

# Closing the gap: Effect of a gender quota on women's access to education in Afghanistan

Rafiuddin Najam

The code in this replication package constructs the dataset, analytical tables and figures for this paper using Stata. Five do files are used to create the dataset(s), tables (excluding Table A2.1), and figures (only figure #1 and event study results), which can be run by one `main.do` file. The entire code should run in less than 3 minutes on Stata 17 and 18 SE—macOS, M1 with Memory 16GB.

## Data Accessibility Statement

### Data sources

Three data sources were used to construct the analytical sample for this paper.

1. Kankor matriculation raw datasets were obtained from publicly available sources in late 2020, including [here](#) and [here](#)—some of the information has been removed from these websites after the significant changes in August 2021 in Afghanistan. See Data section and Data Appendix for further details about the cleaning process. The Kankor matriculation datasets can be accessed [here](#); however, the cleaned applicant level matriculation data compiled by the author—which includes treatment status indicator for the concentration and matches applicants to their district—is not provided<sup>1</sup>. The `01_PrepData.do` file which clean and construct concentration/major level datasets is provided, but cannot be run. Similarly, the chunk of codes to produces Panels E and F for Table 1 in `02.1_Analysis_Descriptive.do` and Table B1.1 in `03_Appendix.do`, can not be run, as it uses the cleaned individual l level data.
2. List of school database with district identifiers was obtained from Ministry of Education website, which can be accessed from the <https://moe.gov.af/> → Access to Information → Statistics → Profiles<sup>2</sup>. Kankor applicants were manually matched to their origin district by the author using the name of school and province reported in the Kankor matriculation dataset. For further details, see Data section.
3. District level monthly nightlight data was acquired from the Goggle Earth Engine [here](#)—the cleaned nightlight data is provided in `data` folder. You might need to have a Google Earth Engine account to view the script. The code cannot run, as it uses National Statistics and Information Authority district level shape file for which I do not have explicit permission to distribute.

---

<sup>1</sup>To minimize the risk of identifying applicants at the district level. However, individual-level de-identified data can be shared by the author on a case-by-case basis.

<sup>2</sup>Once on this page, you need to change the site into one of the two local languages (Dari and Pashto) to access the list of schools for each province. The list of schools is published in local languages (Dari and Pashto).

## Summary of Data

File	Source	Notes	Data
<i>FD_Uni_eng.xlsx</i> <i>FD_Non_uni_eng.csv</i> <i>uni_admitted_DIST.dta</i> <i>uni.dta</i> <i>nonuni.dta</i> <i>kankor_all.dta</i>	Raw Kankor datasets	Cleaning .do provided	No
<i>uni_by_concentration.dta</i>	constructed from above datasets	Cleaning .do provided	Yes
<i>uni_by_major.dta</i>	constructed from above datasets	Cleaning .do provided	Yes
<i>Descriptive_tabel.xlsx</i>	Compiled from: National Yearly Statistics Books, MoHE, and MoE	N/A	Yes

## Description of folder contents

The folder structure of this document is as follows:

- **data:** Holds the cleaned dataset at the concentration and major level for primary outcomes, including cleaned data for descriptive statistics.
- **outputs:** Has two subfolders (1) figures (2) tables, which holds all the analytical tables and figures produced in the analysis, respectively. Additionally, the `outputs/tables` encompasses all regression results and has two subfolders (a) `outputs/tables/appendix`—store all appendix related results except event study figures—and (b) `outputs/tables/table_1`—store descriptive output for table 1.
- **dofiles:** Holds all the Stata .do files. Including Stata ado files.

## Computational requirements

### Software requirements

Stata is used for the analysis in this project. This project is written in Stata 17 and 18 SE and has not been tested on older versions. The codes are ran on macOS, M1.

## Packages dependencies

All the dependencies packages for the Stata are listed in the `main.do` file.

## Instruction for replication

Note: The final formatting of the Tables are done in a LaTeX file using the exported results. For example, I exported coefficients for each estimator type in Table 1, but for the purpose of presentation and referencing those results are compiled in one Table in the main paper. In this process, some of the values are rounded to the nearest integer. Any differences between the tables produces in this folder and what appears in the paper should be due to this rounding.

All the `.do` files can be run by the `main.do`. You only need to change the path directory at the top of the `main.do` file.

File	Remarks
<code>01_PrepData.do</code>	Prepare primary dataset for analysis from the cleaned version of the Kankor matriculation
<code>02.1_Analysis_Descriptive.do</code>	Produces Figure 1 and Table 1 (excluding the values on columns 3 and 4 in Table 1, which are manually computed—see below for computation direction).
<code>02.2_Analysis_EventStudy.do</code>	Create event study Figures reported in Appendix B.4.
<code>02.3_Analysis_Regression.do</code>	Create regression results for all type of estimators reported in Tables 2, 3, and 4. The results for each type of estimator is stored in a separate file in the <code>outputs/tables</code> , which are clearly labeled with the Table number and estimator name. Additionally, p-value is exported for Sun-Abraham (2020) estimator in the output but for the purpose of consistency indicated by * in main paper.
<code>03_Appendix.do</code>	Create all appendix tables except Tables A2.1; A4.1, and A4.2, which store the results in <code>outputs/tables/appendix</code> .

Table 1 *manual calculations*:

1. **Column 3** in Table 1, represents share of female applicants for Panels A–D. It is computed by this formula  $(\text{female} * 100) / \text{Total}$  for each panel. Panels E and

F, show the average score ratio for female-to-male applicants computed by (female score / male score).

2. **Column 4** represents the share of each row (female, male, total) in Panels B–D, compared to the number of female, male, and/or total students graduates from high school reported in Panel A. For example; 79% in Column 4 for Panel B is calculated by  $(62.24 * 100) / 78.53$ .