

# Package ‘SGPdata’

May 7, 2026

**Type** Package

**Title** Exemplar Data Sets for Student Growth Percentiles (SGP) Analyses

**Version** 28.0-0.0

**Date** 2024-7-14

**Maintainer** Damian W. Betebenner <dbetebenner@nciea.org>

**Depends** R (>= 4.0)

**Suggests** SGP (>= 1.9-0.0), knitr, rmarkdown

**Imports** crayon, data.table (>= 1.14.0)

**Description** Data sets utilized by the 'SGP' package as exemplars for users to conduct their own student growth percentiles (SGP) analyses.

**URL** <https://CenterForAssessment.github.io/SGPdata/>,  
<https://github.com/CenterForAssessment/SGPdata/>,  
<https://cran.r-project.org/package=SGPdata>

**BugReports** <https://github.com/CenterForAssessment/SGPdata/issues/>

**VignetteBuilder** knitr

**LazyLoad** Yes

**LazyData** Yes

**LazyDataCompression** xz

**License** GPL-3

**NeedsCompilation** no

**Author** Damian W. Betebenner [aut, cre],  
Adam R. Van Iwaarden [aut],  
Ben Domingue [aut]

**Repository** CRAN

**Date/Publication** 2024-07-15 07:40:05 UTC

## Contents

SGPdata-package . . . . .	2
sgpData . . . . .	2
sgpData_INSTRUCTOR_NUMBER . . . . .	3
sgpData_LONG . . . . .	4
sgpData_LONG_COVID . . . . .	6
sgptData_LONG . . . . .	7

<b>Index</b>	<b>9</b>
--------------	----------

---

SGPdata-package	<i>Exemplar Data Sets for Student Growth Percentile (SGP) Analyses</i>
-----------------	--

---

### Description

The 'SGPdata' package contains five exemplar data sets utilized by the 'SGP' package for testing and examples.

### Details

Package:	SGPdata
Type:	Package
Version:	28.0-0.0
Date:	2024-7-14
License:	GPL-3
LazyLoad:	Yes
LazyData:	Yes

---

sgpData	<i>Longitudinal Student Assessment Data</i>
---------	---

---

### Description

A dataset comprising a subset of five years of state assessment reading data suitable for student growth percentile and percentile growth projection/trajectory analyses. The dataset is used in examples provided in the documentation for the `studentGrowthPercentiles` and `studentGrowthProjections` functions, and serves as an exemplar for user construction of their own datasets for growth percentile analyses.

**Format**

A data frame of student level observations for five years on the following variables.

ID Student ID  
GRADE\_2020 Student Grade Tested 2020, possibly missing  
GRADE\_2021 Student Grade Tested 2021, possibly missing  
GRADE\_2022 Student Grade Tested 2022, possibly missing  
GRADE\_2023 Student Grade Tested 2023, possibly missing  
GRADE\_2024 Student Grade Tested 2024, possibly missing  
SS\_2020 Student Scale Score 2020, possibly missing  
SS\_2021 Student Scale Score 2021, possibly missing  
SS\_2022 Student Scale Score 2022, possibly missing  
SS\_2023 Student Scale Score 2023, possibly missing  
SS\_2024 Student Scale Score 2024, possibly missing

**Note**

All datasets used with `studentGrowthPercentiles` and `studentGrowthProjections` must be specifically formatted as *wide* format files. The first variable/column is the student ID variable. The next set of columns provide the grade of the student across all the years provided in the dataset (possibly missing). The last set of columns provide the scales scores of the student in the respective grades. Multi-year operational analyses benefit from putting data in long format. See `sgpData_LONG` and associated documentation for a comprehensive account.

**Source**

Anonymized student level state assessment data

**See Also**

[sgpData\\_LONG](#)

---

sgpData\_INSTRUCTOR\_NUMBER

*Lookup table associating INSTRUCTOR\_NUMBER with student ID by YEAR and CONTENT\_AREA*

---

**Description**

A lookup table comprising student IDs associated with INSTRUCTOR\_NUMBER a subset of five years of state assessment reading and mathematics data suitable for student growth percentile and percentile growth projection/trajectory analyses. The dataset is in LONG format with each record representing a unique teacher by student by year by content area combination.

**Format**

A data frame of student level observations for five years and two content areas for the following variables.

ID Unique Student Identification Number

CONTENT\_AREA Content area for student/teacher record (Reading or Mathematics)

YEAR Year for student/teacher record

INSTRUCTOR\_NUMBER Unique instructor number identifier

INSTRUCTOR\_LAST\_NAME Instructor last name

INSTRUCTOR\_FIRST\_NAME Instructor first name

INSTRUCTOR\_WEIGHT Proportional weight associated with student/teacher record

INSTRUCTOR\_ENROLLMENT\_STATUS Indicator of full enrollment status of student with teacher

**Source**

Anonymized student/teacher lookup table in LONG format

**See Also**

[sgpData\\_LONG](#)

---

sgpData\_LONG

*Longitudinal Student Assessment Data in LONG format*

---

**Description**

A dataset comprising a subset of five years of state assessment reading and mathematics data suitable for student growth percentile (SGP) and percentile growth projection/trajectory analyses. The dataset is in LONG format with each record representing a unique student, by year by content area combination. The example data is used in examples and illustration in the enclosed documentation to show how SGPs can be operationalized across years with the results being used to produce a variety of high quality visualizations both at the aggregate (e.g., school) and individual level.

**Format**

A data frame of student level observations for five years and two content areas for the following variables.

ID Unique Student Identification Number

LAST\_NAME Student last name

FIRST\_NAME Student first name

CONTENT\_AREA Content area for student observation (Reading or Mathematics)

YEAR Year for student observation

GRADE Grade level of student observation

SCALE\_SCORE Student Scale Score  
ACHIEVEMENT\_LEVEL Achievement level associated with student scale score  
GENDER Student gender  
ETHNICITY Student ethnicity  
FREE\_REDUCED\_LUNCH\_STATUS Student free/reduced lunch status  
ELL\_STATUS Student English Language Learner status  
IEP\_STATUS Student Individual Education Plan status  
GIFTED\_AND\_TALENTED\_PROGRAM\_STATUS Student Gifted and Talented Program status  
SCHOOL\_NUMBER School number associated with student record  
SCHOOL\_NAME School name associated with school number and student record  
EMH\_LEVEL Elementary, Middle, High school indicator for school attended  
DISTRICT\_NUMBER District number associated with student record  
DISTRICT\_NAME District name associated with school number and student record  
SCHOOL\_ENROLLMENT\_STATUS Indicator of full academic year student enrollment in school  
DISTRICT\_ENROLLMENT\_STATUS Indicator of full academic year student enrollment in school  
STATE\_ENROLLMENT\_STATUS Indicator of full academic year student enrollment in school  
VALID\_CASE Indicator of whether the case is valid or invalid

**Note**

All datasets used with `studentGrowthPercentiles` and `studentGrowthProjections` must be specifically formatted as *wide* format files. The first variable/column is the student ID variable. The next set of columns provide the grade of the student across all the years provided in the dataset (possibly missing). The last set of columns provide the scales scores of the student in the respective grades. Operational users of this package are advised to use data formatted similar to the `sgpData_LONG` with the higher level functions `prepareSGP`, `analyzeSGP`, `combineSGP`, `summarizeSGP`, `visualizeSGP`, and `outputSGP`

**Source**

Anonymized student level state assessment data in LONG format

**See Also**

[sgpData](#)

---

sgpData_LONG_COVID	<i>Longitudinal Student Assessment Data in LONG format with COVID-19 2020 interruption</i>
--------------------	--

---

## Description

A dataset comprising a subset of seven years of state assessment ELA and mathematics data across an 8 year span suitable for student growth percentile (SGP) and percentile growth projection/trajectory analyses. The dataset is missing 2020 data to help users model COVID related interruptions to student growth. The dataset is in LONG format with each record representing a unique student, by year by content area combination. The example data is used in examples and illustration in the enclosed documentation to show how SGPs can be operationalized across years with the results being used to produce a variety of high quality visualizations both at the aggregate (e.g., school) and individual level.

## Format

A data frame of student level observations for seven years (across an 8 year span) and two content areas for the following variables.

VALID\_CASE Indicator of whether the case is valid or invalid

CONTENT\_AREA Content area for student observation (ELA or Mathematics)

YEAR Year for student observation

ID Unique Student Identification Number

IEP\_STATUS Student Individual Education Plan status

GRADE Grade level of student observation

SCALE\_SCORE Student Scale Score

SCALE\_SCORE\_without\_COVID\_IMPACT Student Scale Score without COVID impact. Original scale score

ACHIEVEMENT\_LEVEL Achievement level associated with student scale score

ETHNICITY Student ethnicity

FREE\_REDUCED\_LUNCH\_STATUS Student free/reduced lunch status

ELL\_STATUS Student English Language Learner status

GENDER Student gender

SCHOOL\_NUMBER School number associated with student record

SCHOOL\_NAME School name associated with school number and student record

DISTRICT\_NUMBER District number associated with student record

DISTRICT\_NAME District name associated with school number and student record

**Note**

All datasets used with `studentGrowthPercentiles` and `studentGrowthProjections` must be specifically formatted as *wide* format files. The first variable/column is the student ID variable. The next set of columns provide the grade of the student across all the years provided in the dataset (possibly missing). The last set of columns provide the scales scores of the student in the respective grades. Operational users of this package are advised to use data formatted similar to the `sgpData_LONG_COVID` with the higher level functions `prepareSGP`, `analyzeSGP`, `combineSGP`, `summarizeSGP`, `visualizeSGP`, and `outputSGP`

**Source**

Anonymized and altered student level state assessment data in LONG format

**See Also**

[sgpData\\_LONG](#)

---

sgptData_LONG	<i>Longitudinal Student Assessment Data in LONG format with DATE for SGPt analyses.</i>
---------------	---

---

**Description**

A dataset comprising a subset of 3 years (and time periods) of interim assessment early literacy, reading and mathematics data suitable for time dependent student growth percentile (SGPt) and percentile growth projection/trajectory analyses. The dataset is in LONG format with each record representing a unique student, by year/time period by content area combination. The example data is used in examples and illustrations in the enclosed documentation to show how SGPTs can be operationalized across years with the results being used to produce a variety of high quality visualizations both at the aggregate (e.g., school) and individual level.

**Format**

A data frame of student level observations for five years and two content areas for the following variables.

COUNTRY Variable indicating the COUNTRY associated with test record.

STATE State associated with the test record.

VALID\_CASE VALID\_CASE variable.

CONTENT\_AREA Variable indicating content area associated with test record.

YEAR Variable indicating the year associated with the test record.

DATE Variable indicating the year associated with the test record: Must be in YYYY-MM-DD format.

ID Unique student identifier associated with test record.

GRADE Variable indicating the grade associated with the test record.

SCALE\_SCORE Variable providing the scale score associated with the test record.

SCALE\_SCORE\_RASCH Variable providing the Rasch score associated with the test record.

SEM Variable indicating the standard error of measurement associated with the scale score.

ACHIEVEMENT\_LEVEL Variable indicating the achievement level associated with the scale score.

**Source**

Anonymized student level interim assessment data in LONG format

**See Also**

[sgpData\\_LONG](#)

# Index

## \* datasets

- sgpData, [2](#)
- sgpData\_INSTRUCTOR\_NUMBER, [3](#)
- sgpData\_LONG, [4](#)
- sgpData\_LONG\_COVID, [6](#)
- sgptData\_LONG, [7](#)

## \* package

- SGPdata-package, [2](#)

SGPdata (SGPdata-package), [2](#)

sgpData, [2](#), [5](#)

SGPdata-package, [2](#)

sgpData\_INSTRUCTOR\_NUMBER, [3](#)

sgpData\_LONG, [3](#), [4](#), [4](#), [7](#), [8](#)

sgpData\_LONG\_COVID, [6](#)

sgptData\_LONG, [7](#)