README: Replication for "Testing the promise of digital scaling: in-person vs. app-based training for women entrepreneurs" (Cassidy et al 2024)

Overview:

The code in this replication package constructs all tables and figures from the working paper using Stata. The package consists of four Stata do-files. All figures and tables but two can be constructed using the code from the primary Stata do-file. Two raw datasets from the World Bank Microdata Library should be included in the package, one for the main panel, and one for the screening applicants. The "clean" do file will generate variables needed for analysis, and two new datasets with these variables. The replicator should expect the code to run for 15 minutes or less.

Data and Code Availability Statement:

This paper uses data from the baseline and endline survey of an impact evaluation of an app-based v/s in-person business training for women entrepreneurs.

Figure A3 uses a screenshot from a confidential administrative portal to track user activity. This data source cannot be made publicly available.

Statement about Rights:

- I certify that the author(s) of the manuscript have legitimate access to and permission to use the data used in this manuscript.

Summary of availability:

- Some data cannot be made publicly available.

Data Sources:
Two raw datasets should be added in the reproducibility package: the main panel survey including a baseline and endline phone survey, carried out with women entrepreneurs, and a screening survey, carried out with women entrepreneurs who were contacted to determine eligibility into the study. Additional information on each dataset can be found in the table below:

File name:	**Location:**	**Description:**
application_data_anon.dta	`user'/Data	N=6036 women entrepreneurs. Phone survey consisting of some characteristics and other variables needed to determine eligibility for the app-based training
elearning_bl_fu_anon_clean.dta	`user/Data	N=2000 women entrepreneurs

		Raw merged dataset that includes in person baseline survey data, administrative data from the trainings, and the endline phone survey data.

Software Requirements:

This replication package requires Stata (code was last run with version 14). The Stata do-file will require the following packages to be installed (which the master do file should take care of):

Winsor; mat2txt; estout; ietoolkit; kmatch; ritest; moremata; ivreg2; ranktest; pdslasso; sg97_5; st0085_2; lassopack

Description of code:

This package consists of four Stata do-files. All figures and data in the paper can be generated using only the *master_replication.do* file. The files and their purposes are as follows:

- Code in *master_replication.do,* specified directories, installs packages, and generates all figures and tables in the paper. It references *clean.do, analysis.do,* and *appendix.do* and is the only Stata file that needs to run for the replication.
- Code in *clean.do* generates the variables and datasets used for analysis from the two raw data files described above. It is called in *master_replication.do* and does not need to be run separately.
- Code in *analysis.do* generates one figure and four tables in the paper. It is called in *master_replication.do* and does not need to be run separately.
- Code in *appendix.do* generates two figures and 13 tables in the online appendix of the paper. It is called in *master_replication.do* and does not need to be run separately.

Instructions for replicators:

- Download the two datasets in the replication package and place in the 'Data folder'
- Edit the file path in line 36 in the *master_replication.do* file to reflect the folder that will be used for replication.
- Install required packages as necessary.
- Run only *master_replication.do* to generate all tables and results.
- NOTE: One table and one Figure in the online appendix are constructed by hand, and are therefore not included in this package. See additional details below:

Figure/Table #	**Program**	**Line Number**	**Output file**	**Notes**
Table 1	analysis.do	220	Table_1_balance_mat.tex	Table is manually

				formatted for optimal size
Table 2	analysis.do	262	Table_2_Impacts_takeup.tex	Table is manually formatted for optimal size
Table 3	analysis.do	377	Table_3_Impacts_all.tex	Table is manually formatted for optimal size
Table 4	analysis.do	419	Table_4_Takeup_predict.tex	Table is manually formatted for optimal size
Figure 1	analysis.do	455	Fig_1_Course_hist.png	
Table A1	appendix.do	161	Table_A1_application_ttests.tex	Table is manually formatted for optimal size
Table A2	appendix.do	173	Table_A2_balance_all4.tex	Table is manually formatted for optimal size
Table A3	appendix.do	215	Table_A3_Impacts_takeup_4g.tex	Table is manually formatted for optimal size
Table A4	appendix.do	266	Table_A4_Impacts_primary_4g.tex	Table is manually formatted for optimal size
Table A5	appendix.do	338	Table_A5_Impacts_secondary_4g.tex	Table is manually formatted for optimal size
Table A6	appendix.do	383	Table_A6_Impacts_primary_nocontrol.tex	Table is manually formatted

				for optimal size
Table A7	appendix.do	435	Table_A7_Impacts_secondary_nocontrol.tex	Table is manually formatted for optimal size
Table A8	appendix.do	517	Table_A8_Impactsmatching.tex	Table is manually formatted for optimal size
Table A9	appendix.do	569	Table_A9_attrition.tex	Table is manually formatted for optimal size
Table A10	appendix.do	636	Table_A10_Impactshetero_services.tex	Table is manually formatted for optimal size
Table A11	appendix.do	636	Table_A11_Impactshetero_profitsmed.tex	Table is manually formatted for optimal size
Table A12	appendix.do	636	Table_A12_Impactshetero_digskillsmed.tex	Table is manually formatted for optimal size
Table A13	appendix.do	697	Table_A13_Impactshetero_post_secondary.tex	Table is manually formatted for optimal size
Figure A1	appendix.do	845	Fig_A1_quantile_reg_app.png	Table is manually formatted for optimal size
Figure A2	appendix.do	851	Fig_A2_quantile_reg_person.png	Table is manually formatted for optimal size

Table A14	N/A	N/A	N/A	Manually generated, no code used
Figure A3	N/A	N/A	N/A	Manually generated, no code used. See raw excel file.