

Elementary and Secondary School Test Score Datasets

CST

Contains student-level data on the California Standards Test. The CST is taken in grades 2-11. The CST was discontinued following the 2012-2013 school year.

Raw Files

The raw CST files contain a single year of data and are named according to the year in which the spring semester occurs.

Sapper Location: `/secure/ca_ed_lab/data/restricted_access/raw/cst`

Naming Convention: `star"year of spring semester"_ucd.dta`

Years: Spring 2003 (2002-2003 School Year) - Spring 2013 (2012-2013 School Year)

Clean File

The `cst_clean.dta` file contains all of the cleaned individual years of the CST data appended together.

Sapper Location: `/secure/ca_ed_lab/data/restricted_access/clean/cst/cst_clean.dta`

Do File Location: `/secure/ca_ed_lab/data/do_files/clean_cst.do`

SBAC

Contains student-level data on the Smarter Balanced Assessment Consortium. The SBAC is taken in grades 3-8 and 11. The SBAC was first implemented in the 2014-2015 school year.

Raw Files

The raw SBAC files contain a single year of data and are named according to the year in which the spring semester occurs.

Sapper Location: `/secure/ca_ed_lab/data/restricted_access/raw/sbac`

Naming Convention: `ucd_caaspp"year of spring semester".dta`

Years: Spring 2015 (2014-2015 School Year) - Spring 2016 (2015-2016 School Year)

Clean File

The `sbac_clean.dta` file contains all of the cleaned individual years of the SBAC data appended together.

Sapper Location: `/secure/ca_ed_lab/data/restricted_access/clean/sbac/sbac_clean.dta`

Do File Location: `/secure/ca_ed_lab/data/do_files/clean_sbac.do`

K-12 Test Scores

Combines data from the CST and SBAC datasets. The dataset contains all of the observations from the `cst_clean.dta` and `sbac_clean.dta` datasets but further divides the dataset into three subsamples: the *all students sample*, *all scores sample*, and *first scores sample*.

These subsamples are not mutually exclusive and, with the exception of minor differences due to randomly selected duplicates, *all students sample* \supset *all scores sample* \supset *first scores sample*.

This dataset has precalculated lagged test scores (by grade using the *first scores sample*) that begin with the prefix `L#_` (where # denotes the number of lagged grades) and peer effects (by school and grade using the *all students sample* for demographics and the *all scores sample* for test scores) that begin with the prefix `peer_`. This saves time when using these variables, as you avoid the computation time of creating the variables each time you run a do file.

The downside is that the large number of variables combined with the large number of observations creates a very large final dataset. For this reason, I would recommend **never loading the entire dataset into memory** and instead loading only the variables and subsample needed from the dataset using the command

```
use varlist [if] using k12_test_scores_clean.dta, clear
```

Clean File

The `k12_test_scores_clean.dta` file contains all of the cleaned individual years of the CST and SBAC data appended together.

Sapper Location:

```
/secure/ca_ed_lab/data/restricted_access/clean/k12_test_scores/k12_test_scores_clean.dta
```

Do File Location: `/secure/ca_ed_lab/data/do_files/clean_k12_test_scores.do`

CST Sample

This subsample contains every observation in the `cst_clean.dta` dataset.

Identifying Variable: `dataset=="CST"`

SBAC Sample

This subsample contains every observation in the `sbac_clean.dta` dataset.

Identifying Variable: `dataset=="SBAC"`

All Students Sample

This subsample keeps **one observation per student per grade per year**. If there are multiple observations per student per grade per year then the observation with the most demographic information is kept. **This subsample should be used when calculating demographic summary statistics.**

Identifying Variable: `all_students_sample==1`

All Scores Sample

This subsample keeps **one test score observation per student per grade per year**.

From the CST dataset the following observations are dropped:

- CAPA, STS, and CMA scores
- If the include indicator is not yes
- If the student is labeled as not tested by parent/guardian request

From the SBAC dataset the following observations are dropped:

- If the include indicator is not yes

If a student has multiple score observations per grade per year then observations are dropped in the following order. First, if an observation is missing both ELA and math test scores. If there are still duplicates, then an observation is dropped if an observation is missing either the ELA or math test score. If there are still duplicates, then an observation is randomly selected using a random number generator.

This subsample is used to create z-scores. z-scores are standardized at the grade by year level for ELA and grade by year by subject level for math. z-scores are calculated prior to dropping to one observation per student per year to account for the fact that a student may have multiple observations that contain a valid score for only a single subject.

Identifying Variable: `all_scores_sample==1`

First Scores Sample

This subsample keeps the **first test score observation per student per grade**. From the CST dataset the following observations are dropped:

- CAPA, STS, and CMA scores
- If the include indicator is not yes
- If the student is labeled as not tested by parent/guardian request

From the SBAC dataset the following observations are dropped:

- If the include indicator is not yes

If a student has multiple score observations per grade then observations are dropped in the following order. First, if an observation is missing both ELA and math test scores. If there are still duplicates, then an observation is dropped if an observation is missing either the ELA or math test score. If there are still duplicates, then the observation from the earliest year is selected.

Identifying Variable: `first_scores_sample==1`