

README

Code and Data for "Improving Student Outcomes through Adaptive Learning Platforms: Experimental Evidence from the Dominican Republic"

by Carolina Lopez and Astrid Pineda

Overview

This document provides the information needed to replicate all tables and figures in "Improving Student Outcomes through Adaptive Learning Platforms: Experimental Evidence from the Dominican Republic".

The replication package is organized into the following folders:

- Tables/
- Figures/
- Data/
- Output/

Files and Contents

All analysis is conducted in Stata. The root folder of the replication package contains the following do files and auxiliary files:

00_master_replication.do: Master script. Running this file from top to bottom reproduces all tables and figures in the paper. Before running, set the global main path at the top of the file to the location of the Reproducibility_Package folder on your computer (see "How to Run" below).

00_RI.do: Creates the 1,000 permuted datasets used for randomization inference. Takes approximately 5–10 minutes. Called automatically by 00_master_replication.do.

02_cal_tutoring_sessions.do: Prepares the implementation fidelity dataset by importing and reshaping CAL and tutoring session records. Called automatically by 00_master_replication.do if the deposit dataset does not already exist.

replication_outputs.tex: A standalone LaTeX file that compiles all tables and figures from the paper in order (main text, Appendix A, Appendix B). To use it, create two subfolders next to this file: tables/ (copy all .tex files from Output/tables/) and figures/ (copy all .png files from Output/graphs/). Then compile with pdflatex. Note: two static figures (Timeline_CAL.png and rand_design.png) must also be placed in figures/ manually.

A. Tables/

Contains one do file per table in the paper (22 scripts total, including one script for in-text statistics). Each script loads CAL_DR_Data_deposit.dta and the RI permuted datasets, runs OLS regressions on the observed data and on each of the 1,000 permutations, computes randomization inference standard errors and p-values, and exports a LaTeX table to Output/tables/.

B. Figures/

Contains one do file per figure (9 scripts total). Each script produces one code-generated figure and saves it as a .png file to Output/graphs/: hist_hours_cal.do (Figure 3), hist_topics_learned.do (Figure 4), math_beliefs_by_group.do (Figure 5), te_perc_bl.do (Figure 6), density_tutees.do (Figure 7), math_prop_by_group.do (Figure 8), hist_CAL_groups.do (Figure 9), self_efficacy_by_group.do (Figure A1), binscatter_returns.do (Figure A2).

C. Data/

Contains all datasets required to run the replication code:

CAL_DR_Data_deposit.dta: The main analysis dataset used in all regressions. This is a student-level dataset combining information from baseline, midline, and endline surveys, administrative records, and ALEKS platform data. Each row corresponds to one student. Key variables include: treat (=1 if classroom assigned to CAL treatment), treat2_class (=1 if classroom assigned to CAL + Tutoring), treat2 (=1 if student selected to receive tutoring), std_math_bl/ml/el (standardized math test scores at baseline, midline, and endline), pair (classroom pair identifier used in RI), School_ID and Section (school and classroom identifiers).

cal_tutoring_merged_long_deposit.dta: Student-session-level dataset with weekly records of CAL sessions delivered and tutoring sessions delivered, covering 19 CAL classrooms over 21 weeks of implementation.

RI/ (subfolder): Contains CAL_data_estimation_RI_1.dta through CAL_data_estimation_RI_1000.dta. These are permuted versions of CAL_DR_Data_deposit.dta, each with a different random treatment assignment drawn from the same pairwise randomization procedure used in the study. Generated by 00_RI.do (seed 12345). Required by all table scripts.

D. Output/

This folder is created automatically by the replication code (it may be empty or absent initially). It contains two subfolders:

tables/: LaTeX table files (.tex). Produced by the scripts in Tables/. Contains 21 .tex files corresponding to Tables 1–8, A1, and B1–B12 in the paper.

graphs/: Figure files (.png). Produced by the scripts in Figures/. Contains 9 .png files corresponding to Figures 3–9 and A1–A2 in the paper.

Software Requirements

- Stata 16 or later
- The following user-written Stata packages (install via SSC):
 - ssc install cibar
 - ssc install grc1leg2
 - ssc install binscatter

```
ssc install estout
```

```
ssc install blind schemes (provides the plotplainblind graph scheme)
```

How to Run

1. Open 00_master_replication.do in Stata.
2. Set the path at the top of the file:

```
global main "path/to/your/CAL-DR/folder"
```

3. Run 00_master_replication.do.

Step 0 (00_RI.do) generates the 1,000 RI datasets and takes approximately 5-10 minutes. Once the datasets are created, this step can be commented out if the package is re-run.

Tables are saved to: [main]/Output/tables/

Figures are saved to: [main]/Output/graphs/

Note on figures: Two figures in the paper are not code-generated and are included as static image files in the package: Timeline_CAL.png (Figure 1: study timeline) and rand_design.png (Figure 2: randomization design).

Relevant Sample

The intervention was implemented between October 2023 and May 2024 in 38 ninth-grade classrooms across 9 public schools in Santo Domingo and Santiago, Dominican Republic. Treatment was assigned via pairwise randomization at the classroom level: classrooms were ranked by size, paired with the nearest-size classroom, and one classroom in each pair was randomly assigned to treatment. A second stage of pairwise randomization within the treatment group assigned the tutoring component.

The final study sample comprised 38 classrooms: 19 assigned to CAL treatment (ALEKS software), 9 of those also assigned to receive small-group tutoring, and 19 control classrooms. Of the approximately 20% of at-risk students identified in the 9 tutoring classrooms, 61 students were selected to receive tutoring.

Baseline surveys were collected in September 2023 (N = 1,333 students). Midline surveys were collected in December 2023; 92% of the baseline sample completed surveys and tests at midline. Endline surveys were collected in May 2024; 90% of the baseline sample completed surveys and tests at endline. Attrition checks are reported in Table 2 of the paper; Lee bounds are reported in Appendix Table A1.

Variable Construction

All variables are described in the main text and appendices of the paper. Key constructed variables are created within each do file prior to estimation. In particular:

- Math test scores are standardized to have mean 0 and standard deviation 1 in the control group at each time period.

- Index variables (grit, locus of control, self-efficacy, math anxiety, teacher behavior, etc.) are constructed within each respective table script and standardized to have mean 0 and standard deviation 1 in the control group at baseline, unless otherwise noted.
- The overconfidence measure is the difference between a student's predicted test score (reported before taking the endline test) and their actual test score.
- Missing covariate values are replaced with the sample mean and a missing-value indicator is added to the covariate vector.
- Randomization inference p-values are computed as the share of 1,000 simulated treatment effect estimates (under the null of no treatment effect) that are at least as extreme in absolute value as the observed estimate.

References

Lopez, Carolina, and Astrid Pineda. "Improving Student Outcomes through Adaptive Learning Platforms: Experimental Evidence from the Dominican Republic." AEA RCT Registry: AEARCTR-0012383. IRB Protocol: Teachers College Columbia University, #24-444.

Data Availability Statement

The data used in this paper are primary data collected by the authors through a randomized controlled trial conducted between September 2023 and May 2024 in 9 public schools in the Dominican Republic. The datasets were not obtained from third-party sources and cannot be reproduced from publicly available materials.

The replication package includes de-identified versions of all datasets required to reproduce the results. Restricted access to the datasets is available through the World Bank Microdata Library (Licensed Access). Researchers may apply for access at: <https://microdata.worldbank.org>

Detailed information for each dataset:

1. File name: CAL_DR_Data_deposit.dta

Description: Student-level analysis dataset. Combines baseline, midline, and endline survey data, administrative records (grades, attendance, enrollment), and ALEKS platform usage data for 1,333 ninth-grade students across 38 classrooms in 9 schools.

Citation: Lopez, Carolina, and Astrid Pineda (2024). "Improving Student Outcomes through Adaptive Learning Platforms: Experimental Evidence from the Dominican Republic." AEA RCT Registry: AEARCTR-0012383.

Access: Licensed access via World Bank Microdata Library. Data collection took place September 2023 - May 2024.

Access restrictions: Licensed access. Researchers must apply through the World Bank Microdata Library. The dataset has been de-identified: open-text responses, phone numbers, and direct school identifiers have been removed; School_ID has been replaced with an anonymous sequential code.

2. File name: cal_tutoring_merged_long_deposit.dta

Description: Section-week-level dataset with weekly records of CAL sessions delivered and tutoring sessions delivered, covering 19 CAL classrooms over 21 weeks of implementation. Includes session counts, rescheduling flags, and reasons for cancelled sessions.

Citation: Lopez, Carolina, and Astrid Pineda (2024). Ibid.

Access: Licensed access via World Bank Microdata Library. Data collection took place October 2023 - May 2024.

Access restrictions: Licensed access. School names have been removed; school_id has been replaced with an anonymous sequential code.

List of Exhibits

The following table maps each output file generated by the replication code to the corresponding table or figure in the paper.

Output file	Paper exhibit	Description
Tables		
Output/tables/balstab.tex	Table 1	Balance checks
Output/tables/attrition_table.tex	Table 2	Attrition analysis
Output/tables/std_math_ml_el.tex	Table 3	Treatment effects on math scores (midline and endline)
Output/tables/std_math_at_below_el.tex	Table 4	Treatment effects by content level (below / at grade)
Output/tables/std_math_area_ml_el.tex	Table 5	Treatment effects by TIMSS content area
Output/tables/std_math_domain_ml_el.tex	Table 6	Treatment effects by TIMSS cognitive domain
Output/tables/overconfidence_ml_el.tex	Table 7	Overconfidence in anticipated test performance

Output/tables/dose_response.tex	Table 8	Dose-response: IV estimates of platform engagement
Output/tables/lee_bounds_single_reg_std_math_ml_el.tex	Table A1	Lee bounds (Appendix A)
Output/tables/math_intrinsic_motivation_el.tex	Table B1	Intrinsic motivation (Appendix B)
Output/tables/math_preferences_ml_el.tex	Table B2	Math preferences (Appendix B)
Output/tables/absence_dropout_el.tex	Table B3	Absence and dropout (Appendix B)
Output/tables/math_self_concept_el.tex	Table B4	Math self-concept (Appendix B)
Output/tables/self_efficacy_ml_el.tex	Table B5	Self-efficacy (Appendix B)
Output/tables/grit_el.tex	Table B6	Grit (Appendix B)
Output/tables/logic_task_el.tex	Table B7	Perseverance / real-effort logic task (Appendix B)
Output/tables/locus_of_control_el.tex	Table B8	Locus of control (Appendix B)
Output/tables/depression_el.tex	Table B9	Depression index (Appendix B)
Output/tables/math_anxiety_el_aux.tex	Table B10	Math anxiety (Appendix B)
Output/tables/teacher_behavior_el.tex	Table B11	Teacher behavior indices (Appendix B)
Output/tables/time_use_el.tex	Table B12	Student time use (Appendix B)
Figures		
[static] Timeline_CAL.png	Figure 1	Study timeline (not code-generated)
[static] rand_design.png	Figure 2	Randomization design (not code-generated)
Output/graphs/hist_hours_cal_notes.png	Figure 3	Time spent on CAL platform
Output/graphs/hist_topics_learned_notes.png	Figure 4	Topics learned on CAL platform
Output/graphs/math_beliefs_by_group.png	Figure 5	Expected vs. actual performance by group
Output/graphs/te_perc_bl.png	Figure 6	Non-parametric treatment effects by baseline performance

Output/graphs/density_plot_std_math_bl_tutees.png	Figure 7	Selection of tutees and baseline performance distribution
Output/graphs/math_prop_bl_el_by_group_tutees_pcontrol.png	Figure 8	Share of correct answers by performance group
Output/graphs/hist_CAL_groups_bnw.png	Figure 9	Platform use by treatment group
Output/graphs/self_efficacy_hist_CAL_groups_bnw.png	Figure A1	Self-efficacy by performance group (Appendix A)
Output/graphs/binscatter_time_test_scores_el.png	Figure A2	Returns to hours of CAL use (Appendix A)