The Colorado Student Growth Model
A Technical Overview of the 2017-2018 Student Growth Percentile Calculations

Adam R. VanIwaarden
Damian W. Betebenner
National Center for the Improvement of Educational Assessment (NCIEA)

October 2018

Abstract

DRAFT REPORT - DO NOT CITE! This report provides details about the Colorado Student Growth Model methodology and presents a descriptive analysis of the 2018 SGP calculation process and results.

1 Introduction

This report contains details on the 2017-2018 implementation of the student growth percentiles (SGP) model for the state of Colorado. The National Center for the Improvement of Educational Assessment (NCIEA) contracted with the Colorado Department of Education (CDE) to implement the SGP methodology using data derived from the Colorado Measures of Academic Success (CMAS), as well as the PSAT9, PSAT10 and SAT assessments to create the Colorado Growth Model. The goal of the engagement with CDE is to create a set of open source analytics techniques and conduct analyses that will eventually be conducted exclusively by CDE in following years.

The SGP methodology is an open source norm- and criterion-referenced student growth analysis that produces student growth percentiles for each student in the state with adequate longitudinal data. The methodology is currently used for many purposes. States and districts have used the results in various ways including parent/student diagnostic reporting, institutional improvement, and school and educator accountability. Specifics about the manner in which growth is included in school and educator accountability can be found in documents related to those accountability systems.

This report includes four sections:

- ***Data -*** includes details on the decision rules used in the raw data preparation and student record validation.
- ***Analytics -*** introduces some of the basic statistical methods and the computational process implemented in the 2018 analyses.[1]
- ***Goodness of Fit -*** investigates how well the statistical models used to produce SGPs fit Colorado students' data. This includes discussion of goodness of fit plots and the student-level correlations between SGP and prior achievement.
- ***SGP Results -*** provides basic descriptive statistics from the 2018 analyses at both the state and school levels.

This report also includes multiple appendices. Appendix A displays Goodness of Fit plots for each analysis conducted in 2018. Appendix B provides a technical description of the SGP Methodology. Appendix C is an investigation of potential ceiling and/or floor effects present in the Colorado assessment data and growth analyses.

[1] More in-depth treatment of the SGP Methodology can be found here and in Appendix B of this report.

2 Data

CDE supplied CMAS and PSAT/SAT data used in the 2018 SGP analyses to the NCIEA in summer of 2018. These test records were added to prior-years data from the Partnership for Assessment of Readiness for College and Careers (PARCC) consortium assessments to create the longitudinal data set from which the 2018 SGPs were calculated. Subsequent years' analyses will augment this multi-year data set allowing CDE to maintain comprehensive longitudinal data for all students taking the CMAS and PSAT/SAT assessments.

Student Growth Percentiles are produced for students that have a current score and at least one prior score in either the same subject or a related content area. For the 2018 academic year SGPs were produced for both CMAS and PSAT/SAT in English Language Arts (ELA) and Mathematics.

2.1 Longitudinal Data

Growth analyses on assessment data require data that are linked to individual students over time. Student growth percentile analyses require, at a minimum two, and preferably three years of assessment data for analysis of student progress. To this end it is necessary that a unique student identifier be available so that student data records across years can be merged with one another and subsequently examined. Because some records in the assessment data set contain students with more than one test score in a content area in a given year, a process to create unique student records in each content area by year combination was required in order to carry out subsequent growth analyses. Furthermore, student records may be invalidated for other reasons. The following business rules were used to either invalidate particular student records or select the appropriate record for use in the analyses.

2.1.1 General business rules

1. Student records are invalidated if the student identifier is missing.
2. Student records with missing ("NA") scores or scale scores outside of the possible range are invalidated.
3. Student records with any administrative invalidation flag (for example, identifying test irregularities, students that did not attempt the test, or other issues) are invalidated.
4. If a student has multiple records from the same subject and administration period, their highest score was selected.

Table 1 shows the number of valid CMAS student records available for analysis after applying the record invalidation business rules.[2]

Table 1: Number of Valid CMAS Student Records by Grade and Subject for 2018

	Grades					
Content Area	3	4	5	6	7	8
ELA	63,016	64,789	65,359	63,647	60,907	58,684
Mathematics	64,714	65,995	65,516	63,765	59,983	49,189

Table 2 shows the total number of valid PSAT/SAT student records available for analysis after applying the record invalidation business rules.

Table 2: Total Number of Valid PSAT/SAT Student Records by Subject for 2018

Content Area	Valid Records
ELA PSAT 9	63,295
ELA PSAT 10	59,703
ELA SAT	57,423
Mathematics PSAT 9	63,295
Mathematics PSAT 10	59,703
Mathematics SAT	57,423

[2] This does not represent the number of SGPs produced, however, because students are required to have at least one prior score available as well.

3 Analytics

This section provides basic details about the calculation of student growth percentiles from Colorado state assessment data using the R Software Environment (R Development Core Team, 2017) in conjunction with the SGP package (Damian W. Betebenner, VanIwaarden, Domingue, & Shang, 2018).

Broadly, the SGP analysis of the Colorado longitudinal student assessment data takes place in two steps:

1. Data Preparation
2. Data Analysis

Those familiar with data analysis know that the bulk of the effort in the above two step process lies with Step 1: Data Preparation. Following thorough data cleaning and preparation, data analysis using the `SGP` package takes clean data and makes it as easy as possible to calculate, summarize, output and visualize the results from SGP analyses.

3.1 Data Preparation

The data preparation step involves taking data provided by CDE and producing a `.Rdata` file that will subsequently be analyzed in Step 2. This process is carried out annually as new data becomes available from the state assessment program.

The text files that CDE provides were cleaned and processed by first reading them into R and then slightly modifying the variables. The result is a `.Rdata` file containing data in the format suitable for analysis with the SGP package. This data is combined with a subset of prior years' PARCC data necessary to complete the 2018 analyses. With an appropriate longitudinal data set prepared, we move to the calculation of student-level SGPs.

3.2 2018 Data Analysis

The objective of the student growth percentile (SGP) analysis is to describe how (a)typical a student's growth is by examining his/her current achievement relative to students with a similar achievement history; i.e his/her *academic peers* (see this presentation). This norm-referenced growth quantity is estimated using quantile regression (Koenker, 2005) to model curvilinear functional relationships between student's prior and current scores. One hundred such regression models are calculated for each separate analysis (defined as a unique ***year*** *by* ***content area*** *by* ***grade*** *by* ***prior order*** combination). The end product of these 100 separate regression models is a single coefficient matrix, which serves as a look-up table to relate prior student achievement to current achievement for each percentile. This process ultimately leads to tens of thousands of model calculations during each of Colorado's annual batch of analyses. For a more in-depth discussion of SGP calculation, see Betebenner (2009) and Appendix B of this report.

The 2018 Colorado SGP analyses follow a work flow that includes the following 4 steps:

1. Update the Colorado assessment meta-data required for SGP calculations using the `SGP` package.

2. Create annual SGP configurations for analyses.
3. Conduct all CMAS and PSAT/SAT SGP analyses.
4. Combine results into the master longitudinal data set, summarize results and output data.

3.2.1 Update Colorado meta-data

The use of higher-level functions included in the `SGP` package (e.g. `analyzeSGP`) requires the availability of state specific assessment information. This meta-data is compiled in a `R` object named `SGPstateData` that is housed in the package. Given the transition from the PARCC assessment program, significant updates to Colorado's metadata were required for the 2018 analyses. Information was required from both PARCC and CMAS assessments, including knot and boundary values, proficiency level cutscores and other assessment program information used to configure analyses and facilitate student- and group-level reporting.

Knots and boundaries

Cubic B-spline basis functions are used in the calculation of SGPs to more adequately model the heteroscedasticity and non-linearity found in assessment data.[3] These functions require the selection of boundary and interior knots. Boundary knots (i.e. "boundaries") are end-points outside of the scale score distribution that anchor the B-spline basis. These are typically selected by extending the entire range of scale scores by 10%. That is, they are defined as lying 10% below the lowest obtainable/observed scale score (LOSS) and 10% above the highest obtainable/observed scale score (HOSS). The interior knots (i.e. "knots") are the *internal* breakpoints that define the spline. The default choice in the `SGP` package is to select the 20^{th}, 40^{th}, 60^{th} and 80^{th} quantiles of the observed scale score distribution.

In general the knots and boundaries are computed from a distribution comprised of several years of test data (i.e. multiple cohorts combined) so that any irregularities in a single year are smoothed out. This is important because subsequent annual analyses use these same knots and boundaries as well. All defaults were used to compile the knots and boundaries for Colorado from the PARCC tests in previous years, and were also used in 2018 to compute the CMAS knots and boundaries required for the calculation of student growth projections. Official knots and boundaries will be recalculated for Colorado CMAS assessments in 2018 when two years of test data are available and at which point they will be used as the dependent variables in quantile regressions.

Achievement level cutscores

Cutscores, which are set externally through standard-setting processes (previously by PARCC and currently by CDE), are mainly required for student growth projections and goodness of fit plots which use prior achievement level information. Both PARCC and CMAS cutscores and other achievement level metadata (such as labels and descriptions) were also updated to reflect the new CMAS standards.

[3] It should be noted that the independent estimation of the regression functions can potentially result in the crossing of the quantile functions. This occurs near the extremes of the distributions and is more likely to occur given the use of non-linear functions. A potential result of allowing the quantile functions to cross would be *lower* percentile estimations of growth for *higher* observed scale scores at the extremes (give all else equal in prior scores) and vice versa. In order to deal with these contradictory estimates, quantile regression results are isotonized to prevent quantile crossing following the methods derived by Chernozhukov, Fernandez-Val and Glichon (2010).

3.2.2 Create SGP configurations

Unlike CMAS analyses, PSAT/SAT analyses are specialized enough so that it is necessary to specify the analyses to be performed via explicit configuration code. Configurations were used for CMAS SGP analyses as well, allowing for a more consistent and concise analytic process in which all analyses (including growth percentiles and projections) were run concurrently.

Configurations are `R` code scripts that are used as part of the larger SGP analysis to be discussed later. They are broken up into separate R scripts based on content domain (ELA and Mathematics). Each configuration code chunk specifies a set of parameters that defines the norm group of students to be examined. Every potential norm group is defined by, at a minimum, the progressions of content area, academic year and grade-level. Therefore, every configuration must contain the first three elements listed below. The PSAT/SAT analyses may also contain some or all of the fourth through seventh elements:

- `sgp.content.areas`: The progression of content areas to be looked at and their order.
- `sgp.panel.years`: The progression of the years associated with the content area progression (`sgp.content.areas`), potentially allowing for skipped or repeated years, etc.
- `sgp.grade.sequences`: The grade progression associated with the configuration content areas and years. The value **'EOCT'** stands for 'End Of Course Test'. The use of the generic 'EOCT' allows for secondary students to be compared based on the pattern of course taking rather than being dependent upon grade-level designation.
- `sgp.exact.grade.progression`: When set to `TRUE`, this element will force the lower level functions to analyze *only* the progression as specified in its entirety. Otherwise these functions will analyze subsets of the progression for every possible order (i.e. each number of prior time periods of data available). When set to `TRUE`, a norm group preference system is usually required as well.
- `sgp.norm.group.preference`: Because a student can potentially be included in more than one analysis/configuration, multiple SGPs will be produced for some students and a system is required to identify the preferred SGP that will be matched with the student in the `combineSGP` step. This argument provides a ranking that specifies how preferable SGPs produced from the analysis in question is relative to other possible SGPs. ***Lower numbers correspond with higher preference.*** Higher preference is typically given to:
  - Progressions with the greatest number of prior scale scores.
  - Progressions in which a student has repeated a course.
  - Progressions that do not include a skipped year (i.e. a gap in the scale score history).
- `sgp.projection.grade.sequences`: This element is used to identify the grade sequence that will be used to produce straight and/or lagged student growth projections. It can either be left out or set explicitly to `NULL` to produce projections based on the values provided in the `sgp.content.areas` and `sgp.grade.sequences` elements. Alternatively, when set to "`NO_PROJECTIONS`", no projections will be produced. For PSAT/SAT analyses, only configurations that correspond to the canonical course progressions can produce student growth projections. Canonical progressions are codified in the `SGP` package here: `SGPstateData[["RI"]][["SGP_Configuration"]][["content_area.projection.sequence"]]`.
- `sgp.exclude.sequences`: Lookup table containing the grade, subject, and year combinations of students that should be excluded from a cohort. This element is used in

progressions in which a year or similar time period is skipped (i.e. a gap in time exists). For example, in a progression that goes from 8^{th} grade Mathematics to PSAT 9 Mathematics with a skipped year in between one may want to exclude kids that repeated either 8^{th} grade Mathematics or took another math related subject (e.g. Algebra I) in the skipped year. Students with different course progressions may be inappropriate to include with the cohort of students who truly had no mathematics related course in the intervening year(s). No "skipped year" analyses were included in the 2018 analyses, and so this element was not used in any of the configurations.

As an example, here is one SAT Mathematics configuration used to define a 2018 SGP analysis:

```
...

 MATHEMATICS_SAT.2018 = list(
  sgp.content.areas=c("ALGEBRA_I", "MATHEMATICS_PSAT_10",
                      "MATHEMATICS_PSAT_10"),
  sgp.panel.years=c("2016", "2017", "2018"),
  sgp.grade.sequences=list(c("EOCT", "10", "11")),
  sgp.exact.grade.progression=TRUE,
  sgp.norm.group.preference=0,
  sgp.projection.grade.sequences=list("NO_PROJECTIONS")),
...
```

3.2.3 Conduct SGP analyses

The CMAS and PSAT/SAT data files were validated and made available at the same time allowing for all 2018 analyses to be conducted concurrently. The `abcSGP` function was used to perform all SGP analytic steps. This included the final preparation of the 2018 formatted data and creation of a new `SGP` class object in the `prepareSGP` step and calculation of SGP estimates in the `analyzeSGP` step. The `combineSGP` step merges the results into the master longitudinal data set once all analyses are completed. A pipe delimited version of the complete long data is saved in the `outputSGP` step. The `summarizeSGP` function is used to produce many tables of descriptive statistics that are disaggregated at the state, district and school levels, as well as other factors of interest. Finally, visualizations (such as bubble charts) are produced from the data and summary tables using the `visualizeSGP` function.

The following code was used to run all 2018 CMAS and PSAT/SAT analyses, including growth percentiles, projections, and summaries:

```
###  Read in 2018 P/SAT and CMAS SGP config scripts and combine

COLO_2018.config <- c(
    ELA.2018.config,
    ELA_PSAT_9.2018.config,
    ELA_PSAT_10.2018.config,
    ELA_SAT.2018.config,

    MATHEMATICS.2018.config,
    ALGEBRA_I.2018.config,
    GEOMETRY.2018.config,
    MATHEMATICS_PSAT_9.2018.config,
    MATHEMATICS_PSAT_10.2018.config,
    MATHEMATICS_SAT.2018.config
)


###  Use abcSGP to produce growth percentiles and projections

Colorado_SGP <- abcSGP(
        Colorado_Data_LONG,
        sgp.config = COLO_2018.config,
        steps=c(
            "prepareSGP",
            "analyzeSGP",
            "combineSGP",
            "summarizeSGP",
            "outputSGP"),
        sgp.percentiles = TRUE,
        sgp.projections = TRUE,
        sgp.projections.lagged = TRUE,
        sgp.percentiles.baseline=FALSE,
        sgp.projections.baseline = FALSE,
        sgp.projections.lagged.baseline = FALSE,
        simulate.sgps = FALSE,
        save.intermediate.results=FALSE,
        parallel.config = list(
            BACKEND="PARALLEL",
            WORKERS=list(PERCENTILES=14,
              PROJECTIONS=12, LAGGED_PROJECTIONS=10,
              SUMMARY=12)))
```

4 Goodness of Fit

Assessment data are generally imperfect and require sophisticated statistical methods to deal with the various issues they present. Cubic B-spline basis functions are used in the calculation of SGPs to more adequately model the heteroscedasticity, non-linearity and skewness. Despite this, assumptions that are made in the statistical modeling process can impact how well the percentile curves fit the data.[4] Accordingly a thorough evaluation of the models' fit is always required.

Examination of the Colorado Student Growth Model goodness-of-fit was conducted by first inspecting model fit plots the `SGP` software package produced for each analysis, and subsequently inspecting student level correlations between growth and achievement. Discussion of the model fit plots in general and examples of them are provided below, as are tables of the correlation results. The complete portfolio of model fit plots is provided in Appendix A.

4.1 Model Fit Plots

Using all available PARCC, CMAS and PSAT/SAT scores as the variables, estimation of student growth percentiles was conducted for each possible student (those with a current score and at least one prior score). Each analysis is defined by the grade and content area for the grade-level analyses and exact content area (and grade when relevant) sequences for the PSAT/SAT subjects. A goodness of fit plot is produced for each unique analysis run in 2018 and are all provided in Appendix A to this report.

As an example, Figure 1 shows the results for 5^{th} grade ELA as an example of good model fit. Figure 2 is an example of minor model misfit from the PSAT 9 Mathematics analysis that uses Geometry as the most recent prior.

The "Ceiling/Floor Effects Test" panel is intended to help identify potential problems in SGP estimation at the Highest and Lowest Obtainable (or Observed) Scale Scores (HOSS and LOSS). Most often these effects are caused when it is relatively typical for extremely high (low) achieving students to consistently score at or near the HOSS (LOSS) each year leading to the SGPs for these students to be unexpectedly low (high). That is, for example, if a sufficient number of students maintain performance at the HOSS over time, this performance will be estimated as typical, and therefore SGP estimates will reflect typical growth (e.g. 50th percentile). In some cases small deviations from these extreme score values might even yield low growth estimates. Although these score patterns can legitimately be estimated as a typical or low percentile, it is potentially an unfair description of actual student growth (and by proxy teacher or school, etc. performance metrics that use them). Ultimately this is usually an artifact of the assessments' inability to adequately measure student performance at extreme ability levels.

The table of values here shows whether the current year scale scores at both extremes yield

[4] Independent estimation of the regression functions can potentially result in the crossing of the quantile functions. This occurs near the extremes of the distributions and is more likely to occur given the use of non-linear functions. A potential result of allowing the quantile functions to cross would be *lower* estimated growth percentiles for *higher* observed scale scores at the extremes (give all else equal in prior scores) and vice versa. Quantile regression results are isotonized to prevent these contradictory estimates from quantile crossing following the methods derived by Chernozhukov, Fernandez-Val and Glichon (2010).

the expected SGPs[5]. The expectation is that the majority of SGPs for students scoring at or near the LOSS will be low (preferably less than 5 and not higher than 10), and that SGPs for students scoring at or near the HOSS will be high (preferably higher than 95 and not less than 90). Because few students may score *exactly* at the HOSS/LOSS, the top/bottom 50 students are selected and any student scoring within their range of scores are selected for inclusion in these tables. Consequently, there may be a range of scores at the HOSS/LOSS rather than a single score, and there may be more than 50 students included in the HOSS/LOSS row if the 50 students at the extremes only contain the single HOSS/LOSS score.

This table is meant to serve more as a "canary in the coal mine" than as a detailed, conclusive indicator of ceiling or floor effects, and a more fine grained analysis that considers the relationship between score histories and SGPs may be necessary. Appendix C of this report provides a more in depth investigation.

The two bottom panels compare the observed conditional density of the SGP estimates with the theoretical (uniform) density. The bottom left panel shows the empirical distribution of SGPs given prior scale score deciles in the form of a 10 by 10 cell grid. Percentages of student growth percentiles between the 10^{th}, 20^{th}, 30^{th}, 40^{th}, 50^{th}, 60^{th}, 70^{th}, 80^{th}, and 90^{th} percentiles were calculated based upon the empirical decile of the cohort's prior year scaled score distribution[6]. With an infinite population of test takers, at each prior scaled score, with perfect model fit, the expectation is to have 10 percent of the estimated growth percentiles between 1 and 9, 10 and 19, 20 and 29, ..., and 90 and 99. Deviations from 10 percent, indicated by red and blue shading, suggests lack of model fit. The further *above* 10 the darker the red, and the further *below* 10 the darker the blue.

When large deviations occur, one likely cause is a clustering of scale scores that makes it impossible to "split" the score at a dividing point forcing a majority of the scores into an adjacent cell. This occurs more often in lowest grade levels where fewer prior scores are available (particularly in the lowest grade when only a single prior is available). Another common cause of this is small cohort size (e.g. fewer than 5,000 students).

The bottom right panel of each plot is a Q-Q plot which compares the observed distribution of SGPs with the theoretical (uniform) distribution. An ideal plot here will show black step function lines that do not deviate greatly from the ideal, red line which traces the 45 degree angle of perfect fit.

4.1.1 Examples of model fit

We provide two examples here of cohort-referenced model fit. Overall the 2018 CMAS and PSAT/SAT results in all subjects are excellent with few exceptions. See Appendix A for all goodness of fit plots.

Figure 1 shows the results for 5^{th} grade ELA as an example of good model fit.

[5] Note that the prior year scale scores are not represented here, but are also a critical factor in ceiling effects.

[6] The total students in each analysis varies depending on grade and subject, and prior score deciles are based only on scores for students used in the SGP calculations.

Figure 1: Goodness of Fit Plot for 2018 5^{th} Grade ELA: Example of good model fit.

Student Growth Percentile Goodness-of-Fit Descriptives

2018 ELA SGP, Grade 5 (N = 59,976)

Ceiling / Floor Effects Test

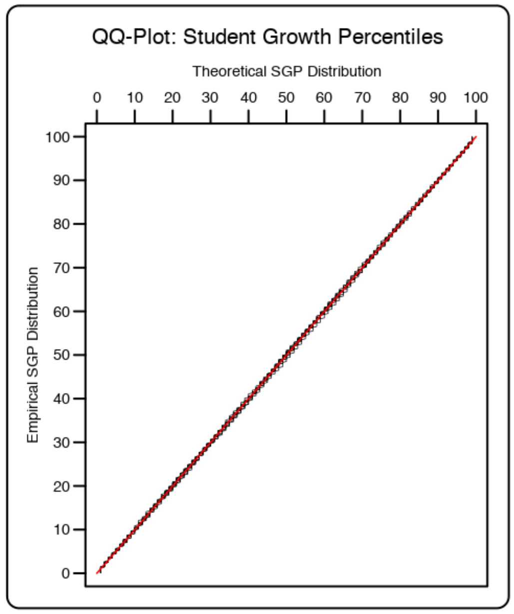
Student Growth Percentile Range

Current Observed Score Extremes (Range)		1 to 5	5 to 9	10 to 19	20 to 29	30 to 39	40 to 49	50 to 59	60 to 69	70 to 79	80 to 89	90 to 94	95 to 99	
Lowest OSS:	[650]	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	N = 202
Highest OSS:	[850]	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	8.9	12.5	78.6	N = 56

Student Growth Percentile Range

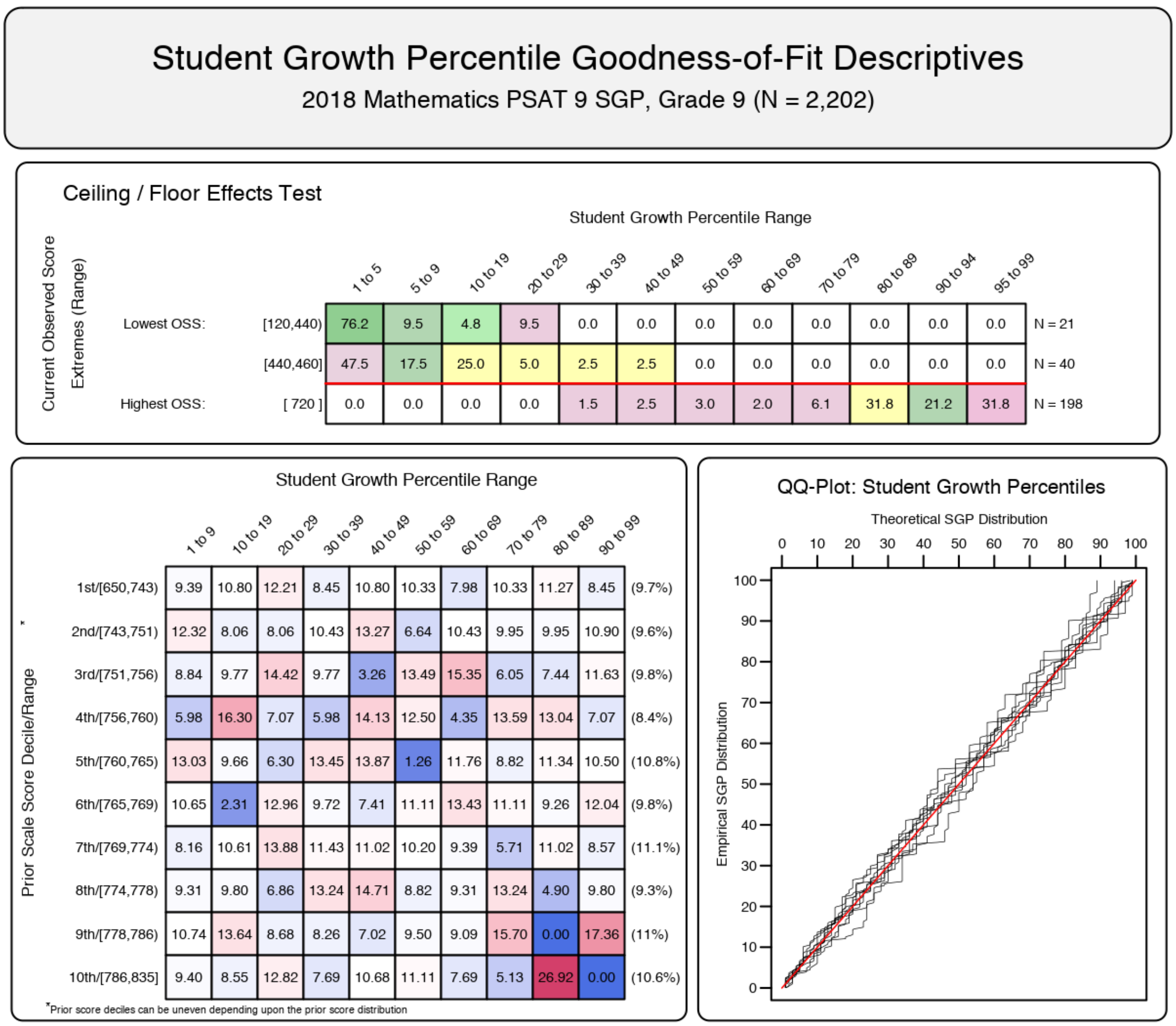
Prior Scale Score Decile/Range*	1 to 9	10 to 19	20 to 29	30 to 39	40 to 49	50 to 59	60 to 69	70 to 79	80 to 89	90 to 99	
1st/[650,695)	9.55	10.28	9.92	9.93	10.13	9.88	9.84	10.15	9.88	10.44	(9.2%)
2nd/[695,713)	9.51	9.38	10.05	9.68	9.68	10.16	10.35	10.38	10.12	10.68	(10.7%)
3rd/[713,725)	9.05	10.45	9.57	10.06	10.71	10.09	9.75	9.72	10.04	10.56	(9.3%)
4th/[725,736)	9.65	9.60	10.15	10.09	9.87	9.76	10.23	10.03	10.17	10.45	(10.6%)
5th/[736,745)	9.36	10.23	9.78	9.90	9.75	10.26	9.61	10.09	10.23	10.79	(10%)
6th/[745,753)	9.81	10.21	9.79	9.81	10.06	9.97	10.03	9.68	10.03	10.63	(9.1%)
7th/[753,763)	9.61	9.71	10.29	9.83	10.58	9.80	10.33	9.81	9.84	10.20	(10.9%)
8th/[763,774)	9.55	10.21	9.83	10.61	9.85	9.81	9.42	10.64	9.78	10.31	(10.1%)
9th/[774,788)	9.09	9.98	9.93	9.68	10.29	10.43	10.11	9.93	9.81	10.75	(9.4%)
10th/[788,850]	9.67	10.03	9.89	9.91	10.05	9.98	10.14	9.88	10.09	10.36	(10.7%)

*Prior score deciles can be uneven depending upon the prior score distribution



Minor misfit in the PSAT 9 Mathematics model is likely due in large part to the relatively small cohort size (2,000+ kids). This cohort is a relatively homogenous cohort of high academic achievers compared to others (more than 80% meeting or exceeding proficiency in Geometry). The indication of a ceiling effect in this plot is a potential concern here as well, likely again due to the cohort's high achievement status resulting in a higher frequency of perfect scores on both prior and current assessments. These results are examined in greater detail in Appendix C of this report.

Figure 2: Goodness of Fit Plot for a 2018 Geometry Progression: Example of slight model mis-fit.



4.2 Growth and Prior Achievement at the Student Level

To investigate the possibility that individual level misfit might impact summary level results, student level SGP results were examined relative to prior achievement. With perfect fit to data, the correlation between students' most recent prior achievement scores and their student growth percentiles is zero (i.e., the goodness of fit tables would have a uniform distribution of percentiles across all previous scale score levels). To investigate in another way, correlations between **a)** prior and current scale scores (achievement) and **b)** prior score and student growth percentiles were calculated. Evidence of good model fit begins with a strong positive relationship between prior and current achievement, which suggests that growth is detectable and modeling it is reasonable to begin with. A lack of relationship (zero correlation) between prior achievement and growth confirms that the model has fit the data well and produced a uniform distribution of percentiles across all previous scale score levels.

Student-level correlations for grade-level CMAS subjects are presented in Table 3, and the results are generally as expected. Strong relationships exist between prior and current scale scores for the grade level analyses (column 3). The correlation between students' most recent prior achievement scores and their student growth percentiles is zero when the model is perfectly fit to the data. This also indicates that students can demonstrate high (or low) growth regardless of prior achievement.

4.2.1 Grade-level CMAS subjects

Table 3: CMAS Student Level Correlations between Prior Standardized Scale Score and 1) Current Scale Score and 2) SGP.

Content Area	Grade	$r_{TestScores}$	r_{SGP}	N Size
ELA	4	0.83	0.00	59,469
	5	0.81	0.00	59,976
	6	0.81	0.00	58,201
	7	0.83	0.00	55,423
	8	0.83	0.00	52,761
Mathematics	4	0.85	0.00	61,206
	5	0.85	0.00	60,818
	6	0.85	0.00	58,262
	7	0.85	0.00	54,569
	8	0.82	0.00	43,294

4.2.2 PSAT/SAT Subjects

PSAT/SAT test subjects may be analyzed using more than one sequence of prior subjects, grades and years, and these unique progressions are disaggregated in Table 4 using the most recent prior available for each norm group (although more prior years' scores are used in SGP calculations when available). The correlations between current and prior scale score here are notably lower than in the CMAS norm groups. Lower correlations may be expected in PSAT/SAT subjects due to lack of instruction in test content (e.g. students who have not yet received instruction in Geometry are tested on Geometry related items) or other issues of misalignment of the test content and Colorado's standards and curriculum.

Table 4: PSAT/SAT Student Level Correlations between Prior Standardized Scale Score and 1) Current Scale Score and 2) SGP - Disaggregated by Norm Group.

Content Area	Most Recent Prior	$r_{TestScores}$	r_{SGP}	N Size
ELA PSAT 9	2017 ELA Grade 8	0.77	0.00	50,794
ELA PSAT 10	2017 ELA 9	0.78	0.01	44,448
ELA SAT	2017 ELA PSAT 10	0.87	-0.01	51,881
Mathematics PSAT 9	2017 Algebra I	0.77	0.05	8,947
	2017 Geometry	0.68	0.00	2,202
	2017 Math Grade 8	0.78	0.00	38,506
Mathematics PSAT 10	2017 Algebra I	0.71	0.01	25,187
	2017 Geometry	0.75	0.01	7,968
	2017 Integrated Math 1	0.80	0.00	7,390
Mathematics SAT	2017 Math PSAT 10	0.68	-0.01	51,881

5 SGP Results

Growth percentiles, being quantities associated with each individual student, can be easily summarized across numerous grouping indicators to provide summary results regarding growth. The median and mean of a collection of growth percentiles are used as measures of central tendency that summarize the distribution as a single number. With perfect data fit, we expect the state median of all student growth percentiles in any grade to be 50 because the data are norm-referenced across all students in the state. Median (and mean) growth percentiles well below 50 represent growth less than the state "average" and median growth percentiles well above 50 represent growth in excess of the state "average".

To demonstrate the norm-referenced nature of the growth percentiles viewed at the state level, Tables 5 and 6 present the cohort-referenced growth percentile medians and means for the CMAS and PSAT/SAT content areas respectively.

Table 5: 2018 CMAS Median (Mean) Student Growth Percentile by Grade and Content Area.

	Grades				
Content Area	4	5	6	7	8
ELA	50 (49.9)	50 (50)	50 (50.1)	50 (50)	50 (49.9)
Mathematics	50 (49.9)	50 (50)	50 (50)	50 (50.1)	50 (49.9)

Table 6: 2018 PSAT/SAT Median and Mean Student Growth Percentile by Content Area.

Content Area	Median SGP	Mean SGP
Ela PSAT 9	51	50.4
Ela PSAT 10	51	50.6
Ela SAT	50	50.2
Mathematics PSAT 9	50	50.0
Mathematics PSAT 10	50	50.1
Mathematics SAT	50	50.3

Based upon perfect model fit to the data, the median of all state growth percentiles in each grade by year by subject combination should be 50. That is, in the conditional distributions, 50 percent of growth percentiles should be less than 50 and 50 percent should be greater than 50. Deviations from 50 indicate imperfect model fit to the data. Imperfect model fit can occur for a number of reasons, including issues with the data (e.g., floor and ceiling effects leading to a "bunching" up of the data) or how the SGP function fits the data. PSAT/SAT results are further complicated by aggregating across multiple analyses (prior course patterns). The results in Tables 5 and 6 are nearly perfect, with almost all values equal to 50.

The results are coarse in that they are aggregated across tens of thousands of students. More refined fit analyses were presented in the Goodness-of-Fit section. Depending upon feedback from CDE, it may be desirable to tweak some operational parameters and attempt to improve fit even further. The impact upon the operational results based on better fit is expected to be extremely minor.

It is important to note how, at the entire state level, the *norm-referenced* growth information returns little information on annual trends due to its norm-reference nature. What the results indicate is that a typical (or average) student in the state demonstrates 50^{th} percentile growth. That is, "typical students" demonstrate "typical growth". One benefit of the norm-referenced results follows when subgroups are examined (e.g., schools, district, demographic groups, etc.) Examining subgroups in terms of the mean or median of their student growth percentiles, it is then possible to investigate why some subgroups display lower/higher student growth than others. Moreover, because the subgroup summary statistic (i.e., the median) is composed of many individual student growth percentiles, one can break out the result and further examine the distribution of individual results.

5.1 Group Level Results

Unlike reporting SGPs at the individual level, when aggregating to the group level (e.g., school) the correlation between aggregate prior student achievement and aggregate growth is rarely zero. The correlation between prior student achievement and growth at the school level is a compelling descriptive statistic because it indicates whether students attending schools serving higher achieving students grow faster (on average) than those students attending schools serving lower achieving students. Results from previous state analyses show a correlation between prior achievement of students associated with a current school (quantified as percent at/above proficient) and the median SGP are typically between 0.1 and 0.3 (although higher numbers have been observed in some states as well). That is, these results indicate that on average, students attending schools serving lower achieving students tend to demonstrate less exemplary growth than those attending schools serving higher achieving students. Equivalently, based upon ordinary least squares (OLS) regression assumptions, the prior achievement level of students attending a school accounts for between 1 and 10 percent of the variability observed in student growth. There are no definitive numbers on what this correlation should be, but recent studies on value-added models show similar results (McCaffrey, Han, & Lockwood, 2008).

5.1.1 School Level Results

To illustrate these relationships visually, the bubble charts in Figures 3 and 4 depict growth as quantified by the median SGP of students at the school against prior achievement status, quantified by percentage of student at/above proficient at the school. "Prior Percent at/above Proficient" in this case is determined by the percent of student's that scored in the "Level 4" or "Level 5" range of the prior year's PARCC test out of all student's that received a score. The charts have been successful in helping to motivate the discussion of the two qualities: student achievement and student growth. Though the figures are not detailed enough to indicate strength of relationship between growth and achievement, they are suggestive and valuable for discussions with stakeholders who are being introduced to the growth model for the first time. Only charts for the CMAS subjects are shown here.

Figure 3: School-level Bubble Plots for Colorado: ELA, 2017-2018.

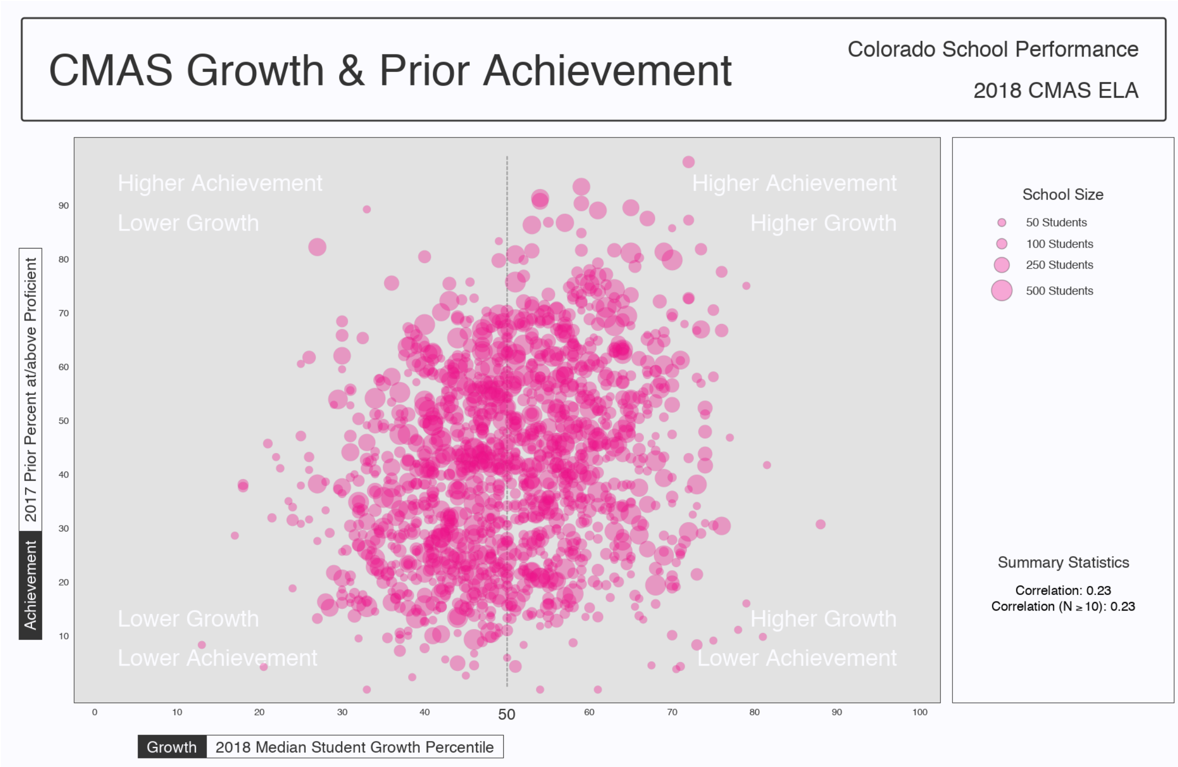


Figure 4: School-level Bubble Plots for Colorado: Mathematics, 2017-2018.

The relationship between average prior student achievement and median SGP observed for Colorado relatively average compared with other states for which the NCIEA has done SGP analyses. Table 7 shows overall correlation between prior achievement (measured here as the mean prior standardized scale score) for the previous three years. All results shown here are for schools with 10 or more students.

Table 7: Correlations between Mean Prior Standardized Scale Score and Aggregate SGPs - (Combined Subjects)

Year	Median SGP	Mean SGP
2016	0.33	0.35
2017	0.36	0.38
2018	0.36	0.38

Correlation tables describing the relationship between prior standardized scale score and aggregate growth percentiles are presented below in separate subsections for CMAS and PSAT/SAT subjects. The first correlation table in the each subsection provides these overall SGP aggregates' relationships with mean prior standardized scale scores. The additional correlation tables are disaggregated by content area, and content area and grade to provide more detail.

CMAS Content Areas

Table 8: School Level CMAS Correlations between Mean Prior Standardized Scale Score and Aggregate SGPs by Content Area.

Content Area	Year	Median SGP	Mean SGP
ELA	2016	0.25	0.27
	2017	0.23	0.24
	2018	0.27	0.29
Mathematics	2016	0.32	0.35
	2017	0.32	0.33
	2018	0.33	0.34

Table 9: 2018 School Level CMAS Correlations between Mean Prior Standardized Scale Score and Aggregate SGPs by Content Area and Grade.

Content Area	Grade	Median SGP	Mean SGP
ELA	4	0.16	0.15
	5	0.12	0.12
	6	0.15	0.16
	7	0.10	0.11
	8	0.16	0.18
Mathematics	4	0.12	0.13
	5	0.11	0.11
	6	0.18	0.20
	7	0.23	0.25
	8	0.30	0.33

PSAT/SAT Subjects

Table 10: School Level PSAT/SAT Correlations between Mean Prior Standardized Scale Score and Aggregate SGPs by Content Area.

Content Area	Year	Median SGP	Mean SGP
ELA PSAT 9	2018	0.21	0.23
ELA PSAT 10	2017	0.29	0.30
	2018	0.29	0.30
ELA SAT	2017	0.40	0.44
	2018	0.50	0.53
Mathematics PSAT 9	2018	0.36	0.38
Mathematics PSAT 10	2017	0.39	0.40
	2018	0.39	0.41
Mathematics SAT	2017	0.58	0.63
	2018	0.47	0.52

References

Betebenner, D. W. (2009). Norm- and criterion-referenced student growth. *Educational Measurement: Issues and Practice*, *28*(4), 42–51.

Betebenner, D. W., VanIwaarden, A., Domingue, B., & Shang, Y. (2018). *SGP: Student growth percentiles & percentile growth trajectories.* Retrieved from sgp.io

Chernozhukov, V., Fernndez-Val, I., & Galichon, A. (2010). Quantile and probability curves without crossing. *Econometrica*, *78*(3), 1093–1125. Wiley Online Library.

Koenker, R. (2005). *Quantile regression.* Cambridge: Cambridge University Press.

McCaffrey, D., Han, B., & Lockwood, J. (2008). From data to bonuses: A case study of the issues related to awarding teachers pay on the basis of their students' progress. In *Performance incentives: Their growing impact on american k-12 education (conference).* Vanderbilt University, Nashville, TN: National Center for Performance Incentives.

R Development Core Team. (2017). *R: A language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org