

Documentation for Data and Programs
“Bringing Education to Afghan Girls: A Randomized-Controlled Trial of Village-Based Schools”
Dana Burde and Leigh L. Linden

This file documents the structure of the programs and data used to create the statistical results presented in the paper.

There are two primary data sets used in the analysis:

1. **Child-Level Data Set.** This is labeled “afghanistan_anonymized_data.dta”. This data set used for all of the analysis except for panels A and B of Table 1. The data is at the child level.
2. **Household-Level Data Set.** This is labeled “afghanistan_anonymized_hh_data.dta”. This data set is only used to produce panels A and B of Table 1. It contains all households that lived in the villages at the time of each survey, and is at the household level.

Both data sets are stored in STATA version 11 format, and each variable is defined in the label.

There are three groups of programs:

1. **Table programs.** This is “Tables.do”. It contains the code to produce all of the tables in the paper.
2. **Wild Cluster Bootstrap.** These five programs use the Wild-Cluster Bootstrap procedure described in the paper to estimate the p-values for the primary specifications in the paper. “WildC_Boys.do” creates the estimates for the fall 2007 results for boys, and “WildC_Boys_s08.do” creates the estimates for the spring 2008 test. The programs labeled “Girls” create the same respective estimates for girls. “WildC_Gender.do” estimates the relative difference between boys and girls for both surveys.
3. **Randomization Inference.** This is “Fisher_Test.do”. It uses randomization inference to estimate the exact p-values for the primary specifications in the paper.

The programs are all independent of each other. They can be run in any order.