Data and Do-files for "Buy the Book? Evidence on the Effect of Textbook Funding on School-level Achievement"

1 Data

This folder contains five final data sets in STATA format which used to produce all tables and figures:

- CA_districts

- CA_schools_es

- CA_schools_ms

- CA_schools_hs

- williams_full_apportionment_schedule

1.1 Variable Descriptions

Variables in CA_districts:

Variable name	Description
year	School year ending with X
distid	Unique district id
num_qual	Number of schools in the district that receive Williams funding
ind_no_qual	Indicator equal to 1 if num_qual is equal to 0.
total	Number of students in the district
value2100	Textbook spending
value4100	Equipment spending
value5100	Instructional Aid spending
value5200	Service spending
value5500	Operations and housekeeping spending
value6400	Travel and Conference spending

Variables in CA_schools_es:

Variable name	Description
cds	Unique school id
year	School year ending with X
cstcapapercentageX_Y	Percent of students who meet performance category X in subject Y
readingscore	School-average mean scale score for reading tests
mathscore	School-average mean scale score for math tests
meanscaledscore_GX_TY	Mean scale score for grade X and test Y (reading = 7, math = 8)
stype	Type of school, Elementary, Middle, or High school
sname	School Name
api_rank	Academic performance rank, determines eligibility for Williams
yrs_teach	School average of teacher experience
yrs_dist	School average of teacher in-district experience
fte_t	School average of full-time equivalence for teachers
fte_a	School average of full-time equivalence for admin
fte_p	School average of full-time equivalence for pupil service staff
frl	Number of students eligible for free or reduced price lunch
amind	Number of American Indian students
asian	Number of Asian students
pac	Number of Pacific Islander students
fil	Number of Filipino students
hisp	Number of Hispanic students
afam	Number of African American students
wht	Number of white students
mnr	Number of students w/ multiple or no response
pct_ai	Percent of American Indian students
pct_as	Percent of Asian students
pct_pi	Percent of Pacific Islander students
pct_fi	Percent of Filipino students
pct_hi	Percent of Hispanic students
pct_aa	Percent of African American students
pct_wh	Percent of white students
pctmnr	Percent of students w/ multiple or no response
total	Number of students within the school
classsize	Average class size

Variables in CA_schools_ms and CA_schools_hs:

Variable name	Description
cds	Unique school id
year	School year ending with X
readingscore	School-average mean scale score for reading tests
mathscore	School-average mean scale score for math tests
stype	Type of school, Elementary, Middle, or High school
sname	School Name
api_rank	Academic performance rank, determines eligibility for Williams

Variables in williams_full_apportionment_schedule:

Variable name	Description
cds	Unique school id
lea	District name
sname	School Name
enroll_list	Enrollment for Williams funding calculation
entitle	Total school entitlement
prior_entitle	Previously paid entitlement
due	Amount due at time of posting

1.2   Intermediate Construction

CA_districts is constructed from intermediate files "200X-200Y SACS Unaudited Actual Data" available at "www.cde.ca.gov/ds/fd/fd". This project uses 2003 through 2009 SACS files. For 2001 to 2003, California used a different reporting system called "J-200 Unaudited Actual Data", which is also available at "www.cde.ca.gov/ds/fd/fd". The variable "object" denotes what the funding is used for (see "www.cde.ca.gov/fg/ac/ac" for a complete discussion of SACS code structure and code definitions) and "value" denotes the dollar amount. The "object" codes used in this project are textbook spending (4100), Equipment (6400), Instructional Aids (2100), Services (5100), Operations and housekeeping (5500), and Travel and Conferences (5200). CA_districts also uses enrollment data available at "www.cde.ca.gov/ds/sd/sd/filesenr.asp" collapsed to the district level.

CA_schools_es is constructed from intermediate files for Standardized Testing and Reporting (STAR) results, Academic Performance Index (API) data, enrollment data, staff demographic data, and free and reduced lunch eligibility data. STAR data is available at "www.star.cde.ca.gov". STAR tests used in this paper are based on CST English-Language Arts (Test ID 7) and CST Mathematics (Test ID 8) for subgroup "all students". Variables "mathscores", "readingscores", and all performance categories are unweighted averages across grades within schools. Variables "meanscaledscore_GX_TY" are the CDE reported averages for grade X on test Y. API data is available at "www.cde.ca.gov/ta/ac/ap/apidatafiles.asp". This uses file "2003 Base API - Data File". The variable api03 determines assignment of Williams funding in 2005; this is renamed "api_rank".

CA_schools_ms and CA_schools_hs are similarly constructed as CA_schools_es, but use

different STAR tests for math because most students take different STAR tests for different grades. CA_schools_ms uses an average of CST Mathematics (Test ID 8) and CST Algebra 1 (Test ID 9). CA_schools_hs uses an average of CST Mathematics, CST Algebra 1, CST Integrated Math 1, CST Geometry, CST Integrated Math 2, CST Algebra 2, CST Summative High School Mathematics, and CST General Mathematics (Test ID's 8-13, 15, and 28).

williams_full_apportionment_schedule is not constructed, but was posted on CDE's website. The PDF is available on request from CDE or myself.

Datasets are merged together using unique school identifiers county code, district code, and school code. Some datasets concatenate these codes as county code+district code+school code, where leading zeros make all codes the same string length. This is labeled "CDS" in the completed data sets. All district-level records are dropped for school-level data files. Alternative schools and CDE "small" schools are dropped. Schools that are not observed each year between 2002 and 2009 are also dropped.

2   Do Files

This folder contains four STATA do files:

- results_district_spending

- results_elementaryschools

- results_middleschools

- results_highschools

Each requires the correct file path to the relevant data set on line 3; the file results_elementaryschools also requires the correct file path to williams_full_apportionment_schedule for Figure 3, panel A.

The file results_district_spending produces Figure 3, Panel B and Figure C.1 for Appendix C.

The file results_elementaryschools produces regressions for Tables 1 through 8 and Figures 1 through 8. Additionally, this file produces results for tables and figures in Appendices A and B.

The file results_middleschools produces regressions for Table 9, panel A.

The file results_highschools produces regressions for Table 9, panel B.