

README.DOC

The data used in the paper are owned by Chicago Public Schools (CPS). Interested researchers may contact CPS to request access to the student-level data that we used in this study (<http://cps.edu/Research/Pages/Research.aspx>).

STATA v15.1 was used throughout with the exception of the map which was created using R.

*** Data files used in estimates ***

- analytic_replication.dta – All data including non-applicant students.
- fullestsample_max.dta – Applicants only. Centers each student around the application school with the highest centered score, calculates indicators for half-standard deviation sample, generates tier indicators and interactions.

*** Paper table creation files ***

Program: **tables1and2and3.do** – data description tables

Output created:

- tables1and2and3.log
- table1.txt
- table3.xml

Data files accessed:

- fullestsample_max.dta
 - analytic_replication.dta
-

Programs: **tables4and5_overall.do**—main estimates for the overall impacts
 tables4and5.do—main estimates for the tier-level impacts

Output created:

- tables4and5_overall.log
- tables4and5_overall.xml
- tables4and5.log
- tables4and5.xml

Data files accessed:

- fullestsample_max.dta
-

Programs: **table6_overall.do**—overall estimates using the free-/reduced-price lunch sample
 tables6.do—tier-level estimates using the free-/reduced-price lunch sample

Output created:

- table6_overall.log
- table6_overall.xml
- table6.log
- table6.xml

Data files accessed:

- fullestsample_max.dta
-

*** Paper figure creation files ***

Program: **figure1.do**—distribution of centered application scores

Output created:

- figure1.log
- figure1.gph
- figure1_ptstotlt895.gph (figure dropping very top scoring students, not in paper)

Data files accessed:

- analytic_replication.dta
-

Program: **figure2.do**—probability of enrolling in a SEHS by centered application score

Output created:

- figure2_tier1.gph
- figure2_tier4.gph

Data files accessed:

- fullestsampl_max.dta
-

Program: **rdplot_tables4and5.do**—Figures 3A – 3C rdplots corresponding to estimates in tables 4 and 5

Output created:

- var_outcome_tier1.gph for all outcome variables *var* for tier 1
- var_outcome_tier4.gph for all outcome variables *var* for tier 4

Data files accessed:

- fullestsampl_max.dta
-

Program: **figure4.do**—boxplots of the admitted scores by SEHS and tier

Output created:

- figure4.gph
- figure4_frpl.gph (figure for using the free-/reduced-price lunch sample, not in paper)

Data files accessed:

- fullestsampl_max.dta
-

*** Appendix table creation files ***

Program: **appendix_table4.do**—sureg test for discontinuities in pretreatment characteristics

Output created:

- appendix_table4.log

Data files accessed:

- fullestsampl_max.dta
-

Programs: **appendix_table5_overall.do**—cutoffs similar across tiers, overall estimates

appendix_table5.do—cutoffs similar across tiers, tier-level estimates

Output created:

- appendix_table5_overall.log
- appendix_table5_overall.xml
- appendix_table5.log
- appendix_table5.xml

Data files accessed:

- fullestsampl_max.dta
-

*** Appendix figure creation files ***

Program: **tiermaps.R**—Appendix Figure 1 map of Chicago census tracts with tier shading and SEHS locations. Map excludes Hancock which is converted to SEHS after period analyzed.

Public data files accessed: (from <https://github.com/open-city/cps-tiers/tree/master/data>)

- tiers1213_1516.csv combines data from publicly available files below:
 - Chicago Public Schools 2013 Tiers and 2010 Chicago Census Tract Data.csv
 - Chicago Public Schools 2015-2016 Enrollment Tiers and Census Tract Socioeconomic Data for the City of Chicago PopStats 2013 Estimates.csv
 - CensusTractsTIGER2010.dbf
 - CensusTractsTIGER2010.prj
 - CensusTractsTIGER2010.sbn
 - CensusTractsTIGER2010.sbx
 - CensusTractsTIGER2010.shp
 - CensusTractsTIGER2010.shx
 - sehsxy.csv (latitude and longitude of SEHSs)
-

Program: **appendix_figure2.do**—Appendix Figure 2 predicted ACT score figures

Output created:

- appendix_figure2.log
- appendix_figure2_tier1.gph
- appendix_figure2_tier4.gph

Data files accessed:

- analytic_replication.dta
-

Program: **rdplot_pretreat_all.do**—Appendix Figures 3A – 3C rdplots of pretreatment characteristics

Output created:

- black8_pretreat_tier1.gph and black8_pretreat_tier4.gph
- hisp8_pretreat_tier1.gph and hisp8_pretreat_tier4.gph
- male8_pretreat_tier1.gph and male8_pretreat_tier4.gph
- frpl8_pretreat_tier1.gph and frpl8_pretreat_tier4.gph
- assignedsch8_pretreat_tier1.gph and assignedsch8_pretreat_tier4.gph

Data files accessed:

- fullestsampl_max.dta
-

Program: **rdplot_tables4and5.do**—Appendix Figures 4A – 4E rdplots corresponding to estimates in tables 4 and 5.

Output created:

- *var_outcome_tier1.gph* for all outcome variables *var* for tier 1
- *var_outcome_tier4.gph* for all outcome variables *var* for tier 4

Data files accessed:

- *fullestsample_max.dta*
-

*** In text information creation files ***

Program: **hsgrowth_byes.do**—national high school growth percentiles for high performing students from low- and high-SES neighborhoods

Output created:

- hsgrowth_byes.log
- HSgrowth_highach1.gph
- HSgrowth_highach4.gph

Data files accessed:

- epasgrowth.dta (created from publicly available data)
 - enrollment and demographic data files from CPS for falls 2010, 2011, and 2012:
 - allm910.dta
 - allm911.dta
 - allm912.dta
 - latent8.dta (predicted grade 8 test scores from UChicago Consortium based on CPS data)
-

Program: **exitentry.do**—text and footnote 11 numbers

Output created:

- exitentry.log

Data files accessed:

- applications_20180120.dta (needed to capture applicant students who never get assigned a student identifier)
 - cutscores.csv (data from Appendix Table 2)
 - analytic_replication.dta
-

Program: **porder_selection.do**—polynomial order selection, footnote 14

Output created:

- porder_select.txt
- porder_select_tier1.txt
- porder_select_tier4.txt

Data files accessed:

- fullestsampl_max.dta
-

Program: **firststage.do**—in text information on first stage

Output created:

- firststage.log

Data files accessed:

- fullestsampl_max.dta
-

Program: **footnote16.do**—information on non-admitted student enrollment in SEHSs

Output created:

- footnote16.log

Data files accessed:

- analytic_replication.dta
-

Program: **selectivecollege.do**—using alternative definitions of selective college

Output created:

- selectivecollege.log

Data files accessed:

- fullestsample_max.dta
-

*** Mapping between information in each table, figure, and in-text number to programs ***

Information	Program
Table 1	tables1and2and3.do
Table 2	tables1and2and3.do
Table 3	tables1and2and3.do
Table 4	tables4and5_overall.do and tables4and5.do
Table 5	tables4and5_overall.do and tables4and5.do
Table 6	table6_overall.do and table6.do
Figure 1	figure1.do
Figure 2	figure2.do
Figures 3A – 3C	rdplot_tables4and5.do
Figure 4	figure4.do
Text info on HS performance for high achieving students from high- and low-SES neighborhoods	hsgrowth_byes.do
Footnote 5	figure1.do
Footnote 11 and attached sentence	exitentry.do
Footnote 12	appendix_figure2.do
Footnote 13	tables1and2and3.do
Footnote 14	porder_selection.do
Footnote 15	appendix_table4.do
Text on firststage	firststage.do
Footnote 16	footnote16.do
Footnote 18	selectivecollege.do
Footnote 21	figure4.do
Footnote 22	appendix_table5.do
Appendix Table 4	appendix_table4.do
Appendix Table 5	appendix_table5.do
Appendix Figure 1	tiermaps.R
Appendix Figure 2	appendix_figure2.do
Appendix Figures 3A – 3C	rdplot_pretreat_all.do
Appendix Figures 4A – 4E	rdplot_tables4and5.do

