Obtaining and Building Dataset:

Data for this project can be obtained through an agreement with the Washington DC Public Schools (DCPS). The data were provided to the University of Virginia as part of several agreed projects, though that agreement has ended and the data are no longer stored at the University of Virginia.

DCPS provided four main datasets:

1. Detailed data on teacher observations for each year, including who conducted them, the day they were conducted, and which of the five annual evaluations was being conducted.

2. Teacher panel with details about teacher characteristics, such as limited race categories, teacher experience, and teacher pay grade.

3. Student panel: student data, including their performance on test scores starting in third grade, student characteristics, and limited data on student expulsions for 2011 and 2012.

4. Student-to-Teacher Roster: certified records of which teachers taught which students in each year, and for what portion of the school year.

Additionally, the district was able to provide suspension and expulsion records for 2011 and 2012.

These datasets are combined in `BUILD_MAIN.do` (and the do-files it calls) to create the analysis dataset. In the final dataset, each row was a student-teacher-year observation, maintaining the student information as well as their teacher characteristics, that teacher's observation information, etc. The main build file provides clarity on what each do-file does.

Analysis Files:

All of the analysis is conducted in the main analysis file, "`00_MAIN_ANALYSIS.do`." Once the data is constructed as described, the main analysis file will call all of the appropriate do-files. The do-files with `_Figure` prefixes are self-contained and called from the main analysis file and produce figures, and similarly for `_Table` prefixes. These files are labeled in Dataverse to match the corresponding table or figure in the manuscript, even though the file names remain the original name. "`00_CreateInsample.do`" is a short program that establishes a consistent sample for all analyses, and "`00_generateFrames.do`" does the final data sorting and variable creation and stores various data frames for use.

Helper functions:

`calculate_q_values.do` – provides a contained program that creates sharpened q-values
`stack_results.do` – a help function to stack items in the summary statistics table
`Randomization Inference.do` – a function that uses a regression call and generates randomization inference