The PARCC Consortium Student Growth Model

A Technical Overview of the Spring 2018 Student Growth Percentile Calculations

Damian W. Betebenner

Adam R. VanIwaarden

National Center for the Improvement of Educational Assessment (NCIEA)

December 31, 2018

1 Introduction

This report contains details on the Spring 2018 implementation of the student growth percentiles (SGP) model for the Partnership for Assessment of Readiness for College and Careers (PARCC) consortium and Pearson Education (Pearson). The National Center for the Improvement of Educational Assessment (NCIEA) contracted with Pearson to implement the SGP methodology using data derived from the PARCC assessments to create the PARCC Student Growth Model. The goal of the engagement with Pearson is to create a set of open source analytic techniques to conduct growth analyses for both Fall and Spring test administration periods.

The SGP methodology is an open source norm- and criterion-referenced student growth analysis that produces student growth percentiles and student growth projections/targets for each student in the PARCC consortium with adequate longitudinal data. This methodology can be used for many purposes including parent/student diagnostic reporting, institutional improvement, and school and educator accountability. Specifics about the manner in which growth is included in each PARCC member states' accountability systems can be provided by their respective state education agencies.

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 PARCC consortium students' data. This includes discussion of goodness of fit plots and the student-level correlations between SGPs and prior achievement.
- ***SGP Results -*** provides basic descriptive statistics from the 2018 analyses at the consortium and state levels.

This report also includes multiple appendices. Appendix A displays Goodness of Fit plots for each analysis conducted for Spring 2018. Appendix B provides a more comprehensive discussion of the SGP methodology. Appendix C is an investigation of potential ceiling and/or floor effects present in the PARCC consortium 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

Data for the PARCC grade level and end-of-course tests (EOCT) used in the SGP analyses were supplied by Pearson to the NCIEA for analysis in the summer of 2018. These PARCC test records were added to PARCC assessment program data from previous test administrations to create the longitudinal data set from which the Spring 2018 SGPs were calculated. Subsequent SGP analyses will augment this multi-year data set allowing Pearson to maintain a comprehensive longitudinal data set for all students taking the PARCC assessments.

Student Growth Percentiles have been 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 grade-level English Language Arts (ELA) and Mathematics, as well as for EOCT subjects including Algebra I, Algebra II and Geometry. There were no viable Integrated Mathematics analyses for the Spring 2018 data.

2.1 Longitudinal Data

Growth analyses of assessment data require test scores that are linked to individual students over time. Student growth percentile analyses require, at a minimum two (and preferably more) 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.

Table 1 shows the number of valid grade level student records available for analysis after applying a set of general business rules. The final raw data submissions provided by Pearson contained no duplicate records or other problematic cases, and therefore these records are all considered valid for use in the SGP models.[2]

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

	Grades								
Content Area	3	4	5	6	7	8	9	10	11
ELA	339,820	346,292	349,382	345,686	340,251	340,747	125,605	190,570	104,428
Mathematics	348,683	354,648	356,658	346,876	324,971	265,252			

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

Table 2 shows the total number of valid EOCT student records from spring 2018 available for analysis.

Table 2: Total Number of Valid EOCT Student Records by Subject for 2018

Content Area	Valid Records
Algebra I	220,753
Algebra II	125,419
Geometry	131,825
Integrated Math 1	757
Integrated Math 2	885
Integrated Math 3	193

3 Analytics

This section provides basic details about the calculation of student growth percentiles from PARCC consortium assessment data using the R Software Environment (R Development Core Team, 2018) in conjunction with the SGP package (Damian W. Betebenner, VanIwaarden, Domingue, & Shang, 2018). More in depth treatment of the data analysis process with code examples has been provided to Pearson staff directly.

Broadly, the SGP analysis of the PARCC consortium 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, output and visualize the results from SGP analyses.

3.1 Data Preparation

The data preparation step involves taking PARCC consortium assessment data provided by the Pearson Education Corporation and producing a single file that will be subsequently analyzed in Step 2. This process is carried out as new data becomes available after each assessment interval. Data from the spring 2018 testing period was provided to the Center for Assessment in July 2018. An individual comma separated values (.csv) file was provided for each state in the PARCC consortium.

A custom function was used to read each state specific file into R and simultaneously combine them into a single data table. The data was then cleaned up variable by variable to ensure that it conforms to data naming and variable format conventions used in the SGP software package. The result is a .Rdata file containing data in the format suitable for analysis with the SGP package. With an appropriate longitudinal data set prepared, we continued on with the calculation of student-level SGPs.

3.2 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* - those students beginning at the same place (see this presentation for a detailed description of this concept). 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 the PARCC consortium's seasonal batch of analyses (and

many more when simulation-extrapolation, or SIMEX, methods of measurement error corrections are performed). For a more in-depth discussion of SGP calculation, see Betebenner (2009) and Appendix B of this report, and see Shang, VanIwaarden and Betebenner (2015) for further information on the SIMEX measurement error correction methodology.

The Spring 2018 PARCC SGP analyses follow a work flow that includes the following four steps:

1. Update the PARCC assessment meta-data required for SGP calculations using the `SGP` package.
2. Create SGP configurations for analyses.
3. Conduct all grade level and EOCT SGP analyses.
4. Combine results into the master longitudinal data set and output data in the format requested by Pearson.

3.2.1 Update PARCC meta-data

The use of higher-level functions included in the `SGP` package (e.g. `analyzeSGP`) requires the availability of assessment specific information. The meta-data for all assessment programs is compiled in a `R` data file named `SGPstateData` that is housed in the package. The initial PARCC meta-data entry included *a)* the knots and boundary values for each assessment score distribution, *b)* the proficiency level cutscores designated by Pearson and the PARCC consortium members, and *c)* a table that identifies an ordered preference for norm groups from which SGPs might be produced. General information about the grades and content areas assessed and other analysis configuration metadata was ultimately included in the PARCC `SGPstateData` entry, which can be viewed here on Github.

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. These functions require the selection of boundary and interior knots. Boundary knots (i.e. "boundaries") are end-points outside of the test score distribution that anchor the B-spline basis. These are typically selected by extending the entire range of test 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 into a single distribution) 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 PARCC using the first two years of test data.

Proficiency level cutscores

Cutscores, which are set externally by the PARCC consortium through standard-setting processes, are mainly required for student growth projections. Cutscores for all content areas and grade levels were added. Additionally, cutscores corresponding to both the scaled scores and the base scores (Item Response Theory "theta" values) are included in the PARCC meta-data entry.

Norm group preferences

The process through which some SGP analyses are run can produce multiple SGPs for some students (particularly in EOCT math where students may have multiple prior math test scores). In order to identify which quantity will be used as the students' "official" SGP and subsequently merged into the master longitudinal data set, a system of norm group preferencing is established and is encoded into a lookup table included in the `SGPstateData`. In general, preference is given to:

- Progressions with the most recent prior scale score.
- Progressions with the most common course progression, which is generally also used in student growth projections (in order to ensure correspondence between their reported growth percentile and projection).
- Progressions with the greatest number of prior scale scores.

The next section describes the process that establishes individual course progressions with analyses configuration code, which incorporates the preferencing system within it.

3.2.2 Specify SGP configurations

Many SGP analyses are specialized enough that it is necessary to specify each analyses to be performed via explicit configuration code. Such configurations were required to conduct the EOCT mathematics SGP analyses for the PARCC consortium. Although not necessary for the grade level content areas, explicit configuration code were also used for these analyses to provide greater coding consistency and allow for all analyses to be run concurrently.

Configurations are `R` code scripts that are used to define each SGP analysis. They are broken up into four separate `R` scripts based on content domain (ELA and mathematics) and test score metric (IRT theta and scaled scores). Each configuration code chunk specifies a set of parameters that identifies 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 used contains the first three elements and some analyses also contain the fourth through sixth elements:

- `sgp.content.areas:` The ordered progression of content areas to be looked at.
- `sgp.panel.years:` The progression of the years associated with the content area progression (`sgp.content.areas`), potentially allowing for skipped or repeated years, fall to spring (within academic year) analyses, etc. Because some PARCC states administer tests in both Fall and Spring, a season or period indicator is required as well. This is encoded here as either a ".1" or ".2" suffix to the academic year (i.e. "2017_2018.1" for Fall 2017 and "2017_2018.2" for Spring 2018).
- `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' label 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.projection.grade.sequences:` This element is used to identify the grade sequence that will be used to produce 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 EOCT analyses, only configurations that correspond to the canonical course progressions can produce student growth projections. The canonical progressions are codified in the `SGP` package here: `SGPstateData[["PARCC"]][["SGP_Configuration"]][["content_area.projection.sequence"]]`. Note also that grade level Mathematics can have two different EOCT trajectories - one that follows the Integrated Math courses and another that goes from Algebra I to Geometry to Algebra II.
- `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 numeric ranking that specifies how preferable SGPs produced from the analysis in question is relative to other possible analyses. ***Lower numbers correspond with higher preference***.

As an example, here is a Geometry configuration used to define a 2018 SGP analysis with a skipped year (i.e. using 2016 Grade 8 Math as a prior for 2018 Geometry):

```
...

    GEOMETRY.2017_2018.2 = list(
        sgp.content.areas=c("MATHEMATICS", "GEOMETRY"),
        sgp.panel.years=c("2015_2016.2", "2017_2018.2"),
        sgp.grade.sequences=list(c("8", "EOCT")),
        sgp.exact.grade.progression=TRUE,
        sgp.projection.grade.sequences=list("NO_PROJECTIONS"),
        sgp.norm.group.preference=8L),

...
```

3.2.3 Conduct SGP analyses

All cohort-referenced (unadjusted) and SIMEX corrected grade level and EOCT SGPs were calculated concurrently. Data analysis is conducted using the R Software Environment in conjunction with the SGP package. Broadly, the analysis takes place in 6 steps.

1. `prepareSGP`
2. `analyzeSGP`
3. `combineSGP`
4. `summarizeSGP`
5. `outputSGP`

6. `visualizeSGP`

Because these steps are almost always conducted simultaneously, the `SGP` package has wrapper functions `abcSGP` and `updateSGP` that "wrap" the above 6 steps into a single function call and simplify the source code associated with the analysis. Documentation for all SGP functions are available online.

SGP analyses were conducted at both the consortium and state level. For all analyses we use the `abcSGP` function to complete all data management, analytic and visualization steps. In the `prepareSGP` step the final preparation and addition of the cleaned and formatted Spring 2018 data to prior years data is completed at which point we can calculate SGP estimates (`analyzeSGP` step).

Once all analyses were completed the results were merged into the master longitudinal data set (`combineSGP` step). A pipe delimited version of the complete long data is output (`outputSGP` step) and submitted to Pearson after some additional formatting to add requested information. Finally, visualizations (such as growth achievement charts) are produced from the results data using the `visualizeSGP` function.

4 Goodness of Fit

Despite the use of B-splines to accommodate heteroscedasticity and skewness of the test score distributions, assumptions that are made in the statistical modeling process can impact how well the percentile curves fit the data. Examination of how well the SGP model fit the PARCC assessment data was conducted by first inspecting goodness-of-fit plots the `SGP` software package produces 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 goodness-of-fit plots is provided in Appendix A of this report.

4.1 Model Fit Plots

Using all available grade level and EOCT 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 EOCT 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 8^{th} grade ELA as an example of good model fit. Figure 2 demonstrates minor model misfit in the Algebra I analyses that uses 2017 Geometry for its prior scores.

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 typical or low growth percentile, it is potentially an unfair description of student growth (and by proxy teacher or school, etc. performance metrics that use them). Ultimately this is 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 scores at both extremes yield the expected SGPs.[3] 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.

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

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 of potential ceiling effects.

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.[4] 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).

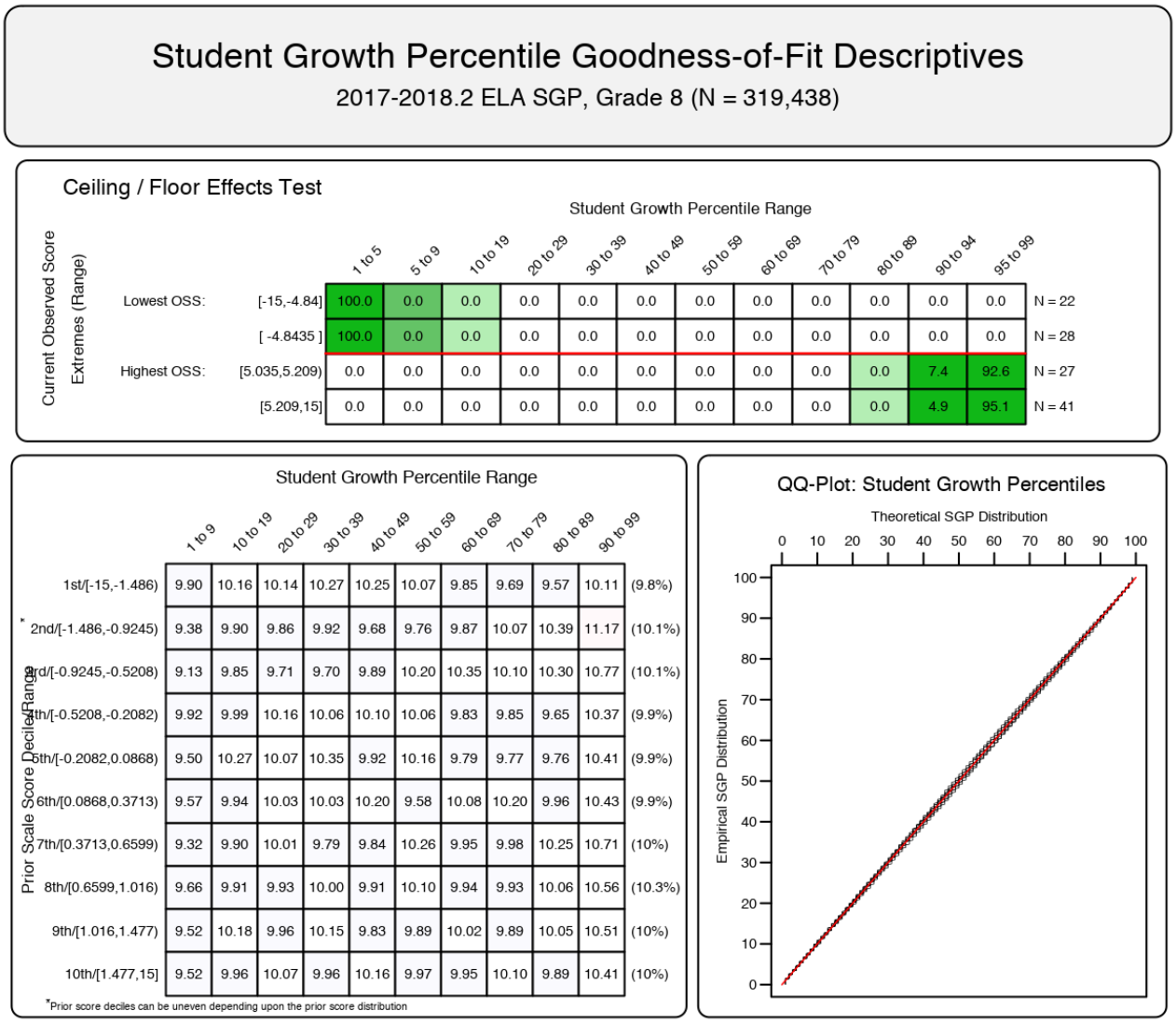
The bottom right panel of each plot is a Q-Q plot which compares the observed distribution of SGPs with the theoretical (uniform) distribution. An ideal plot here will show black step function lines that do not deviate greatly from the ideal, red line which traces the 45 degree angle of perfect fit.

4.1.1 Unadjusted model fit

The model fit plots for the unadjusted model results in all subjects are excellent with few exceptions (see Appendix A for all model fit plots). Figure 1 displays the 8^{th} grade ELA model as an exemplar of model fit, and Figure 2 is an Algebra I progression which exemplifies the extent to which model misfit is present in the PARCC SGP analyses.

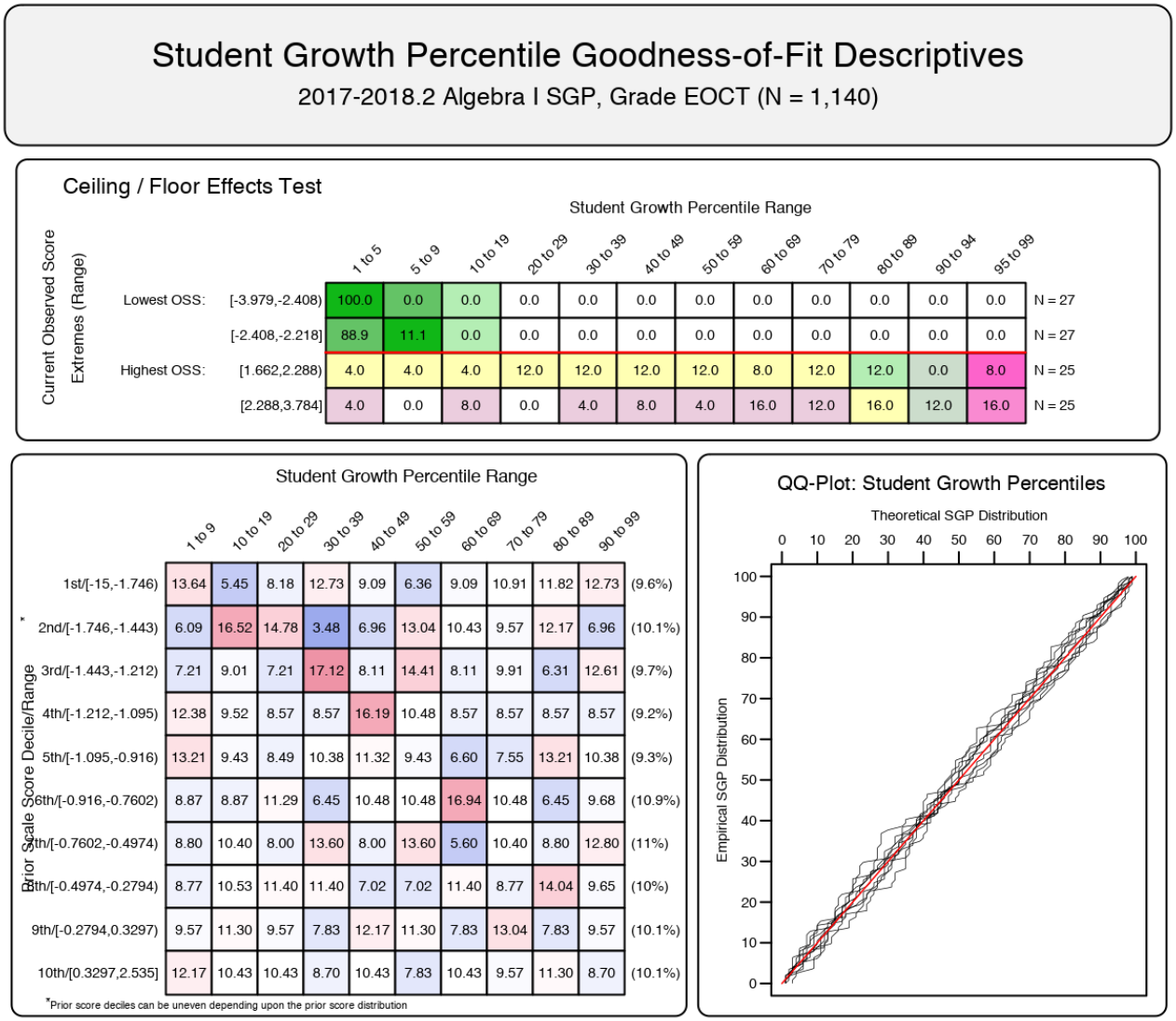
[4]The total students in each analysis varies depending on grade and subject, and prior score deciles are based only on scores for students used in the SGP calculations.

Figure 1: Goodness of Fit Plot for 2018 ***Uncorrected*** 8^{th} Grade ELA: Example of good model fit.



Misfit in the Algebra I model is likely due to the relatively small cohort size (1,140 students). The atypical course progression from Geometry to Algebra I also likely may mean the students in the cohort are atypical and may be more difficult to model (e.g. the decile score ranges in the SGP Range plot suggest the cohort is largely low achieving with a small group of high achieving outliers). The indication of a potential ceiling effect in this plot is probably overstated given the range of scores included in the HOSS (bottom) row. That is, the low SGPs are likely attributed to students scoring just below the HOSS, and therefore less concerning. The possibility of ceiling and floor effects in this, and all other, analyses is investigated in greater detail in Appendix C of this report.

Figure 2: Goodness of Fit Plot for 2018 ***Uncorrected*** Algebra I (Geometry Prior) Progression: Example of model mis-fit.



4.1.2 SIMEX model fit

Although the *official* PARCC SGPs do not incorporate SIMEX measurement error correction, we provide the adjusted model fit plots here to compare with unadjusted models. Note that both models presented here use the same student data as those in the section above.

SIMEX model fit is not expected to be perfect (i.e. uniformly distributed regardless of prior achievement). In these models we ***expect misfit*** in the form of increased high SGPs for students with lower prior performance (and a complementary decrease in low SGPs for those students), and the reverse expectation for high achieving students. This effect is alleviated to some extent through the ranking of the SIMEX SGPs (Castellano & McCaffrey, 2017). This shift is visible in the goodness of fit plots in Figure 3 where the Ranked SIMEX correction method has been applied to the 8^{th} grade ELA model, and Figure 4 for the same Algebra I progression as presented above. Most notable is the shift from blue to red in the top half (and red to blue in the bottom half) of the "Student Growth Percentile Range" panel.

The patterns observed in these two charts suggest that the SIMEX model adjustments are behaving as expected.

Figure 3: Goodness of Fit Plot for 2018 ***SIMEX Corrected*** 8^{th} Grade ELA: Example of good model fit.

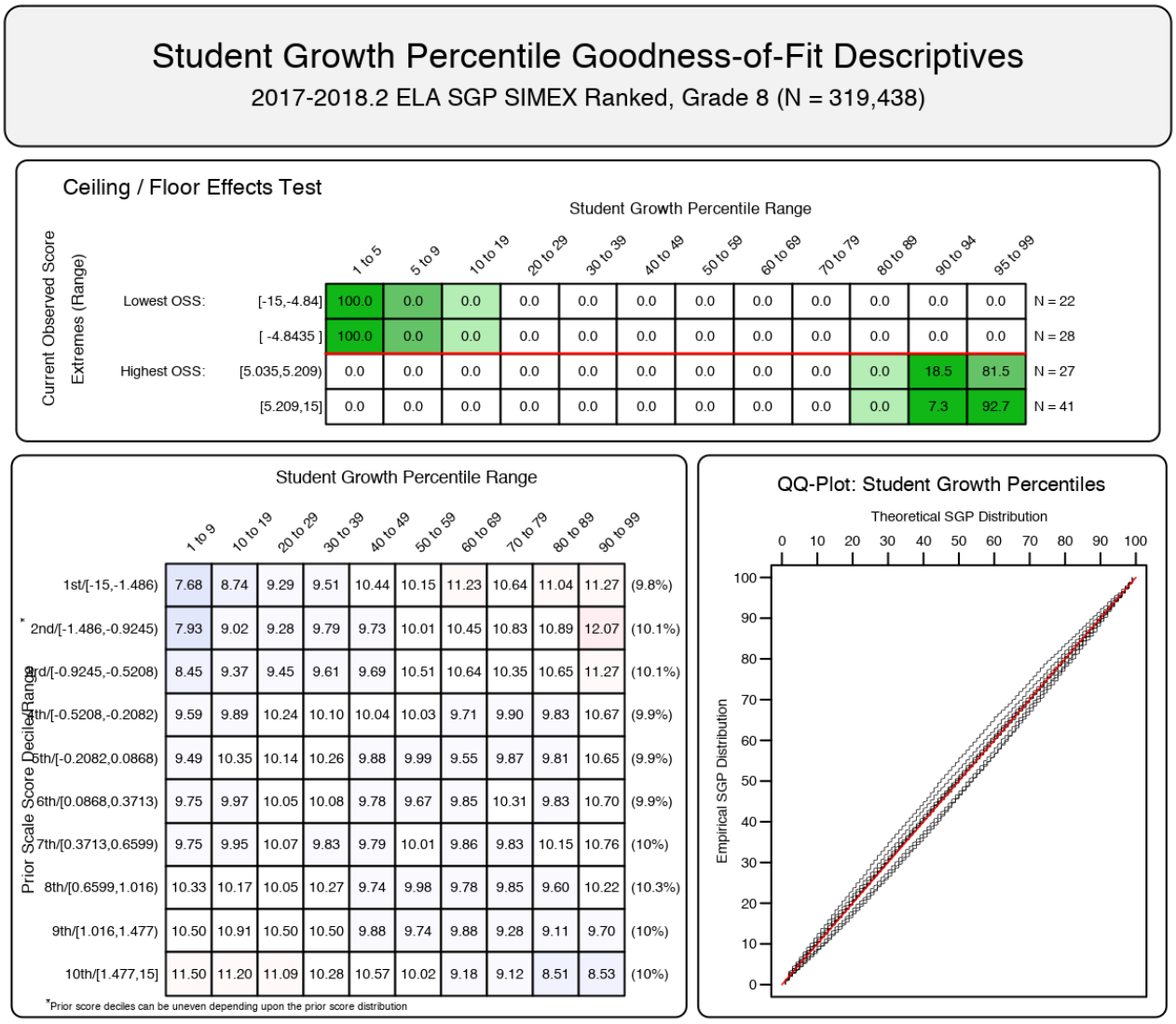
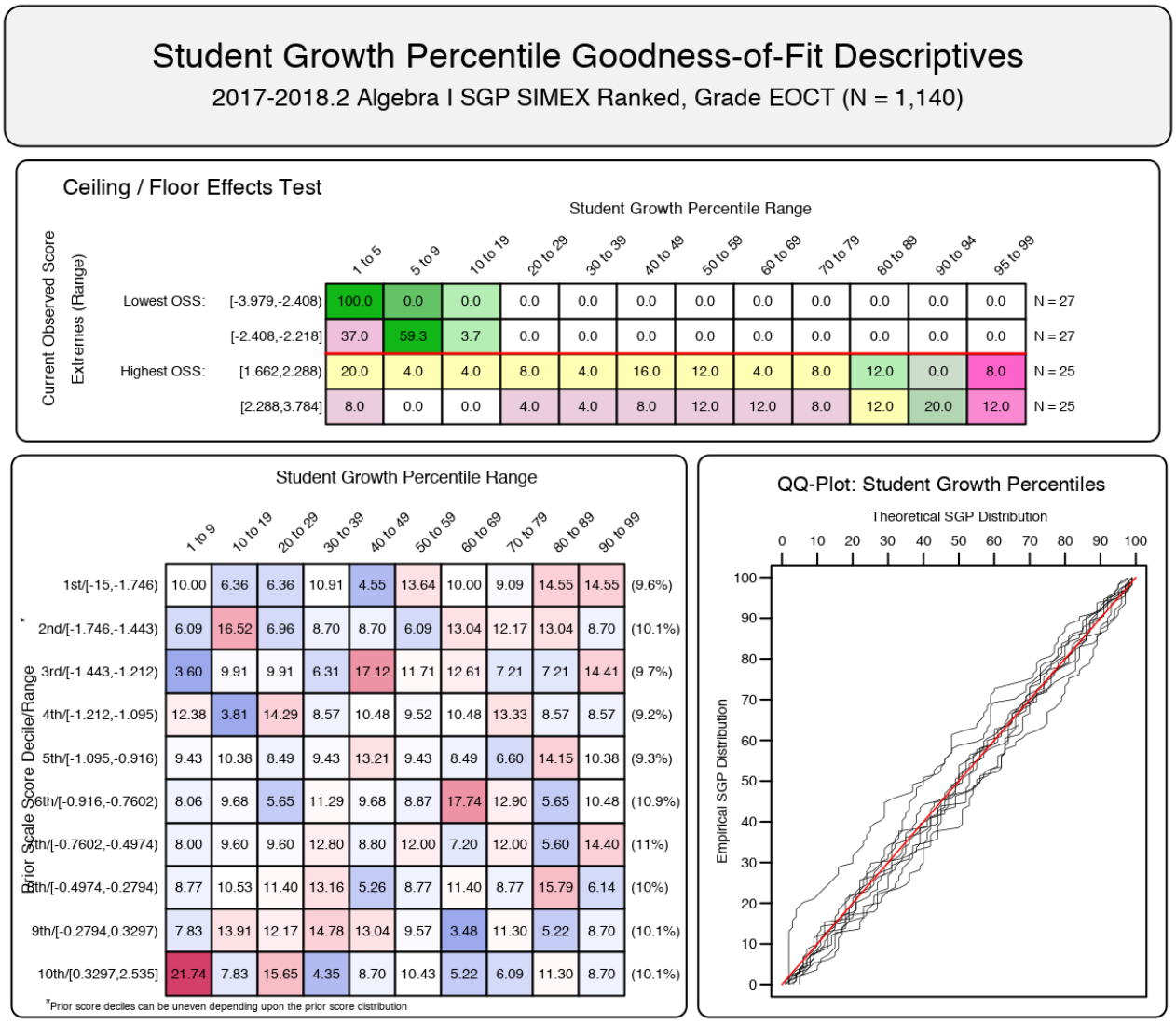


Figure 4: Goodness of Fit Plot for 2018 ***SIMEX Corrected*** Algebra I (Geometry Prior) Progression: Example of model mis-fit.



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. 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 correlation coefficients (r) for grade level 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). With cohort-referenced (unadjusted) percentiles, 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. Correlations for the PARCC consortium's unadjusted SGPs are all essentially zero (column 4).

SIMEX corrected SGPs induce a negative correlation between growth and prior achievement. Rather than a uniform distribution, SIMEX produces a distribution in which growth for lower prior achieving students' is weighted upward and higher achieving students' growth is weighted down. In theory, this bias in student-level SGPs may decrease the bias in aggregate growth measures (as documented in previous SIMEX reports and Shang et al., 2015). Subsequently, the correlations between both student- and aggregate-level "unadjusted" SGPs and prior achievement are typically higher than those with SIMEX corrected SGPs.[5]

[5]Correlations that are initially negative will become increasingly negative rather than return to zero.

4.2.1 Grade level subjects

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

Content Area	Grade	$r_{TestScores}$	r_{SGP}	$r_{SIMEXSGP}$	N Size
ELA	4	0.83	0.00	-0.08	328,358
	5	0.83	0.00	-0.06	331,928
	6	0.83	0.00	-0.06	321,697
	7	0.85	0.00	-0.06	317,804
	8	0.86	0.00	-0.06	319,438
	9	0.85	0.00	-0.05	110,582
	10	0.80	0.02	-0.04	164,530
	11	0.72	0.01	-0.04	92,622
Mathematics	4	0.82	0.00	-0.07	330,982
	5	0.84	0.00	-0.05	333,838
	6	0.83	0.00	-0.06	323,018
	7	0.82	0.00	-0.07	307,460
	8	0.76	0.00	-0.07	248,714

4.2.2 EOCT Subjects

Each EOCT subject is 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 can be used in SGP calculations when available). Unless otherwise noted, all priors are from the Spring of 2017. The correlations between current EOCT and prior test score are generally lower than in the grade level norm groups, and overall lower correlations may be expected in EOCT subjects due to the change in specific subject from one course to the next.

The relationships between growth and prior achievement reported in Table 4 are still non-existent for cohort referenced SGPs and slightly negative after SIMEX correction. These results are as expected for appropriate fit to the respective models as discussed in the grade level section above.

Table 4: EOCT Student Level Correlations between Prior Standardized Scale Score and 1) Current Scale Score, 2) SGP and 3) SIMEX SGP - Disaggregated by Norm Group.

Content Area	Most Recent Prior	$r_{TestScores}$	r_{SGP}	r_{SIMEX}	N Size
Algebra I	Geometry	0.59	0.01	-0.08	1,140
	Mathematics Grade 6	0.67	0.00	-0.06	11,734
	Mathematics Grade 7	0.83	0.00	-0.06	60,886
	Mathematics Grade 8	0.70	0.00	-0.08	97,237
Algebra II	Algebra I	0.72	0.05	-0.04	12,900
	Geometry	0.78	0.00	-0.06	82,403
	Mathematics Grade 8	0.81	0.01	-0.10	1,850
	Fall 2017 Geometry	0.84	-0.01	-0.10	945
Geometry	Spring 2016 Algebra I	0.51	-0.05	-0.13	3,355
	Spring 2016 Mathematics Grade 8	0.56	0.04	-0.06	2,196
	Algebra II	0.85	0.02	-0.05	3,640
	Algebra I	0.81	0.02	-0.06	110,573
	Mathematics Grade 8	0.81	0.03	-0.06	1,265
	Fall 2017 Algebra I	0.81	0.04	-0.09	696

5 SGP Results

In the following sections basic descriptive statistics from the 2018 analyses are provided, including the consortium-level mean and median growth percentiles. Currently the PARCC consortium uses unadjusted cohort-referenced SGPs as the official student-level growth metric. Descriptive statistics from the unadjusted and SIMEX corrected SGP results are both presented here. The interested reader can find more in depth discussions of the SGP methodology in Appendix B of this report as well as other available literature, and information about the SIMEX measurement error correction methodology is available in recent academic articles (see Shang et al., 2015).

5.1 Unadjusted SGPs

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 consortium-wide median of all student growth percentiles in any grade to be 50 because the data are norm-referenced across all students in PARCC. Median (and mean) growth percentiles well below 50 represent growth less than the consortium "average" and median growth percentiles well above 50 represent growth in excess of the PARCC "average".

To demonstrate the norm-referenced nature of the growth percentiles viewed at the consortium level, Tables 5 and 6 present the cohort-referenced growth percentile medians and means for the grade level and EOCT content areas respectively.

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

	Grades							
Content Area	4	5	6	7	8	9	10	11
ELA	50 (50)	50 (50.1)	50 (50)	50 (50)	50 (50)	50 (49.9)	54 (52.5)	50 (50.1)
Mathematics	50 (50)	50 (50.1)	50 (50)	50 (50)	50 (50)			

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, some due to issues with the data (e.g., floor and ceiling effects leading to a "bunching" up of the data) as well as issues due to the way that the SGP function fits the data. The results in Tables 4 and 5 are close to perfect, with almost all values equal to 50.

The notable exception to this is 10^{th} Grade ELA. In this case the deviation from 50 is due to aggregation of SGPs across two different analyses: the traditional course progression of grades 8, 9 and 10 and a second analysis that ignores the 9^{th} grade test score. Although all PARCC

students were used in the skipped grade analysis, states that use these SGPs for their students had higher typical SGPs (60+) relative to the entire PARCC cohort. Since only these states' have these SGPs included in and this is reflected in

Table 6: Spring 2018 EOCT Median and Mean Student Growth Percentile by Content Area.

Content Area	Median SGP	Mean SGP
Algebra I	50	50.0
Algebra II	50	49.7
Geometry	50	50.3

The results are coarse in that they are aggregated across hundreds of thousands of students. More refined fit analyses were presented in the Goodness-of-Fit section. Depending upon feedback from Pearson, 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 consortium 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.

5.2 Ranked SIMEX Adjusted SGPs

The use of error-prone standardized test scores in statistical models can lead to numerous problems. Understanding the effects of measurement error and correcting for it is particularly difficult in the SGP model given that it is based on non-parametric quantile regression. Preliminary investigations suggest that the use of error-prone measures may lead SGP estimates to be inflated for students with high prior achievement and underestimated for students with lower prior achievement. This bias at the individual student-level can translate to bias in aggregate measures (such as median and mean SGPs) when student sorting based on prior achievement exists at the level of aggregation (e.g. classrooms or schools). As a result, growth and prior achievement are (positively) correlated, giving schools and teachers with higher achieving students an undue advantage and disadvantaging those with lower achieving students (Shang et al., 2015).

Simulation-extrapolation (SIMEX) methods of correcting for measurement error induced bias in the SGP estimation has been proposed and tested in other states. This correction has been applied to the 2018 SGPs in addition to a system of ranking the SIMEX SGPs in

order to improve some student level properties of the growth measure (Castellano & McCaffrey, 2017). Descriptive statistics from the uncorrected and Ranked SIMEX corrected SGP results are both presented here. The interested reader can find more in depth discussions of the SGP methodology in the available literature, and information about the SIMEX measurement error correction methodology is available in recent academic articles (Castellano & McCaffrey, 2017; Shang et al., 2015).

Table 7: SIMEX Corrected Grade Level Median (Mean) Student Growth Percentile by Grade and Content Area for Spring 2018

	Grades							
Content Area	4	5	6	7	8	9	10	11
ELA	50 (50)	50 (50.1)	50 (50.1)	50 (50)	50 (50)	50 (49.9)	54 (52.7)	50 (50.2)
Mathematics	50 (50)	50 (50.1)	50 (50.1)	50 (50.1)	50 (50.1)			

Table 8: SIMEX Corrected EOCT Median (Mean) Student Growth Percentile by Content Area for Spring 2018

Content Area	Median SGP	Mean SGP
Algebra I	50	50.0
Algebra II	50	49.8
Geometry	50	49.9

A comparison of the unadjusted (Tables 5 and 6) and Ranked SIMEX corrected (Tables 7 and 8) shows very little difference in the medians and means. This is not surprising as the majority of the growth percentiles for students in the middle of the prior score distributions change very little, and the larger changes that occur for students in the extremes of the prior score distributions tend to even out.

5.3 Group Level Results

Unlike when 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 (D. McCaffrey, Han, & Lockwood, 2008).

5.3.1 States

In this section we aggregate test score and SGP results by state. At present, district, school or other group identification has not been attached to the student records provided to the Center for Assessment. Smaller group level analyses are therefore left to the individual consortium members to perform. In addition to mean SGP values, mean prior test scores are given using both the IRT Theta and PARCC scale metrics. This mean prior achievement value is given further context by providing percentile value it corresponds with in that test score distribution.

Summary values are provided in the following tables only when a state has *1,000 students or greater* in a norm group. A randomized (i.e. non-alphabetical) letter has been attached to each state in order to protect the identity of the member states, and results are shown only when there is more than one state that meets the 1,000 student threshold. As in previous tables of SGP results in this report, ELA and Mathematics results are disaggregated by grade level and EOCT Math tables are disaggregated by the most recent prior ensuring that all SGPs included in the summary statistics include only results from one analysis (norm group). Tables for the Integrated Mathematics SGPs are not presented as only two states had any one cohort with greater than 1,000 students, and therefore presenting these tables may compromise state anonymity.

Table 9: Spring 2018 Typical Prior Achievement and Growth by State: ELA

Grade	State	Mean Prior Scale Score	Mean Prior Theta Score	PARCC %ile (Mean Prior Theta)	Mean SGP	Mean SIMEX
4	A	734.9	-0.02	46	46.8	47.1
	B	725.2	-0.28	38	53.1	53.9
	C	750.1	0.40	60	56.0	55.4
	D	714.2	-0.58	29	40.3	41.8
	E	725.3	-0.28	38	45.7	46.5
	F	737.5	0.05	48	49.5	49.6
5	A	737.7	-0.11	44	45.4	45.7
	B	733.1	-0.25	40	52.1	52.4
	C	754.0	0.41	62	56.2	55.6
	D	716.2	-0.79	24	47.5	48.8
	E	727.3	-0.44	34	54.4	55.2
	F	741.1	0.00	47	49.6	49.7
6	A	737.2	-0.08	43	47.6	48.0
	B	733.4	-0.21	39	50.2	50.7
	C	756.1	0.56	64	53.7	53.0
	D	720.8	-0.64	27	51.4	52.6
	E	731.0	-0.29	37	52.6	53.2
	F	740.1	0.02	46	48.8	49.0

Table 10: Spring 2018 Typical Prior Achievement and Growth by State: ELA (*Continued*)

Grade	State	Mean Prior Scale Score	Mean Prior Theta Score	PARCC %ile (Mean Prior Theta)	Mean SGP	Mean SIMEX
7	A	737.6	-0.04	45	46.1	46.3
	B	732.1	-0.23	39	55.8	56.3
	C	752.4	0.48	62	56.0	55.4
	E	730.5	-0.29	37	45.7	46.3
	F	739.3	0.02	47	50.9	51.1
8	A	739.4	-0.09	44	45.3	45.5
	B	733.1	-0.28	38	52.3	52.7
	C	756.9	0.43	62	56.7	56.1
	E	727.3	-0.45	33	53.1	53.9
	F	741.6	-0.02	46	48.9	49.1
9	C	757.8	0.42	55	49.2	49.0
	E	729.1	-0.42	29	51.3	52.2
	F	730.8	-0.37	30	62.9	63.6
10	B	730.1	-0.39	40	43.1	43.9
	C	744.8	0.04	54	37.6	37.7
	C	751.0	0.33	53	48.4	48.4
	E	730.5	-0.27	31	53.4	54.0
	F	739.5	-0.12	49	60.5	60.8
	F	737.4	-0.07	38	69.2	69.6
11	C	738.0	-0.02	51	46.1	46.1
	E	731.7	-0.16	45	62.0	62.1
	F	724.6	-0.33	39	51.9	52.4

Table 11: Spring 2018 Typical Prior Achievement and Growth by State: Mathematics

Grade	State	Mean Prior Scale Score	Mean Prior Theta Score	PARCC %ile (Mean Prior Theta)	Mean SGP	Mean SIMEX
4	A	740.5	-0.15	46	45.8	46.0
	B	737.8	-0.24	44	50.4	50.8
	C	751.8	0.20	59	56.1	55.6
	D	720.3	-0.78	27	37.9	39.3
	E	731.7	-0.43	37	49.2	49.8
	F	742.6	-0.09	48	50.5	50.6
5	A	733.4	-0.27	45	47.1	47.4
	B	732.4	-0.30	45	51.9	52.0
	C	747.5	0.20	61	54.1	53.6
	D	710.5	-1.03	22	46.0	47.4
	E	727.2	-0.48	39	52.7	53.2
	F	737.8	-0.12	51	49.5	49.5
6	A	733.7	-0.30	46	49.5	49.8
	B	731.2	-0.39	43	45.6	46.1
	C	747.4	0.18	63	52.4	51.9
	D	717.5	-0.85	28	53.9	55.1
	E	728.2	-0.49	40	48.2	48.7
	F	736.8	-0.19	50	48.6	48.7

Table 12: Spring 2018 Typical Prior Achievement and Growth by State: Mathematics (*Continued*)

Grade	State	Mean Prior Scale Score	Mean Prior Theta Score	PARCC %ile (Mean Prior Theta)	Mean SGP	Mean SIMEX
7	A	731.6	-0.27	47	49.7	49.9
	B	724.0	-0.54	38	49.4	50.2
	C	741.7	0.10	59	51.3	50.8
	E	724.1	-0.53	38	45.9	46.7
	F	728.2	-0.39	42	50.6	50.9
8	A	732.5	-0.16	47	48.8	48.6
	B	720.3	-0.66	31	50.5	51.3
	C	729.8	-0.27	43	53.1	52.9
	E	718.5	-0.73	29	51.3	52.2
	F	718.9	-0.72	29	48.7	49.6

Table 13: Spring 2018 Typical Prior Achievement and Growth by State: Algebra I

Most Recent Prior	State	Mean Prior Scale Score	Mean Prior Theta Score	PARCC %ile (Mean Prior Theta)	Mean SGP	Mean SIMEX
Math Grade 6	C	795.3	2.07	97	51.0	49.9
	F	775.2	1.31	89	51.9	52.7
Math Grade 7	A	757.8	0.86	79	38.9	39.0
	C	764.7	1.13	85	50.1	49.6
	E	747.9	0.46	68	42.5	43.3
	F	748.9	0.50	69	53.7	54.4
Math Grade 8	C	730.0	0.02	57	48.7	48.2
	E	713.7	-0.48	39	45.8	46.6
	F	712.6	-0.53	37	54.5	55.3

Table 14: Spring 2018 Typical Prior Achievement and Growth by State: Algebra II

Most Recent Prior	State	Mean Prior Scale Score	Mean Prior Theta Score	PARCC %ile (Mean Prior Theta)	Mean SGP	Mean SIMEX
Algebra I	C	758.7	0.69	66	47.9	48.2
	E	731.8	-0.17	40	52.6	54.3
	F	760.5	0.74	68	52.5	52.7
Geometry	C	737.0	0.12	53	48.0	48.0
	E	728.7	-0.20	42	55.2	55.7
	F	747.9	0.54	67	54.9	54.2

Table 15: Spring 2018 Typical Prior Achievement and Growth by State: Geometry

Most Recent Prior	State	Mean Prior Scale Score	Mean Prior Theta Score	PARCC %ile (Mean Prior Theta)	Mean SGP	Mean SIMEX
Algebra I	B	718.0	-0.64	24	52.0	53.4
	C	744.4	0.23	52	49.3	48.6
	E	727.1	-0.32	34	55.0	55.5
	F	764.4	0.86	71	51.2	49.6

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

Castellano, K. E., & McCaffrey, D. F. (2017). The accuracy of aggregate student growth percentiles as indicators of educator performance. *Educational Measurement: Issues and Practice*, *36*(1), 14–27. Retrieved from http://dx.doi.org/10.1111/emip.12144

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. (2018). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org

Shang, Y., VanIwaarden, A., & Betebenner, D. W. (2015). Covariate measurement error correction for student growth percentiles using the simex method. *Educational Measurement: Issues and Practice*, *34*(1), 4–14. Retrieved from http://dx.doi.org/10.1111/emip.12058