The Utah Student Growth Model
A Technical Overview of the 2014-2015 Student Growth Percentile Calculations

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

April 2016

Abstract

DRAFT REPORT - DO NOT CITE! This report provides details about the Utah Student Growth Model methodology and presents a descriptive analysis of the 2015 SGP results.

1 Introduction

This report contains details on the implementation of the student growth percentiles (SGP) model for the state of Utah. The National Center for the Improvement of Educational Assessment (NCIEA) contracted with the Utah State Office of Education (USOE) to implement the SGP methodology using data derived from the Utah student assessment program to create the Utah Student Growth Model. The goal of the engagement with USOE is to create a set of open source analytic techniques and conduct a set of initial analyses that will eventually be conducted by USOE in following years.

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 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.

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 2015 analyses.[1]
- ***Goodness of Fit -*** describes 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 2015 analyses at both the state and school levels.

This report also includes multiple appendices. Appendix A displays Goodness of Fit plots for all content areas and grades. Appendix B provides a more technical description of the SGP methodology and statistical concepts. Appendix C is an investigation of potential ceiling and/or floor effects present in the Utah student assessment data and growth analyses.

[1] More in-depth treatment of the SGP Methodology can be found here.

2 Data

Data for the Utah Student Assessment of Growth and Excellence (SAGE) used in the SGP analyses were supplied by the USOE to the NCIEA for analysis in August 2015. The current longitudinal data set now includes academic years 2007-2008 through 2014-2015. Subsequent years' analyses will augment this multi-year data set allowing USOE to maintain a comprehensive longitudinal data set for all students taking the SAGE and EOCT assessments.

Student Growth Percentiles have been produced for students that have a current score and at least one prior score in the same subject or a related content area. SGPs were produced for English Language Arts (ELA), Science and Mathematics end-of-grade tests (EOGT). For the 2015 academic year SGPs were produced for end-of-course tests (EOCT) for Earth Science, Biology, Chemistry, Physics, Secondary Math I, Secondary Math II and Secondary Math III courses.

2.1 Longitudinal Data

Growth analyses on assessment data require data which are linked to individual students over time. Student growth percentile analyses require, at a minimum, two 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 is required in order to carry out subsequent growth analyses. The following business rules were used to select the appropriate record for use in the analyses.

2.1.1 Student Record Selection Business Rules

1. Student records with missing ("NA") scores or scale scores outside of the possible range are invalidated.
2. If a student has multiple records (duplicate from the same subject and grade), the score associated with the most advance course is selected.
3. Students who are not enrolled in a school for the full academic year (FAY) are omitted from the production of model coefficient matrices. Unlike previous years, these students do have SGPs created subsequently using the model coefficient matrices produced using FAY/accountability eligible data.
4. Non-standard participation or accommodation students are omitted.

Table 1 shows the number of valid grade level student records available for analysis[2] and Table 2 shows the available number of valid EOCT records.

[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 1: Number of Valid Grade Level Student Records by Grade and Subject for 2015

	Grades								
Content Area	3	4	5	6	7	8	9	10	11
ELA	48,206	46,913	46,839	46,121	44,415	43,577	42,167	40,560	37,438
Mathematics	48,500	47,139	47,103	46,169	43,617	43,487			
Science		47,141	47,149	46,332	44,783	44,033			

Table 2: Total Number of Valid EOCT Student Records by Subject for 2015

Content Area	
Earth Science	25,473
Biology	43,922
Chemistry	24,403
Physics	18,704
Sec Math I	44,955
Sec Math II	41,270
Sec Math III	28,302

3 Analytics

This section provides basic details about the calculation of student growth percentiles and percentile growth trajectories ('projections') from Utah state assessment data using the R Software Environment (R Development Core Team, 2015) in conjunction with the SGP Package (Damian W. Betebenner, VanIwaarden, Domingue, & Shang, 2015). More in depth treatment of the data analysis process with code examples is available to the USOE staff through Github.

Broadly, the SGP analysis of the Utah 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 cleaning and processing data provided by the USOE using R 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. In previous years' analyses the bulk of the data cleaning and implementation of business rule validation was also performed in R by NCIEA staff. However, the division of the cleaning and validation tasks used in 2015 is an important step in moving towards USOE self-sufficiency in calculating growth percentiles in subsequent years.

The cleaned and formatted data was combined with the existing data used up through the 2014 analyses. With longitudinal data prepared appropriately, we continued to the calculation of student-level SGPs.

3.2 2015 Data Analysis

The objective of the student growth percentile (SGP) analysis is to produce a measure which describes 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*. The estimation of this norm-referenced growth quantity is obtained using quantile regression (Koenker, 2005) to model curvilinear functional relationships between student's prior and current scores. One hundred such regression calculations are run for each separate analysis (defined as a unique year, content area, and grade 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 the calculation of thousands of regression models annually for Massachusetts. See Betebenner (2009) and Appendix B of this report for a more in-depth discussion of the SGP estimation methodology.

The 2015 Utah SGP analyses follow a work flow established in previous years that includes the following 4 steps:

1. Create annual SGP configurations for End-of-Grade Test (EOGT) and End-of-Course Test(EOCT) analyses.
2. Update the `SGPstateData` object in the `SGP` package. For 2015 this includes **a)** updating the norm group preferences, **b)** adding meta-data used in the calculation of student growth percentiles and projections, including the SAGE test knots and boundaries for the cubic basis splines and the test specific proficiency cutscores, and **c)** adding new SAGE related meta-data for the production of individual student reports (ISRs).
3. Conduct EOGT and EOCT SGP Analyses.
4. Export data, and produce summaries and visualizations from the `Utah_SGP` data object (including ISRs).

3.2.1 Create annual SGP configurations.

The EOCT analyses are specialized enough so that it is necessary to explicitly specify the analyses to be performed via a configuration code script. For several years, configurations have been employed to conduct EOCT SGP analyses for Utah. Unlike previous years, where EOGT analyses were run separately and using automatic configuration functionality, the 2015 EOGT analyses also used configuration scripts in order to run more efficiently.

Each configuration 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. Other parameters may also be defined. Each configuration used for the Utah EOGT analyses contain the first three elements. The EOCT analyses also contain the fourth and fifth elements:

Configurations are `R` code scripts that are used as part of the larger SGP analysis to be discussed later and to construct the norm group preference object discussed previously. They are broken up into three separate R scripts: one for grade level analyses (ELA, mathematics and science) and two for the EOCT content domains (mathematics and science). 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. Each configuration used for the end-of-grade analyses contain the first three elements. The EOCT analyses also contain the fourth and fifth elements:

- `sgp.content.areas`: A progression of values that specifies the 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`) provided in the configuration, potentially allowing for skipped years, repeated years, etc.
- `sgp.grade.sequences`: The grade progression associated with the content area and year progressions provided in the configuration. **'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/class-designation.
- `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, somewhat counter-intuitively, be left out or set to NULL, in which case projections will be produced and the package functions will populate the grade sequence to use based on the values provided in the `sgp.grade.sequences` element. 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 and stored in the `SGP` package here: `SGPstateData[["UT"]][["SGP_Configuration"]][["content_area.projection.sequence"]]`.

- `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 analyses. ***Lower numbers correspond with higher preference.***

The `sgp.content.areas`, `sgp.panel.years`, and `sgp.grade.sequences` elements all correspond to values found in the Utah data, and are used to select the subset of the longitudinal data set to be analyzed. Only these three elements are needed for the EOGT analyses because they automatically fall into canonical projection sequences and there will not be any duplicates produced given the data cleaning process.

As an example, here is one Biology configuration used to define a 2015 SGP analysis:

```
...

        BIOLOGY.2015 = list(
            sgp.content.areas =
                  c(rep('SCIENCE', 4), 'EARTH_SCIENCE', 'BIOLOGY'),
            sgp.panel.year s= as.character(2010:2015),
            sgp.grade.sequences = list(c(5:8, 'EOCT', 'EOCT')),
            sgp.projection.grade.sequences = NULL,
            sgp.norm.group.preference=1),

...
```

3.2.2 Update Utah assessment 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. The required updates for the 2015 analyses included **a)** updating the norm group preferences, **b)** adding meta-data used in the calculation of student growth percentiles and projections, including the SAGE test knots and boundaries for the cubic basis splines and the test specific proficiency cutscores, and **c)** adding new SAGE related meta-data for the production of individual student reports (ISRs).

Norm group preferences

The process through which EOCT analyses are run can produce multiple SGPs for some students. 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 look-up table and included in the `SGPstateData`. In general, the preference is 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).
- Progressions which fall into the "canonical" course progression, on which student growth projection estimates are produced and adequate growth judgement *potentially* made.

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) 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 Utah from the CRT and EOCT tests in previous years, and were also used in 2015 to compute the SAGE assessments' knots and boundaries using SAGE data from 2014 and 2015. New knots and boundaries will be required for Utah SAGE assessments beginning in 2015 as they are now used as the dependent variables in the quantile regressions.

SAGE proficiency level cutscores

Cutscores, which are set externally by the USOE through standard-setting processes, are mainly required for student growth projections. These growth projection estimates are used in the computation of adequate growth measures and elements of the ISRs. The SAGE cutscore data was provided to the Center for Assessment staff in 2014 and added at that time. This year is the first year in which those values were used in the analyses.

ISR meta-data

Finally, meta data for the ISR production was added. Mainly this entailed updating the `Assessment_Program_Information` and `Student_Report_Information` sections. The entire 2015 Utah entry of the `SGPstateData` can be viewed here.

3.2.3 Conduct SGP Analyses.

Unlike the 2014 analyses, we use the `updateSGP` function to A) do the final preparation and addition of the new long data to the existing `SGP` data object (`prepareSGP` step) and B)

[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 potentially 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).

produce SGPs for ***both*** the grade level and EOCT subjects (`analyzeSGP` step). Also, unlike any previous years' analyses, we produced SGPs for students who were non-continuously enrolled for the full academic year (FAY). Only the FAY students were used to construct the estimating models and their associated coefficient matrices. The non-FAY analyses utilize these model coefficient matrices to produce those students' SGPs. Therefore, there are two analyses being run with the data submitted this year by USOE (in strict order): **1)** 2015 FAY students and **2)** 2015 non-FAY students.

3.2.4 Merge 2015 results into the longitudinal data, and output, summarize and visualize data.

Once all analyses were completed the results were merged into the master longitudinal data set. A pipe delimited version of the complete long data is output and submitted to USOE. The data is also summarized using the `summarizeSGP` function, which produced many tables of descriptive statistics that are disaggregated at the state, district, school and other factors of interest. Finally, visualizations (such as bubble charts and ISRs) are produced from the data and summary tables.

4 Goodness of Fit

Despite the use of B-splines to accommodate heteroscedasticity and skewness of the scale score distributions, assumptions that are made in the statistical modeling process can impact how well the percentile curves fit the data. Examination of 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 of this report.

4.1 Model Fit Plots

Using all available EOGT 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 EOC subjects. A goodness of fit plot is produced for each unique analysis run in 2015 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 is the fit plot for Secondary Math III, and demonstrates minor model misfit.

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). If is is relatively typical for extremely high (low) achieving students to consistently score at or near the HOSS (LOSS) each year, the SGPs for these students may 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). 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 scale scores at both extremes yield the expected SGPs[4]. 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.

These plots are 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

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

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 estimated conditional density with the theoretical uniform density of the SGPs. 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[5]. 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).

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.

The results in all subjects are excellent with few exceptions. Deviations from perfect fit are indicated by red and blue shading. The further *above* 10 the darker the red, and the further *below* 10 the darker the blue. In instances where large deviations from 10 occur, the likely cause is that there is a mass point associated with certain 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 lower grades where fewer prior scores are available (particularly in the lowest grade when only a single prior is available). This is the case with all large deviations observed in the Utah data.

[5] The total students in each for the analyses varies depending on grade and subject.

Student Growth Percentile Goodness-of-Fit Descriptives

2015 ELA SGP, Grade 8 (N = 40,850)

Ceiling / Floor Effects Test

Current Observed Score Extremes (Range) × Student Growth Percentile 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:	[106,174)	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 = 25
	[174,187]	88.0	12.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	N = 25
Highest OSS:	[676,697)	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	12.0	88.0	N = 25
	[697,782]	0.0	0.0	0.0	0.0	0.0	0.0	0.0	4.0	0.0	0.0	4.0	92.0	N = 25

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/[113,324)	9.84	10.07	10.41	10.09	10.07	9.70	9.94	9.52	9.72	10.63	(9.9%)
2nd/[324,363)	9.74	10.61	9.74	9.14	9.74	10.26	9.41	10.26	10.16	10.94	(9.8%)
3rd/[363,391)	9.99	9.94	9.99	10.62	9.86	9.27	10.23	9.94	10.01	10.16	(10%)
4th/[391,414)	9.42	10.21	10.26	9.17	10.39	10.06	9.92	10.11	9.82	10.63	(9.9%)
5th/[414,436)	9.47	9.91	9.58	10.29	9.63	9.94	9.96	10.20	10.24	10.78	(10.4%)
6th/[436,456)	10.65	9.50	10.52	9.81	9.63	10.04	10.09	9.86	9.99	9.91	(9.6%)
7th/[456,478)	9.50	10.49	9.98	10.13	9.76	9.89	9.59	10.06	9.76	10.84	(10.1%)
8th/[478,502)	9.22	9.78	9.66	10.29	10.36	9.78	9.76	10.14	10.55	10.46	(10.1%)
9th/[502,533)	10.11	9.33	9.94	9.57	9.55	10.11	10.97	9.72	10.11	10.58	(10%)
10th/[533,745]	9.38	10.34	9.89	9.86	9.96	10.05	9.79	10.22	10.17	10.34	(10.3%)

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

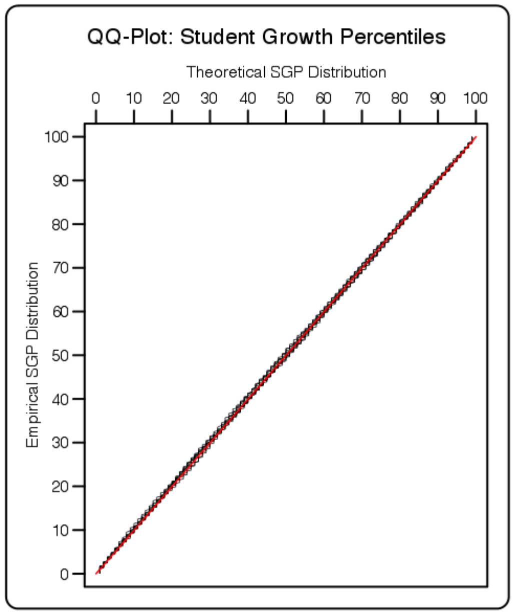

Figure 1: Goodness of Fit Plot for 2015 8^{th} Grade ELA: Example of good model fit.

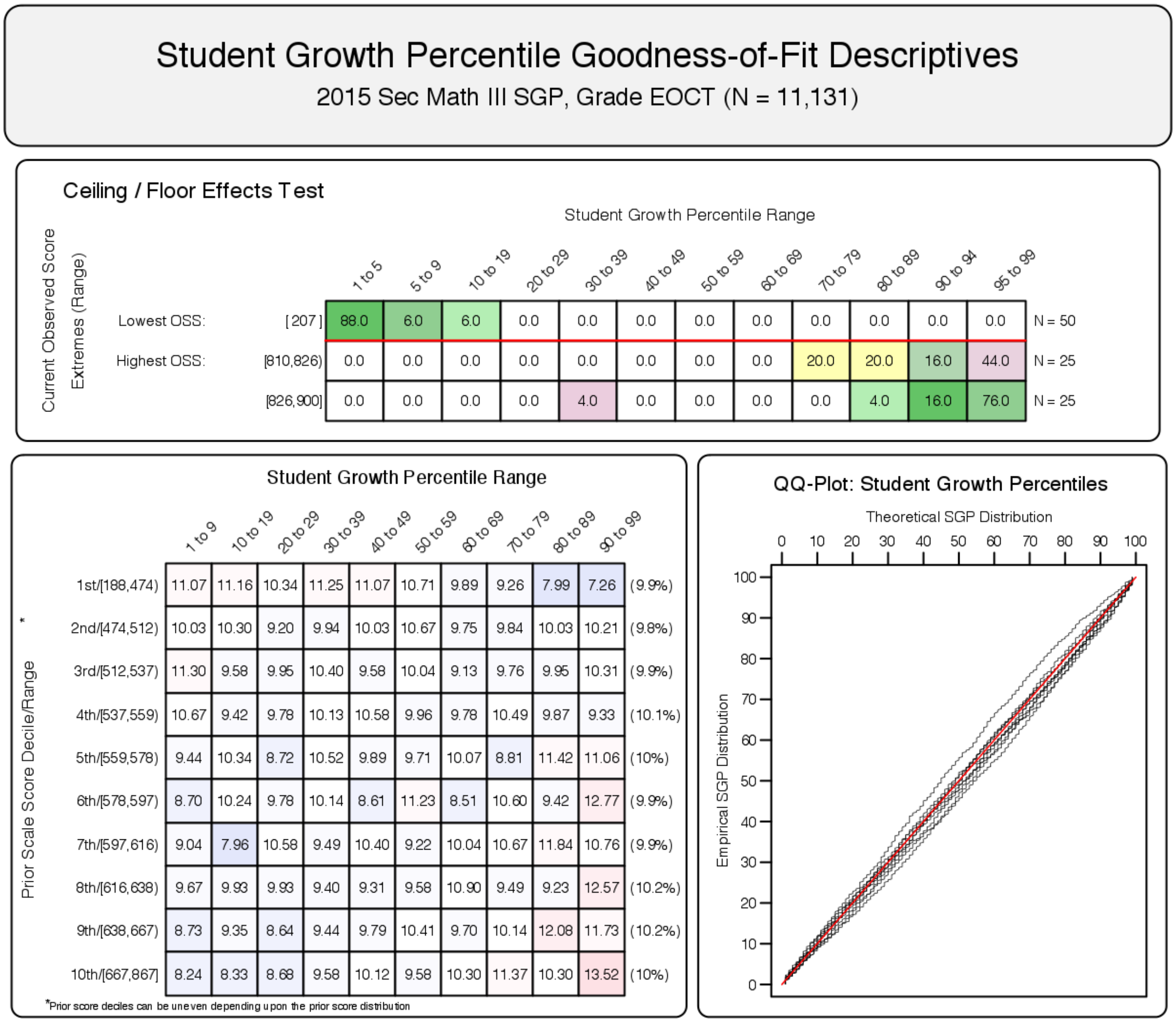

Figure 2: Goodness of Fit Plot for 2015 Secondary Math III: Example of slight model mis-fit.

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 student scale scores 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.

4.2.1 EOGT Subjects

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

Content Area	Grade	$r_{ScaleScore}$	r_{SGP}	N Size
ELA	4	0.82	0.00	44,158
	5	0.83	0.00	44,245
	6	0.83	0.00	43,635
	7	0.83	0.00	41,821
	8	0.85	0.00	40,850
	9	0.84	0.01	39,097
	10	0.83	0.01	37,430
	11	0.83	0.00	34,210
Mathematics	4	0.84	0.00	44,265
	5	0.85	0.00	44,370
	6	0.84	0.00	43,524
	7	0.83	0.00	40,604
	8	0.83	0.00	39,300
Science	5	0.78	0.00	44,421
	6	0.77	0.00	43,786
	7	0.79	0.00	41,900
	8	0.82	0.01	37,714

Student-level correlations for EOG 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 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 Colorado's uncorrected SGPs are all essentially zero (column 4).

4.2.2 EOCT Subjects

The EOCT correlations in Table 5 are disaggregated by the constituent norm groups. That is, each EOCT subject is potentially analyzed using more than one sequence of prior subjects, grades and years. The norm groups here are indicated by the most recent prior available for each norm group (although more prior years' scores are likely used). Although still strong[6], the correlations between these different content areas are slightly lower than those between the grade level subjects.

Table 4: EOC 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	$r_{ScaleScore}$	r_{SGP}	NA
Earth Science	2014 Science Grade 8	0.80	0.01	21,954
Biology	2014 Earth Science	0.77	0.01	19,002
	2014 Science Grade 8	0.80	0.00	14,775
Chemistry	2014 Biology	0.77	0.00	17,165
Physics	2014 Biology	0.76	0.00	6,401
	2014 Chemistry	0.74	0.01	6,936
Sec Math I	2014 Mathematics Grade 8	0.83	0.01	37,664
Sec Math II	2014 Sec Math I	0.79	0.03	35,079
Sec Math III	2014 Sec Math II	0.75	0.03	25,646

[6] As a rule of thumb, we hope to see correlations above 0.60

5 SGP Results

The following sections provide basic descriptive statistics from the 2015 analyses, including the state level mean and median growth percentiles. Currently Utah uses cohort referenced SGPs as the official student level growth metric. The interested reader can find more in depth discussions of the SGP methodology in Appendix B of this report and the available literature.

5.1 Median 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 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, Table 5 presents the cohort-referenced growth percentile medians and means for the EOGT content areas and 6 shows the EOCT subjects.

Table 5: Grade Level SAGE Median (Mean) Student Growth Percentile by Grade and Content Area for 2015

	Grades							
Content Area	4	5	6	7	8	9	10	11
ELA	50 (49.8)	50 (49.8)	50 (49.8)	50 (49.7)	50 (49.9)	49 (49.5)	50 (49.7)	50 (49.7)
Mathematics	50 (49.7)	50 (49.8)	50 (49.9)	50 (49.8)	49 (49.5)			
Science		50 (49.8)	50 (49.8)	50 (49.8)	50 (49.8)			

Table 6: EOCT SAGE Median and Mean Student Growth Percentile by Content Area for 2015

Content Area	Median SGP	Mean SGP
Earth Science	49	49.4
Biology	50	49.7
Chemistry	50	50.0
Physics	50	49.9
Sec Math I	50	49.7
Sec Math II	51	50.8
Sec Math III	50	50.2

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 Table 5 and 6 are close to 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 are presented in the Goodness-of-Fit section that follows. Depending upon feedback from the USOE, it may be desirable to tweak with 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 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 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 (McCaffrey, Han, & Lockwood, 2008).

5.2.1 School Level Results

To illustrate these relationships visually, the bubble charts in Figures 3, 4 and 5 depict growth as quantified by the median SGP of students at the school against achievement/status, quantified by percentage of student at/above proficient at the school[7]. 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 EOGT subjects are shown here.

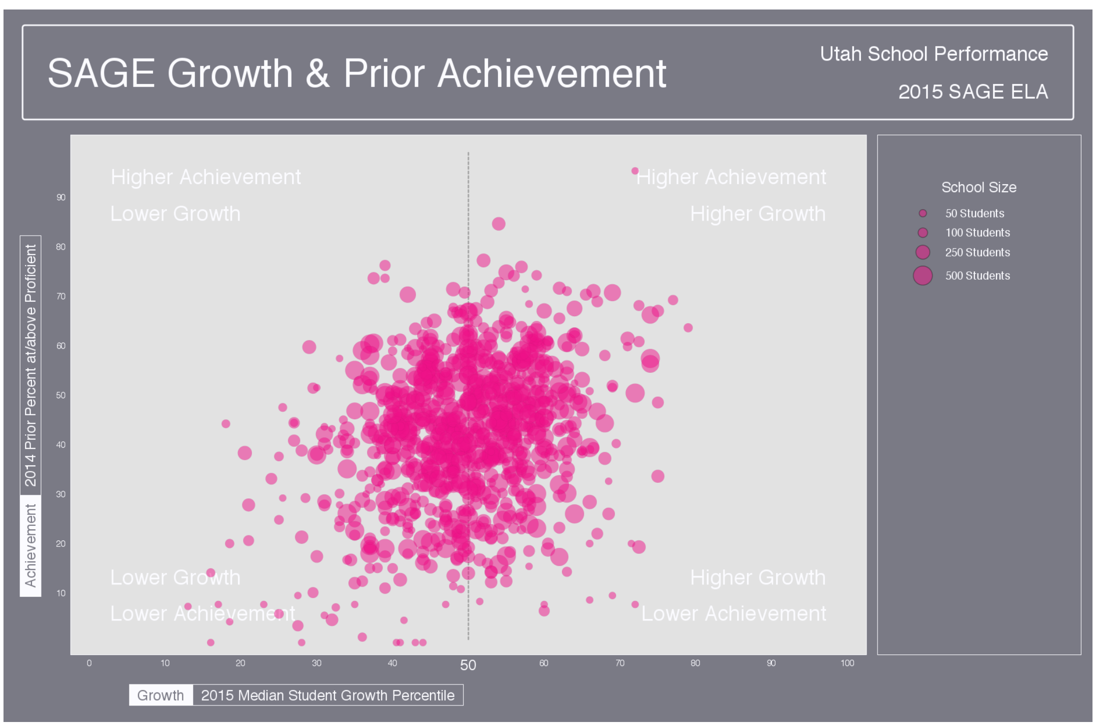

Figure 3: School Level Bubble Plots for Utah: ELA, 2014-2015.

[7]Percent Prior Proficient in this case is determined by the percent of student's who scored in the Proficient or Advanced range of the prior year's SAGE assessment out of all student's that received a score. This measure does not reflect student's that did not receive a score.

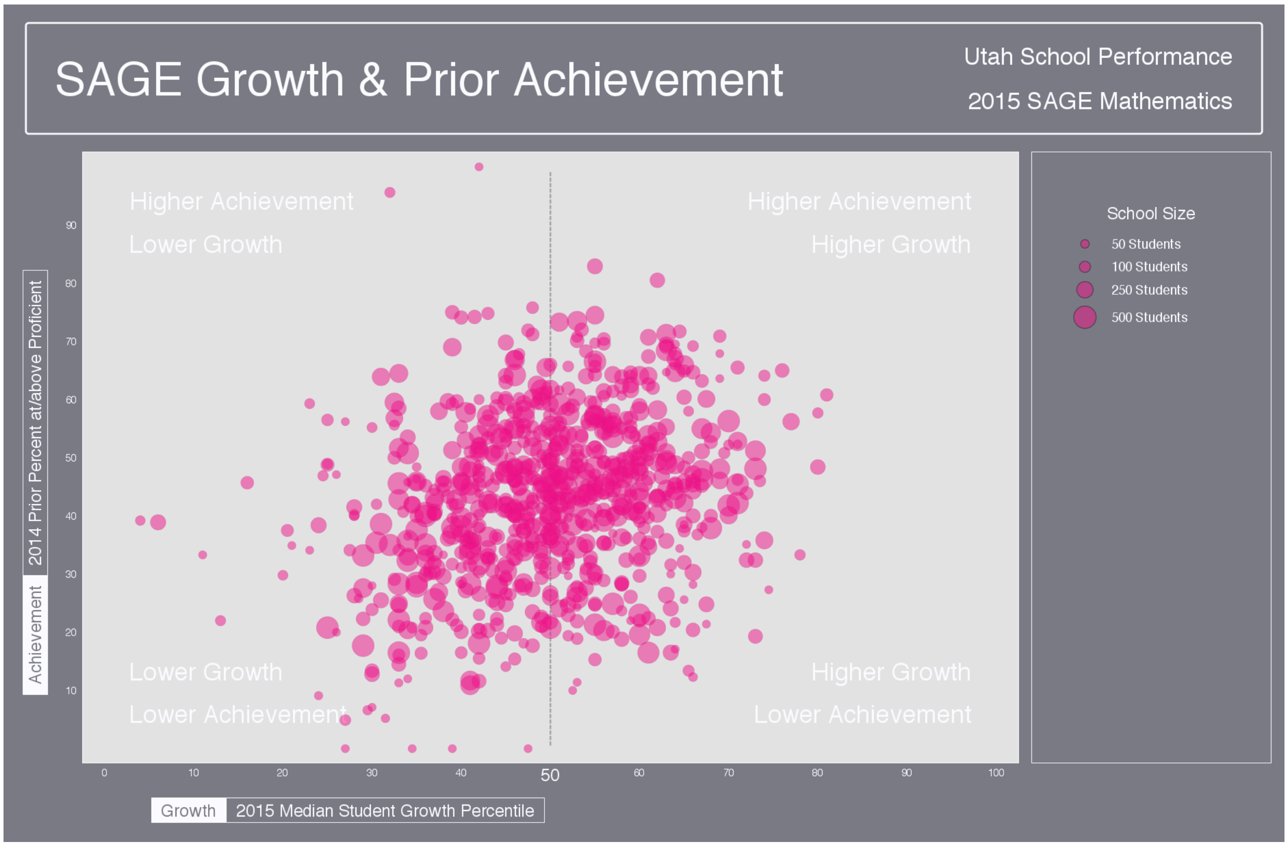

Figure 4: School Level Bubble Plots for Utah: Mathematics, 2014-2015.

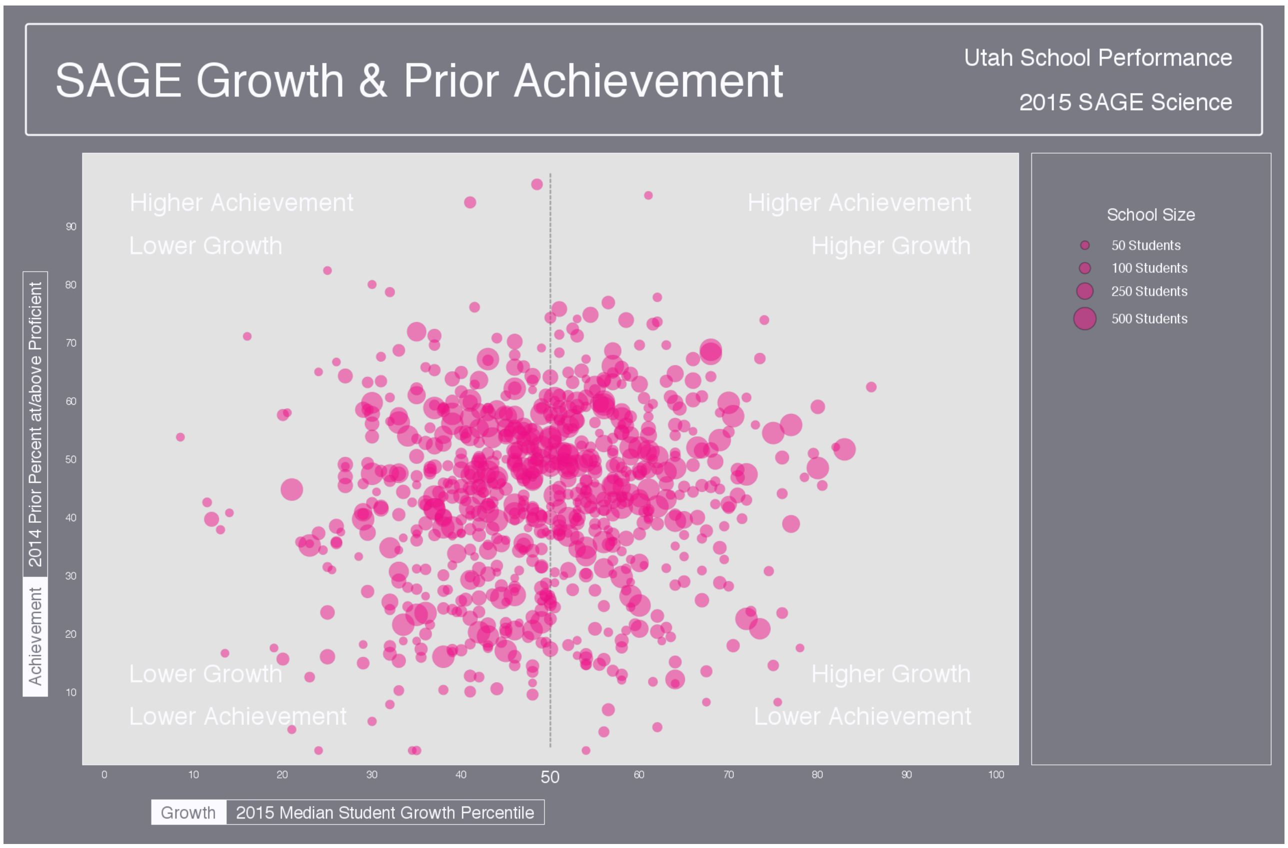

Figure 5: School Level Bubble Plots for Utah: Science, 2014-2015.

The relationship between average prior student achievement and median SGP observed for Utah is relatively strong compared to some other states for which the Center has done SGP analyses. Table 7 shows correlations between prior achievement (measured as the mean prior standardized scale score as well as the percent at/above proficient at the school[8]). All results shown here are for schools with 10 or more students.

Table 7: School Level Correlations between Mean Prior Standardized Scale Score and 1) Aggregate SGPs and 2) Percent Proficient and Above - (Combined Subjects)

Year	Median SGP	Mean SGP	Pct Proficient Or Above
2011	0.47	0.47	0.81
2012	0.45	0.45	0.80
2013	0.30	0.32	0.86
2014	0.41	0.41	0.89
2015	0.34	0.36	0.93

Correlation tables describing the relationship between prior achievement (here defined as mean prior standardized scale score) and aggregate growth percentiles are presented below in separate subsections for grade level and EOCT subjects. Additionally, the correlation between the groups prior achievement and a measure of their current achievement (here the percent of kids that are at or above the proficiency cuts). Typically these correlations are stronger than that between prior achievement and growth, which suggests that school achievement *status* tends to stay the same over time.

This is indeed what we see in the correlation tables. The first table in the each subsection provides these overall SGP aggregates' relationships with mean prior standardized scale scores. The additional correlation tables are dis-aggregated by content area, and content area and grade to provide more detail.

[8]Percent Prior Proficient in this case is determined by the percent of student's that scored in the Proficient or Advanced range of all student's that received a score. This measure does not reflect student's that did not receive a score but are included in the denominator of percent proficient.

Table 8: School Level EOGT Correlations between Mean Prior Standardized Scale Score and 1) Aggregate SGPs and 2) Percent Proficient and Above by Content Area.

Content Area	Year	Median SGP	Mean SGP	Pct Proficient Or Above
ELA	2013	0.36	0.40	0.82
	2014	0.40	0.42	0.87
	2015	0.31	0.32	0.93
Mathematics	2013	0.27	0.28	0.84
	2014	0.24	0.24	0.88
	2015	0.24	0.25	0.89
Science	2013	0.21	0.21	0.90
	2014	0.05	0.07	0.88
	2015	0.11	0.12	0.89

Table 9: 2013 to 2015 School Level EOCT Correlations between Mean Prior Standardized Scale Score and 1) Aggregate SGPs or 2) Percent Proficient and Above by Content Area.

Content Area	Year	Median SGP	Mean SGP	Pct Proficient Or Above
Earth Science	2013	0.11	0.13	0.86
	2014	-0.13	-0.13	0.80
	2015	-0.10	-0.10	0.80
Biology	2013	0.26	0.26	0.80
	2014	0.13	0.14	0.67
	2015	0.14	0.15	0.69
Chemistry	2013	0.15	0.17	0.70
	2014	0.08	0.09	0.56
	2015	0.23	0.25	0.82
Physics	2013	0.42	0.40	0.83
	2014	0.28	0.27	0.78
	2015	0.22	0.21	0.75
Sec Math I	2014	0.22	0.25	0.75
	2015	0.17	0.15	0.83
Sec Math II	2014	0.47	0.48	0.76
	2015	0.41	0.45	0.90
Sec Math III	2014	0.20	0.16	0.71
	2015	0.32	0.36	0.90

The final table disaggregates the 2015 correlations for the EOGT subjects further by grade level.

Table 10: 2015 School Level EOGT Correlations between Mean Prior Standardized Scale Score and 1) Aggregate SGPs or 2) Percent Proficient and Above by Grade.

Content Area	Grade	Median SGP	Mean SGP	Pct Proficient Or Above
ELA	4	0.04	0.05	0.83
	5	0.17	0.16	0.86
	6	0.13	0.14	0.86
	7	0.17	0.18	0.86
	8	0.13	0.14	0.86
	9	0.25	0.24	0.85
	10	0.18	0.13	0.88
	11	0.20	0.21	0.87
Mathematics	4	0.01	0.03	0.79
	5	0.10	0.11	0.83
	6	0.02	0.02	0.79
	7	0.20	0.20	0.85
	8	0.14	0.14	0.81
Science	5	0.06	0.06	0.82
	6	-0.02	-0.02	0.75
	7	0.12	0.13	0.74
	8	-0.03	-0.03	0.84

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. (2015). *SGP: An r package for the calculation and visualization of student growth percentiles & percentile growth trajectories.* Retrieved from https://github.com/CenterForAssessment/SGP

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. (2015). *R: A language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org