5 (self-taught)
- **Level of familiarity with economics:** Scale of 1 (not at all familiar) to 5 (extremely familiar)
- **Note:** To avoid losing data, any missing control variables will be replaced by zero, and indicator dummies for missing variables will be included for each variable.

Part I: Self-efficacy

- **Sample:** We restrict our sample to students that started the entrance survey and reached the point where they were randomly assigned into one of four groups: control, generic message, language treatment, or gender treatment.
- **Notes:** For students that took the entrance survey more than once, we accept the most recently completed survey. If a student never completed the survey, we take the most recently started survey. We flag those few students that did not complete enough of the entrance survey to see the message that relates to their assigned treatment group.
- **Language treatment:** Regress outcomes on assignment to language prompt ($LTreat$), non-native English speaker ($NonEnglish$), and an interaction term ($LTreat \times NonEnglish$) with and without controls for each student i . The hypothesis is that the overall effect for non-native English speakers ($LTreat + LTreat \times NonEnglish$) is positive.

$$Y_i = \alpha + \beta_1 LTreat_i + \beta_2 NonEnglish_i + \beta_3 (LTreat \times NonEnglish)_i + \epsilon_i$$

$$Y_i = \alpha + \beta_1 LTreat_i + \beta_2 NonEnglish_i + \beta_3 (LTreat \times NonEnglish)_i + \gamma X_i + \epsilon_i$$

- **Gender treatment:** Regress outcomes on assignment to gender prompt ($GTreat$), dummy for female student ($Female$) and interaction term ($GTreat \times Female$) with and without controls for each student i . The hypothesis is that the overall effect for female students ($GTreat + GTreat \times Female$) is positive.

$$Y_i = \alpha + \beta_1 GTreat_i + \beta_2 Female_i + \beta_3 (GTreat \times Female)_i + \epsilon_i$$

$$Y_i = \alpha + \beta_1 GTreat_i + \beta_2 Female_i + \beta_3 (GTreat \times Female)_i + \gamma X_i + \epsilon_i$$

Part II: Tutoring

- **Sample:** The sample consists of students that opted in to the tutoring lottery. Students were allocated into one of 14 groups based on mutual scheduling availability between tutors and students, or to an unallocated stratum if no mutual scheduling was possible. Within these groups, half were randomly assigned to receive a tutor, stratified according to poor country.
- **Tutoring treatment:** Regress outcomes on assignment to tutoring. We include a dummy variable for poor country since randomization was stratified by poor country. We also include a second regression with additional controls. We do not cluster by group as the treatment assignment was at the individual level. The hypothesis is that the coefficient on the tutoring treatment is positive.

$$Y_i = \alpha + \beta Tutor_i + \gamma PoorCountry_i + \epsilon_i$$

$$Y_i = \alpha + \beta Tutor_i + \gamma X_i + \epsilon_i$$

Complementary analysis: Who benefits more from tutoring?

- **Tutoring treatment & college education:** Regress outcomes on assignment to tutoring, college education and interaction term, with same controls as above. Our hypothesis is that the treatment effect is greater for students with a college education.

$$Y_i = \alpha + \beta_1 Tutor_i + \beta_2 College_i + \beta_3 (Tutor \times College)_i + \gamma PoorCountry_i + \epsilon_i$$

$$Y_i = \alpha + \beta_1 Tutor_i + \beta_2 College_i + \beta_3 (Tutor \times College)_i + \gamma X_i + \epsilon_i$$

- **Tutoring treatment & poor country:** Regress outcomes on assignment to tutoring, poor country, and interaction term, with same controls as above. Our hypothesis is that the treatment effect is greater for students from a poor country.

$$Y_i = \alpha + \beta_1 Tutor_i + \beta_2 PoorCountry_i + \beta_3 (Tutor \times PoorCountry)_i + \epsilon_i$$

$$Y_i = \alpha + \beta_1 Tutor_i + \beta_2 PoorCountry_i + \beta_3 (Tutor \times PoorCountry)_i + \gamma X_i + \epsilon_i$$

Complementary analysis: Tutoring treatment & engagement

- **Course staff:** Regress engagement with course staff on assignment to tutoring. Engagement with course staff is defined as level of engagement with any course staff member (with tutor via tutoring sessions or email, questions/responses directed to the staff on the discussion forum, or participation in office hours with TAs).

$$EngageStaff_i = \alpha + \beta Tutor_i + \epsilon_i$$

- **Course staff and forum:** Regress engagement with staff or any forum activity on tutoring.

$$EngageAny_i = \alpha + \beta Tutor_i + \epsilon_i$$

- Our hypothesis is that tutoring increases engagement with course staff and on the forum.

Part III: Regular Study Time

- **Sample:** Restrict sample to students that started or completed the entrance survey and were randomized into one of the four groups.
- **Notes:** Duplicates and incomplete surveys handled in same way as with self-efficacy.
- **Variables:**
 - *RSTdummy* Assigned to any of the RST groups (RST=2,3,4)
 - *RSTopt* Opted in to any of the RST groups (0 if didn't have the option)
 - *EM1dummy* Assigned to see message (RST=3,4)
 - *EM2dummy* Assigned to see option for email reminder (RST=4)
 - *EM2opt* Opted in for reminder email (0 if didn't have the option)

- **Regular study time groups:**

RST Group	RST dummy (offered option to sign up for RST)	RST opt in	EM 1 dummy (message that course team can track usage)	EM 2 dummy (option for email reminder)	EM 2 opt in
1 (Control)	0	0	0	0	0
2 (RST)	1	0 or 1	0	0	0
3 (RST + EM1)	1	0 or 1	1	0	0
4 (RST + EM1 + EM2)	1	0 or 1	1	1	0 or 1

Primary Analysis

- **Reduced form 1:**

- Regress outcomes on assigned RST (*RSTdummy*)

$$Y_i = \alpha + \beta RSTdummy_i + \epsilon_i$$

- **Reduced form 2(a):**

- Regress outcomes on assigned RST and assigned RST * assigned EM1 message (*EM1dummy*) and assigned RST * assigned EM2 study reminder (*EM2dummy*)

$$Y_i = \alpha + \beta_1 RSTdummy_i + \beta_2 EM1dummy_i + \beta_3 EM2dummy_i + \epsilon_i$$

- **Reduced form 2(b) (EM1 and EM2 conditional on opting in to RST):**

- Regress outcomes on assigned RST and opted in to RST * assigned EM1 and opted in to RST * assigned EM2

$$Y_i = \alpha + \beta_1 RSTdummy_i + \beta_2 (RSTopt_i \times EM1dummy_i) + \beta_3 (RSTopt_i \times EM2dummy_i) + \epsilon_i$$

- **First stage, Opt in to RST:**

- Regress opted for RST (*RSTopt*) on assigned RST

$$RSTopt_i = \alpha + \beta RSTdummy_i + \epsilon_i$$

- Regress opted for RST on assigned RST and assigned RST * assigned EM1 message and assigned RST * assigned EM2

$$RSTopt_i = \alpha + \beta_1 RSTdummy_i + \beta_2 EM1dummy_i + \beta_3 EM2dummy_i + \epsilon_i$$

- **First stage (a), Opt in to email reminder:**

- Regress opted for email reminder (*EM2opt*) on assigned RST and assigned RST * assigned EM1 message and assigned RST * assigned EM2

$$EM2opt_i = \alpha + \beta_1 RSTdummy_i + \beta_2 EM1dummy_i + \beta_3 EM2dummy_i + \epsilon_i$$

- **First stage (b), Opt in to email reminder (conditional on opting in to RST):**

- Regress opted for email reminder (*EM2opt*) on assigned RST and opted in to RST * assigned EM1 message and opted in to RST * assigned EM2

$$EM2opt_i = \alpha + \beta_1 RSTdummy_i + \beta_2 (RSTopt_i \times EM1dummy_i) + \beta_3 (RSTopt_i \times EM2dummy_i) + \epsilon_i$$

Secondary Analysis: Intermediate variable: Study time

- Regress total time spent on course on assigned to any RST. We repeat the analysis with and without those who drop out after the survey (For those who drop out, we assigned zero study time after they have dropped out. Dropping out is defined as not participating in the HW or one of the FE assignments for the final unit, as in the variable for staying until the end.). We also include regressions restricted to the first half of the course or the second half of the course.

$$TotalTime_i = \alpha + \beta RSTdummy_i + \epsilon_i$$

- Regress total time spent on course on assigned to any RST and assigned RST * assigned EM1 message and assigned RST*assigned EM2, also repeating restricted to the first half and second half of the course.

$$TotalTime_i = \alpha + \beta_1 RSTdummy_i + \beta_2 EM1dummy_i + \beta_3 EM2dummy_i + \epsilon_i$$

- Regress study at planned time on opted in to any RST. Studying at the planned time is defined as the fraction of the number of the weeks of the course when the student studied within 30 minutes of one of their designated study times.

$$StudyAtRST_i = \alpha + \beta RSTopt_i + \epsilon_i$$

- Regress study at planned time on opted in to any RST and assigned RST * assigned EM1 message and opted in to RST * opted into EM2

$$StudyAtRST_i = \alpha + \beta_1 RSTopt_i + \beta_2 EM1dummy_i + \beta_3 EM2opt_i + \epsilon_i$$

- The hypothesis is that the overall effect (on performance and time spent on the course) for students who commit to a regular study time will be positive. Similarly, we hypothesize that the effect will be greater for students assigned to see the enforcement message that the course staff can track usage and greater still for those that also opt in to receive an email reminder to stick to their regular study time.