

STRATEGIC **DATA** PROJECT

# SDP TOOLKIT

## FOR EFFECTIVE DATA USE

---

### A GUIDE FOR CONDUCTING DATA ANALYSIS IN EDUCATION AGENCIES

Please mute your computer speakers and phone microphone.



2. Clean

[www.gse.harvard.edu/sdp/tools](http://www.gse.harvard.edu/sdp/tools)



**STRATEGIC DATA PROJECT**

# **MISSION**

*Transform the use of data in education to improve student achievement.*



**Patty Diaz**

Senior Program Manager, Fellows

**Todd Kawakita**

Manager of Product Development



**Aaron Dow**

Research Analyst



1

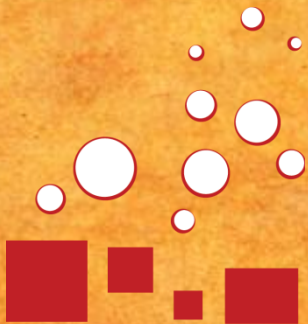
# SDP TOOLKIT

## FOR EFFECTIVE DATA USE

---

A GUIDE FOR CONDUCTING DATA ANALYSIS IN EDUCATION AGENCIES

2



### 2. Clean

1. Review structure of tasks
2. Walk you through one of the five tasks

3

# Q & A

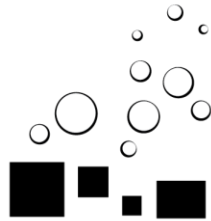
# SDP TOOLKIT

## FOR EFFECTIVE DATA USE

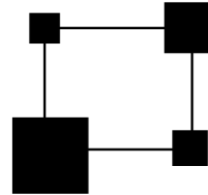
A GUIDE FOR CONDUCTING DATA  
ANALYSIS IN EDUCATION AGENCIES



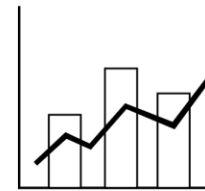
**1. Identify**  
essential data  
elements



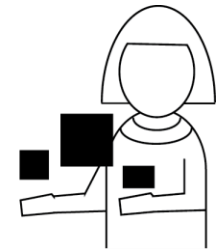
**2. Clean,**  
check, and  
build variables  
for your  
datasets



**3. Connect**  
relevant  
datasets from  
different  
sources



**4. Analyze**  
your datasets

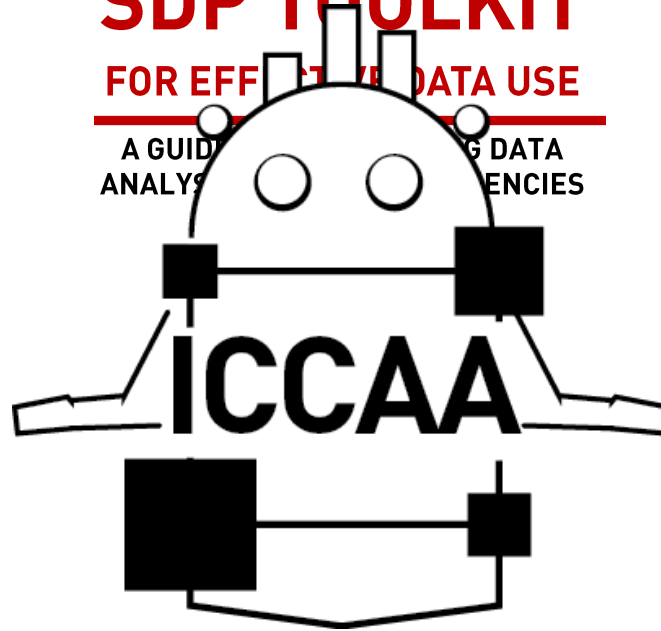


**5. Adopt**  
best practices  
to facilitate  
shared and  
replicable data  
analysis

# SDP TOOLKIT

FOR EFFECTIVE DATA USE

A GUIDE TO MANAGING DATA ANALYSES AND DATA AGENCIES

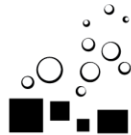


I



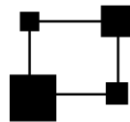
**Identify:** Data Specification Guide

C



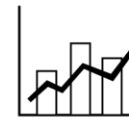
**Clean:** Data Building Tasks

C



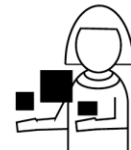
**Connect:** Data Linking Guide

A



**Analyze:** Diagnostic Analyses Guide

A

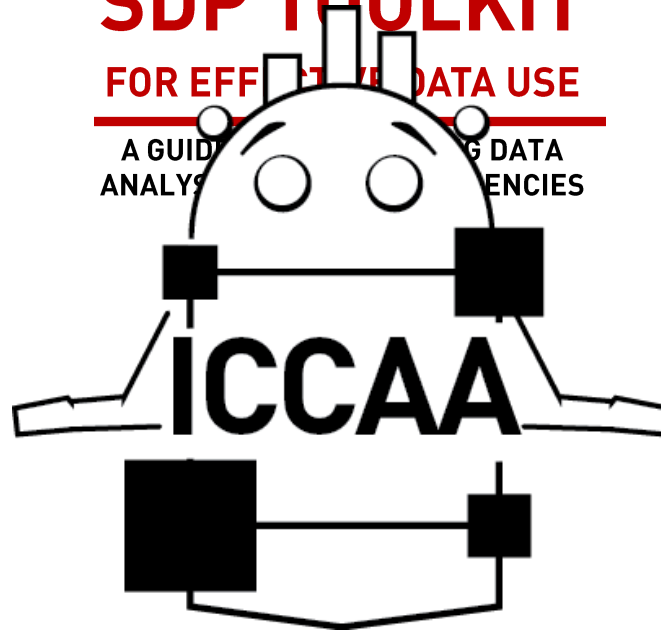


**Adopt:** Coding Style Guide

# SDP TOOLKIT

FOR EFFECTIVE DATA USE

A GUIDE TO MANAGING DATA ANALYSIS AGENCIES



I



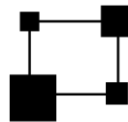
**Identify:** Data Specification Guide

C



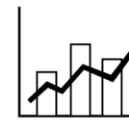
**Clean:** Data Building Tasks

C



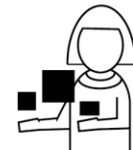
**Connect:** Data Linking Guide

A



**Analyze:** Diagnostic Analyses Guide

A

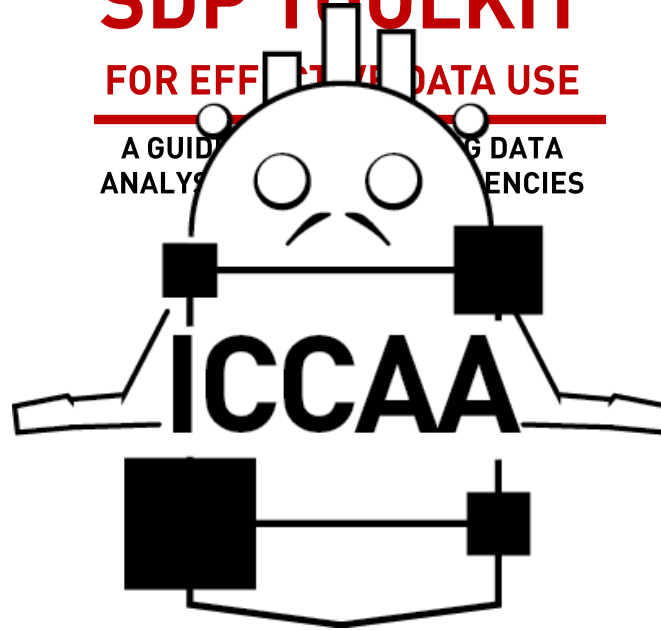


**Adopt:** Coding Style Guide

# SDP TOOLKIT

FOR EFFECTIVE DATA USE

A GUIDE TO MANAGING DATA ANALYSIS AGENCIES



I



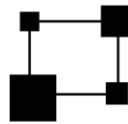
**Identify:** Data Specification Guide

C



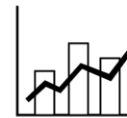
**Clean:** Data Building Tasks

C



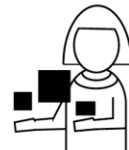
**Connect:** Data Linking Guide

A



**Analyze:** Diagnostic Analyses Guide

A



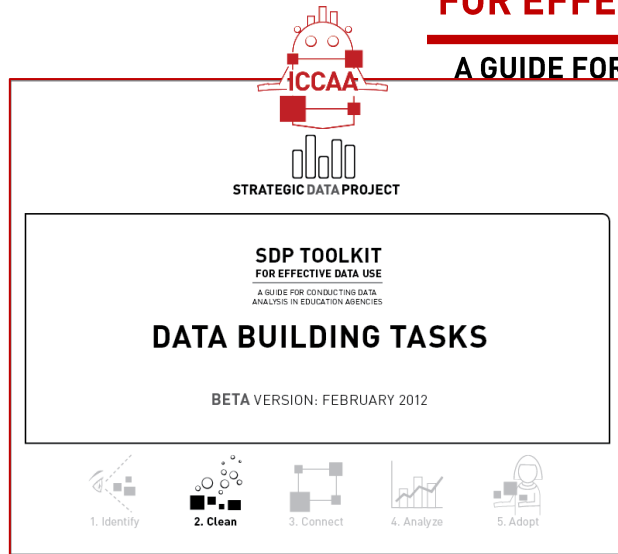
**Adopt:** Coding Style Guide



# SDP TOOLKIT

## FOR EFFECTIVE DATA USE

A GUIDE FOR CONDUCTING DATA ANALYSIS IN EDUCATION AGENCIES



I



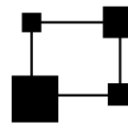
**Identify:** Data Specification Guide

C



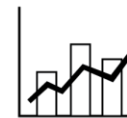
**Clean:** Data Building Tasks

C



**Connect:** Data Linking Guide

A

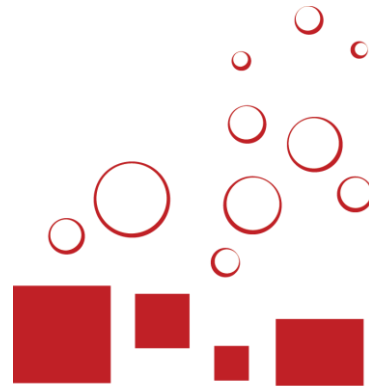


**Analyze:** Diagnostic Analyses Guide

A



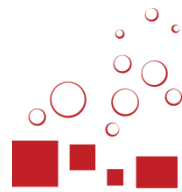
**Adopt:** Coding Style Guide



## 2. Clean

### Data Building Tasks

Upon collecting essential data elements, ensure that the data can be reliably used in future analyses.

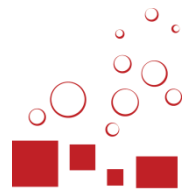


## 2. Clean

Data Building Tasks

### Research file STUDENT ATTRIBUTES




- Time-invariant variables such as **gender**, **race/ethnicity**
- Unique by **student**

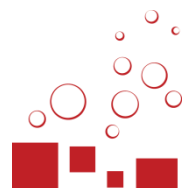


## 2. Clean Data Building Tasks

### Research file STUDENT ATTRIBUTES

- Time-invariant variables such as **gender**, **race/ethnicity**
- Unique by **student**

student	school_year	race_ethnicity
TK	2004	
TK	2005	
TK	2006	





## 2. Clean


Data Building Tasks

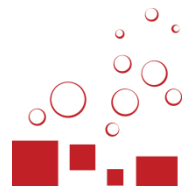
### Research file STUDENT ATTRIBUTES

- Time-invariant variables such as **gender**, **race/ethnicity**
- Unique by **student**

student	school_year	race_ethnicity
TK	2004	
TK	2005	
TK	2006	

Decision Rule?





## 2. Clean

Data Building Tasks

# A Set of **Five** Tasks

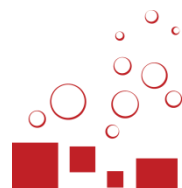
1. STUDENT  
ATTRIBUTES

2. STUDENT  
SCHOOL YEAR

3. IDENTIFYING  
THE NINTH-GRADE  
COHORT

4. STUDENT  
SCHOOL  
ENROLLMENT

5. PRIOR  
ACHIEVEMENT



**2. Clean**  
Data Building Tasks

# A Set of **Five** Tasks

**1. STUDENT  
ATTRIBUTES**

**2. STUDENT  
SCHOOL YEAR**

**3. IDENTIFYING  
THE NINTH-GRADE  
COHORT**

**4. STUDENT  
SCHOOL  
ENROLLMENT**

**5. PRIOR  
ACHIEVEMENT**

**2. Clean: SDP Data Building Tasks**

Practice Files: [Excel](#) [Stata](#)

Upon collecting essential data elements you need to ensure that the data can be reliably used in future analyses.

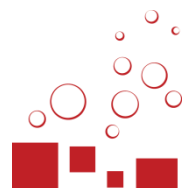
→ *Practice file*  
(input dataset)

→ *Practice output file*  
*clean + ready for*  
*analysis*

- Data explored
- Data cleaned
- Key variables defined
- Solutions checked

***Your own agency's file*** →

***Your own agency's***  
***output file***  
***clean + ready for analysis***



## 2. Clean Data Building Tasks

# A Set of **Five** Tasks

1. STUDENT  
ATTRIBUTES

2. STUDENT  
SCHOOL YEAR

3. IDENTIFYING  
THE NINTH-GRADE  
COHORT

4. STUDENT  
SCHOOL  
ENROLLMENT

5. PRIOR  
ACHIEVEMENT

### 2. Clean: **SDP Data Building Tasks**

Practice Files: [Excel](#) [Stata](#)

Upon collecting essential data elements you need to ensure that the data can be reliably used in future analyses.

→ Practice file  
(input dataset)

→ Practice output file  
ready for analysis

→ Your own agency's file

→ Your own agency's  
output file  
ready for analysis

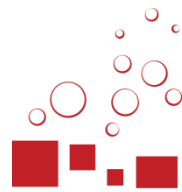
- Data explored
- Data cleaned
- Key variables defined
- Solutions checked



Shout out to:

**Nicole Wagner** SDP Data Fellow in LAUSD





## 2. Clean

Data Building Tasks

# A Set of **Five** Tasks

1. STUDENT  
ATTRIBUTES

2. STUDENT  
SCHOOL YEAR

3. IDENTIFYING  
THE NINTH-GRADE  
COHORT

4. STUDENT  
SCHOOL  
ENROLLMENT

5. PRIOR  
ACHIEVEMENT

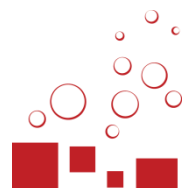
*Purpose* – What is the goal / output? What are the core assignments?

*How to Start* – What is in the input / practice file?

*Data Description*– What are the data elements in the practice file and what is the desired uniqueness?

*Instructions* – What steps should I take to produce the output (logical steps, code, and data snapshot exercises)?

*Solutions* – Did I do the exercises correctly?



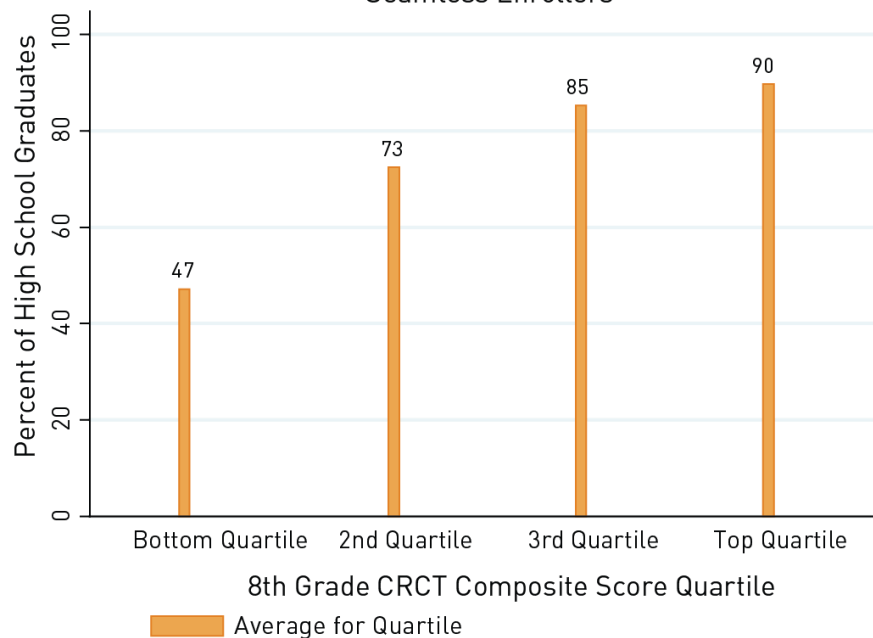
## 2. Clean Data Building Tasks



College-Going  
Success

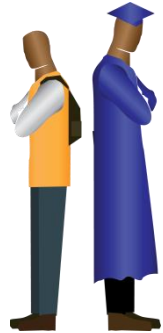
What is the relationship between 8<sup>th</sup> grade test scores and college enrollment rates?

Distribution of College Enrollment Rates  
by Prior Student Achievement  
Seamless Enrollers





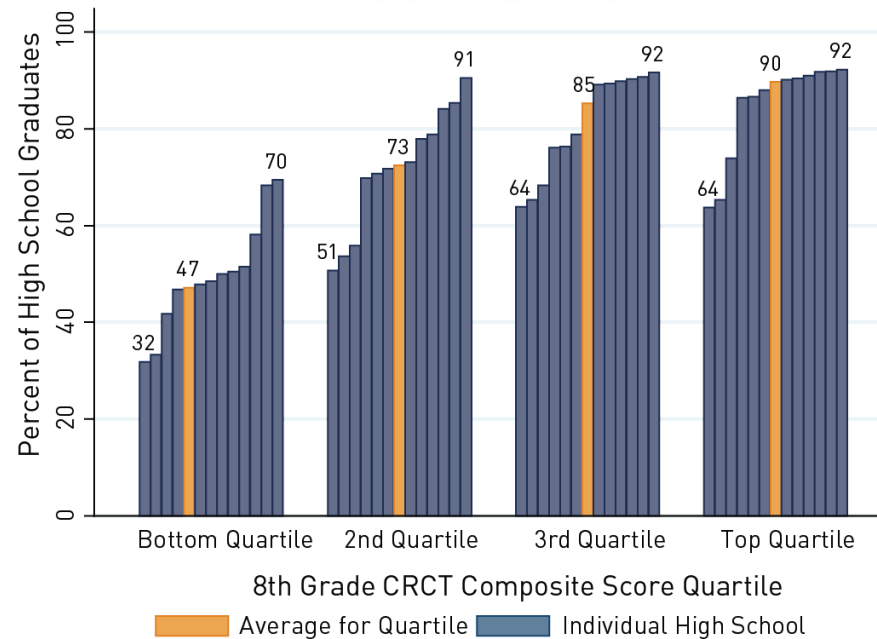
## 2. Clean Data Building Tasks

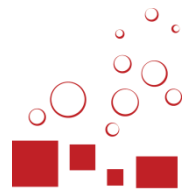


College-Going  
Success

# What is the relationship between 8<sup>th</sup> grade test scores and college enrollment rates?

### Distribution of College Enrollment Rates by Prior Student Achievement Seamless Enrollers





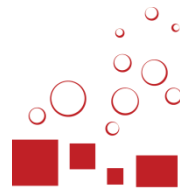
## 2. Clean

Data Building Tasks

# Prior Achievement

**Purpose** Generate a cleaned output file containing students' achievement on state standardized Math and English language arts tests in grade 8.

1. Clean raw and scaled eighth grade test scores and resolve instances in which a student is observed to have taken the same test multiple times.
2. Standardize test scores to have a mean of 0 and a standard deviation of 1.
3. Generate a composite math and English language arts test score



## 2. Clean

Data Building Tasks

# Prior Achievement

## How to Start

To begin, open the provided Student\_Test\_Scores practice file in Stata.

This practice file contains data for assessments administered during school years 2000-01 through 2006-07 for all grades. The file is unique by test taking instance—that is, it contains one observation for each time a student takes a given test.



## 2. Clean Data Building Tasks

# Data Description

Field Name	Values or Data Type	Definition
<b>sid</b>	numeric	Student identifier unique to each student. This identification number is typically assigned to a student upon enrollment in your agency. State agencies may have different identification numbers than district agencies for the same student.
<b>test_subject</b>	1 = math 2 = English language 3 = science 4 = social studies 5 = other	The general subject matter of the course.
<b>school_year</b>	date format (yyyy)	The exact date ( <b>in this case the minimum of school year</b> ) on which the test was completed. Note that students who re-take tests or are retained in grade may have multiple observations for a single test_subject.
<b>grade_level</b>	-9 = ungraded -1 = any pre-kindergarten 0 = kindergarten 1-12 = grades 1-12 13+ = additional grade levels (i.e. vocational training, special education past year 12)	The numeric grade level of the test.
<b>raw_score</b>	numeric	The student's raw score if available.
<b>scaled_score</b>	numeric	The student's scaled score.



## 2. Clean

Data Building Tasks

**NOT** Unique By:  
**sid + test\_subject + school\_year**

Unique By:  
**sid + test\_subject + school\_year**

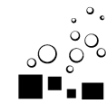
Unique By:  
**sid**



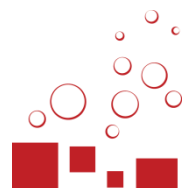
RAW  
research file  
from **Identify**



**Instructions**



CLEAN  
research file  
ready for **Connect**



## 2. Clean Data Building Tasks

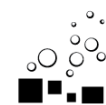
**NOT** Unique By:  
**sid + test\_subject + school\_year**

Unique By:  
**sid + test\_subject + school\_year**

Unique By:  
**sid**



**RAW**  
research file  
from **Identify**



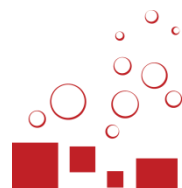
**CLEAN**  
research file  
ready for **Connect**

# Instructions

## 1. Examine your **Student\_Test\_Scores** raw research file input dataset.

sid	school_year	grade_level	test_subject	raw_score	scaled_score
3	2004	8	1	348	258
3	2004	8	2	208	185
3	2005	9	1	264	220
4	2001	8	1	480	320
4	2002	8	1	396	278
4	2002	8	1	.	371





## 2. Clean Data Building Tasks

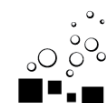
**NOT** Unique By:  
**sid + test\_subject + school\_year**

Unique By:  
**sid + test\_subject + school\_year**

Unique By:  
**sid**



**RAW**  
research file  
from **Identify**

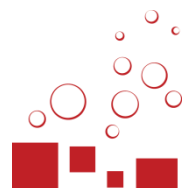


**CLEAN**  
research file  
ready for **Connect**

# Instructions

## 1. Examine your **Student\_Test\_Scores** raw research file input dataset.

sid	school_year	grade_level	test_subject	raw_score	scaled_score
3	2004	8	1	348	258
3	2004	8	2	208	185
3	2005	9	1	264	220
4	2001	8	1	480	320
4	2002	8	1	396	278
4	2002	8	1	.	371



## 2. Clean Data Building Tasks

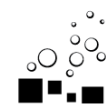
**NOT** Unique By:  
**sid + test\_subject + school\_year**

Unique By:  
**sid + test\_subject + school\_year**

Unique By:  
**sid**



**RAW**  
research file  
from **Identify**

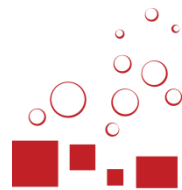


**CLEAN**  
research file  
ready for **Connect**

# Instructions

## 1. Examine your **Student\_Test\_Scores** raw research file input dataset.

sid	school_year	grade_level	test_subject	raw_score	scaled_score
3	2004	8	1	348	258
3	2004	8	2	208	185
3	2005	9	1	264	220
4	2001	8	1	480	320
4	2002	8	1	396	278
4	2002	8	1	.	371

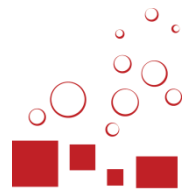


# Instructions

## 2. Clean Data Building Tasks

**2. Narrow down the test score data file to include only the relevant tests and grades.**

sid	school_year	grade_level	test_subject	raw_score	scaled_score
3	2004	8	1	348	258
3	2004	8	2	208	185
3	2005	9	1	264	220
4	2001	8	1	480	320
4	2002	8	1	396	278
4	2002	8	1	.	371

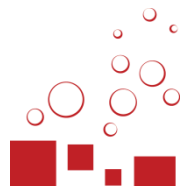


# Instructions

## 2. Clean Data Building Tasks

**2. Narrow down the test score data file to include only the relevant tests and grades.**

sid	school_year	grade_level	test_subject	raw_score	scaled_score
3	2004	8	1	348	258
3	2004	8	2	208	185
3	2005	7	1	264	220
4	2001	8	1	480	320
4	2002	8	1	396	278
4	2002	8	1	.	371

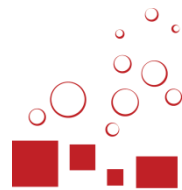


## 2. Clean Data Building Tasks

# Instructions

### 3. Clean up raw and scaled test scores.

sid	school_year	grade_level	test_subject	raw_score	scaled_score
3	2004	8	1	348	258
3	2004	8	2	208	185
3	2005	7	1	264	220
4	2001	8	1	480	320
4	2002	8	1	396	278
4	2002	8	1	.	371



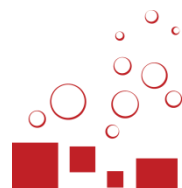
## 2. Clean Data Building Tasks

# Instructions

### 3. Clean up raw and scaled test scores.

sid	school_year	grade_level	test_subject	raw_score	scaled_score
3	2004	8	1	348	258
3	2004	8	2	208	185
3	2005	7	1	264	220
4	2001	8	1	480	320
4	2002	8	1	396	278
4	2002	8	1	.	371

No changes based on  
decision rule



## 2. Clean Data Building Tasks

**NOT** Unique By:  
**sid + test\_subject + school\_year**

Unique By:  
**sid + test\_subject + school\_year**

Unique By:  
**sid**



RAW  
research file  
from **Identify**

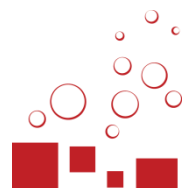


CLEAN  
research file  
ready for **Connect**

## Instructions

### 4. Identify same-year repeat test takers.

sid	school_year	grade_level	test_subject	raw_score	scaled_score
3	2004	8	1	348	258
3	2004	8	2	208	185
3	2005	7	1	264	220
4	2001	8	1	480	320
4	2002	8	1	396	278
4	2002	8	1	.	371



## 2. Clean Data Building Tasks

**NOT** Unique By:  
**sid + test\_subject + school\_year**

Unique By:  
**sid + test\_subject + school\_year**

Unique By:  
**sid**



RAW  
research file  
from **Identify**



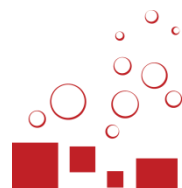
CLEAN  
research file  
ready for **Connect**

## Instructions

### 4. Identify same-year repeat test takers.

sid	school_year	grade_level	test_subject	raw_score	scaled_score
3	2004	8	1	348	258
3	2004	8	2	208	185
3	2005	7	1	264	220
4	2001	8	1	480	320
4	2002	8	1	378	278
4	2002	8	1	.	371





## 2. Clean Data Building Tasks

**NOT** Unique By:  
**sid + test\_subject + school\_year**

Unique By:  
**sid + test\_subject + school\_year**

Unique By:  
**sid**



RAW  
research file  
from **Identify**



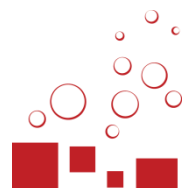
## Instructions



CLEAN  
research file  
ready for **Connect**

### 4. Identify same-year repeat test takers.

sid	school_year	grade_level	test_subject	raw_score	scaled_score
3	2004	8	1	348	258
3	2004	8	2	208	185
4	2001	8	1	480	320
4	2002	8	1	.	371



## 2. Clean

Data Building Tasks

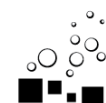
**NOT** Unique By:  
**sid + test\_subject + school\_year**

Unique By:  
**sid + test\_subject + school\_year**

Unique By:  
**sid**



RAW  
research file  
from **Identify**

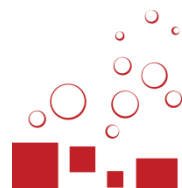


CLEAN  
research file  
ready for **Connect**

## Instructions

### 5. Reshape your data.

sid	school_year	grade_level	test_subject	raw_score	scaled_score
3	2004	8	1	348	258
3	2004	8	2	208	185
4	2001	8	1	480	320
4	2002	8	1	.	371



## 2. Clean Data Building Tasks

**NOT** Unique By:  
**sid + test\_subject + school\_year**

Unique By:  
**sid + test\_subject + school\_year**

Unique By:  
**sid**



RAW  
research file  
from **Identify**



## Instructions



CLEAN  
research file  
ready for **Connect**

### 5. Reshape your data.

sid	school_year	grade_level	math_raw_score	math_scaled_score	ela_raw_score	ela_scaled_score
3	2004	8	348	258	208	185
4	2001	8	480	320	-	-
4	2002	8	-	371	-	-



## 2. Clean Data Building Tasks

**NOT** Unique By:  
**sid + test\_subject + school\_year**

Unique By:  
**sid + test\_subject + school\_year**

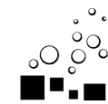
Unique By:  
**sid**



**RAW**  
research file  
from **Identify**



## Instructions

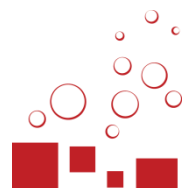


**CLEAN**  
research file  
ready for **Connect**

### 6. Create standardized test scores.

sid	school_year	grade_level	math_raw_score	math_scaled_score	ela_raw_score	ela_scaled_score	s
3	2004	8	348	258	208	185	
4	2001	8	480	320	.	.	
4	2002	8	.	371	.	.	

Let's scooch the data over to add the columns for standardized test scores.



## 2. Clean Data Building Tasks

**NOT** Unique By:  
**sid + test\_subject + school\_year**

Unique By:  
**sid + test\_subject + school\_year**

Unique By:  
**sid**



RAW  
research file  
from **Identify**



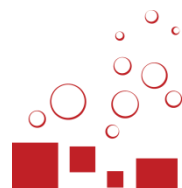
## Instructions



CLEAN  
research file  
ready for **Connect**

### 6. Create standardized test scores.

el	math_raw_score	math_scaled_score	ela_raw_score	ela_scaled_score	std_scaled_math	std_scaled_ela
	348	258	208	185	.36	-1.2
	480	320	.	.	1.6	.
	.	371	.	.	3.20	.



## 2. Clean Data Building Tasks

**NOT** Unique By:  
**sid + test\_subject + school\_year**

Unique By:  
**sid + test\_subject + school\_year**

Unique By:  
**sid**



RAW  
research file  
from **Identify**



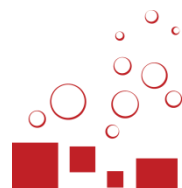
## Instructions



CLEAN  
research file  
ready for **Connect**

### 7. Identify different year repeat test takers.

sid	school_year	grade_level	math_raw_score	math_scaled_score	ela_raw_score	ela_scaled_score
3	2004	8	348	258	208	185
4	2001	8	480	320	.	.
4	2002	8	.	371	.	.



## 2. Clean Data Building Tasks

**NOT** Unique By:  
**sid + test\_subject + school\_year**

Unique By:  
**sid + test\_subject + school\_year**

Unique By:  
**sid**



RAW  
research file  
from **Identify**



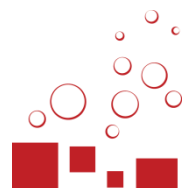
## Instructions



CLEAN  
research file  
ready for **Connect**

### 7. Identify different year repeat test takers.

sid	school_year	grade_level	math_raw_score	math_scaled_score	ela_raw_score	ela_scaled_s
3	2004	8	348	258	208	185
4	2001	8	480	320	.	.
4	2002	8	.	371	.	.



## 2. Clean Data Building Tasks

**NOT** Unique By:  
**sid + test\_subject + school\_year**

Unique By:  
**sid + test\_subject + school\_year**

Unique By:  
**sid**



RAW  
research file  
from **Identify**



## Instructions



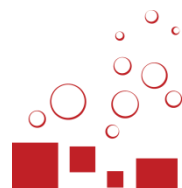
CLEAN  
research file  
ready for **Connect**

### 8. Lastly, generate a composite score.

sid	school_year	grade_level	math_raw_score	math_scaled_score	ela_raw_score	ela_scaled_score
3	2004	8	348	258	208	1
4	2002	8	.	371	.	

Let's scooch the data over to add the column for the composite score.





## 2. Clean Data Building Tasks

**NOT** Unique By:  
**sid + test\_subject + school\_year**

Unique By:  
**sid + test\_subject + school\_year**

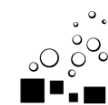
Unique By:  
**sid**



RAW  
research file  
from **Identify**



## Instructions



CLEAN  
research file  
ready for **Connect**

### 8. Lastly, generate a composite score.

caled_score	ela_raw_score	ela_scaled_score	std_scaled_math	std_scaled_ela	composite_scaled_score
258	208	185	.36	-1.2	-0.42
320	.	.	1.6	.	.

# APPENDIX



## 2. Clean: Data Building Tasks CG

Upon collecting essential data elements, ensure that the data can be reliably used in future analyses.

### STATA GLOSSARY

This and other tasks use code that is specific to the statistical program, Stata. If you are unfamiliar with or need to brush up on Stata, please refer to the following **STATA GLOSSARY** of common commands. Note that the glossary is nowhere near comprehensive but outlines many useful commands pertinent to the data cleaning and exploration process. Examples are provided where appropriate. Commands are NOT listed in alphabetical order.

**tab** — examines the distribution of values for a variable. This command is essential for checking data.

- A tab with one variable shows frequencies for each value of the variable.
- A tab with two variables (often called a cross-tab) produces a matrix of frequencies for the values of one variable against the values of another variable.

This is an important tool to use when examining relationships and the distributions of values between variables. The `, mi` option is often added to include the distribution of missing values in the tabulation.

Example: examining the gender distribution within the data set.

```
. tab gender,mi
```

gender	Freq.	Percent	Cum.
	1	20.00	20.00
female	2	40.00	60.00
male	2	40.00	100.00
Total	5	100.00	

Example: examining which students qualify for free lunch.

```
. tab sid free_lunch
```

sid	free_lunch		Total
	0	1	
1	0	1	1
2	0	1	1
4	0	1	1
5	1	0	1
Total	1	3	4

# APPENDIX

---



## 2. Clean: Data Building Tasks CG

Upon collecting essential data elements, ensure that the data can be reliably used in future analyses.

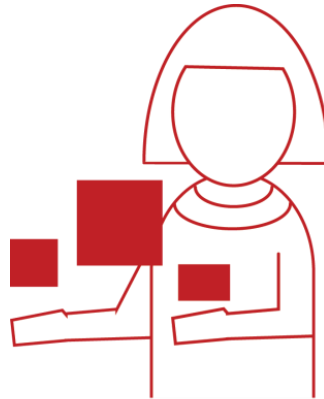
### DECISION RULES GLOSSARY

As noted in the **DATA DESCRIPTION** for each task, many of the tasks use a partial dataset to help you learn the core methodology for cleaning the raw research files. Some of these excluded variables, however, are essential for later analyses. To guide you through the process of cleaning these additional variables, we provide below a **DECISION RULES GLOSSARY** that provides further instruction on how to clean these variables.

Coupled with the specific instructions you've received from each task, we hope that you will be able to extrapolate your knowledge beyond what is covered directly in the tasks and clean these other variables as well.

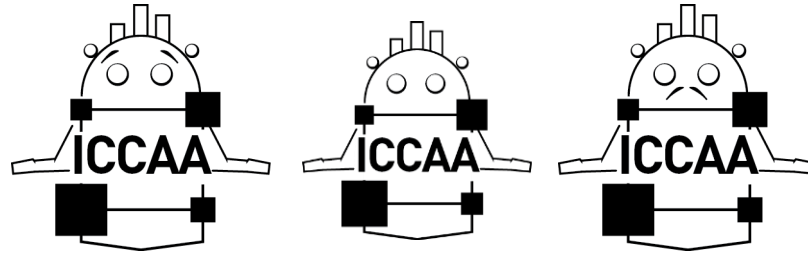
## STUDENT ATTRIBUTES

Data Element		Possible Scenario	SDP Decision Rule	Reference in Data Building Tasks
<b>sid</b>	numeric			
<b>male</b>	0 = female 1 = male	For the purposes of the diagnostic, SDP considers race_ethnicity time invariant. However, for some students more than one race/ethnicity may be observed.	see Task 1	
<b>race_ethnicity</b>	1 = African American 2 = Asian American 3 = Hispanic 4 = American Indian 5 = White, not Hispanic 6 = Other 7 = Multiple	For the purposes of the diagnostic, SDP considers race_ethnicity time invariant. However, for some students more than one race/ethnicity may be observed.	see Task 1	
<b>race</b>	1 = African American 2 = Asian American 3 = American Indian 4 = White 5 = Other 6 = Multiple		<i>in data spec, but not connect phase</i>	
<b>ethnicity</b>	0 = Hispanic 1 = not Hispanic		<i>in data spec, but not connect phase</i>	
<b>birth_date</b>	date format (yyyymm-dd)	For the purposes of the diagnostic, SDP considers birth_date time invariant. However, for some students more than one birth date may be observed.	In this case, report the modal birth date. If multiple modes, report the most recent birth date recorded. When evaluating modal birth date exclude birth dates the fall outside +/- four years of the expected birth date given grade-level and school year.	see Task 1, cleaning code for race_ethnicity variable, or Task 4, code for defining first and last high schools
<b>first_9th_school_year_reported</b>	spring calendar year	This is the school year during which the student was a 9th grader for the first time. Not all systems will record this information.	Report what the system explicitly recorded for first 9th grade school year. If the system does not record this information, leave the information omitted.	
<b>hs_diploma</b>	0 = no high school diploma 1 = has high school diploma			
<b>hs_diploma_type</b>	use local values	For the purposes of the diagnostic, SDP considers hs_diplomas_type to be time invariant. However, for some students more than one diploma type may be observed.	If more than one type is observed, report the type associated with the earliest diploma received. For example, Honors diploma, College Prep diploma, or General Education Diploma (GED) diploma.	
<b>zip_code</b>	xxxxx or xxxxx-yyyy		<i>in data spec, but not connect phase</i>	



## **5. Adopt** Coding Style Guide

To ensure that statistical code is easily shared across a team and is replicable by future users, SDP and the Center for Education Policy Research (CEPR) recommends that you follow best coding, programming, and data management practices.

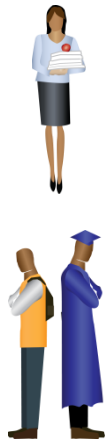


# Q & A

# SDP TOOLKIT

## FOR EFFECTIVE DATA USE

A GUIDE FOR CONDUCTING DATA ANALYSIS IN EDUCATION AGENCIES



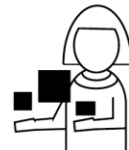
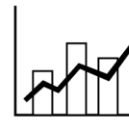
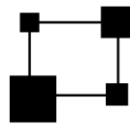
I

C

C

A

A



**Identify:** Data Specification Guide

**Clean:** Data Building Tasks

**Connect:** Data Linking Guide

**Analyze:** Diagnostic Analyses Guide

**Adopt:** Coding Style Guide

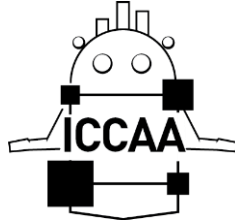
Will Be Released Prior to Webinar On:

Thursday, February 2

Thursday, February 9

Thursday, February 16

# Thank You



*The toolkit is currently in **BETA**.*

*Please send us your feedback at [goo.gl/AAvdF](http://goo.gl/AAvdF).*

*Check [www.gse.harvard.edu/sdp/tools](http://www.gse.harvard.edu/sdp/tools) for the most recent toolkit version.*

*Please contact us at [sdp@gse.harvard.edu](mailto:sdp@gse.harvard.edu) if you have any questions about the toolkit.*