

# The Utah Student Growth Model

## A Technical Overview of the 2013-2014 Student Growth Percentile Calculations

Damian W. Betebenner

Adam R. VanIwaarden

*National Center for the Improvement of Educational Assessment (NCIEA)*

June 2015

### **Abstract**

DRAFT REPORT - DO NOT CITE! This report provides details about the Utah Student Growth Model methodology and presents a descriptive analysis of the 2014 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 for each student with longitudinal data in the state. 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.

The report includes four sections covering Data, Analytics, SGP Results, and Goodness of Fit:

- *Data* includes details on the decision rules used in the raw data preparation and student record validation.
- *Analytics* introduces some of the basic methodological concepts implemented.
- *SGP Results* provides basic descriptive statistics from the 2014 analyses.
- *Goodness of Fit* describes how well the statistical models used to produce SGPs fit Utah students' data. This includes discussion of goodness of fit plots and the student- and school level correlations between SGP and prior achievement.

Additionally, multiple appendices are included. Appendix A includes 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 a report on the transition from the CRT assessments to the Utah Student Assessment of Growth and Excellence (SAGE) that addresses concerns about reported anomalies in the 2014 SGP results.

# 2 Data

Data for the Utah Student Assessment of Growth and Excellence (SAGE) used in the SGP analyses were supplied by the Utah SOE to the NCIEA for analysis in the summer of 2014. The current longitudinal data set includes academic years 2007-2008 through 2013-2014. 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 2014 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.
4. Non-standard participation or accommodation students are omitted.

Table 1 shows the number of valid grade level student records available for analysis<sup>1</sup> and Table 2 shows the available number of valid EOCT records.

**Table 1:** Number of Valid Grade Level Student Records by Grade and Subject for 2014

Content Area	Grades								
	3	4	5	6	7	8	9	10	11
ELA	44,040	44,178	43,891	42,562	41,688	40,939	39,554	36,477	32,982
Mathematics	44,061	44,184	43,870	42,475	40,582	40,278			
Science		44,179	43,908	42,370	38,099	40,698			

<sup>1</sup>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:** Total Number of Valid EOCT Student Records by Subject for 2014

Content Area	
Earth Science	23,717
Biology	40,813
Chemistry	22,371
Physics	16,086
Sec Math I	41,102
Sec Math II	34,683
Sec Math III	16,079

### 3 Analytics

Please see the [Github Utah Repository Wiki](#) for information on the implementation of the open source software package **SGP** (Betebenner, VanIwaarden, Domingue, & Shang, 2014) developed for the **R** software environment (R Development Core Team, 2015) by the National Center for the Improvement of Educational Assessment.

## 4 SGP Results

The following sections provide basic descriptive statistics from the 2014 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.

### 4.1 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 3 presents the cohort-referenced growth percentile medians and means for the grade level content areas and 4 shows the EOCT subjects.

**Table 3:** Grade Level SAGE Median (Mean) Student Growth Percentile by Grade and Content Area for 2014

Content Area	Grades							
	4	5	6	7	8	9	10	11
ELA	50 (49.9)	50 (49.9)	50 (50.1)	50 (50)	50 (49.9)	50 (49.9)	50 (49.9)	50 (49.7)
Mathematics	50 (49.8)	50 (50)	50 (50)	50 (50.2)	50 (49.9)			
Science		50 (49.8)	50 (50)	50 (50)	50 (49.9)			

**Table 4:** EOCT SAGE Median and Mean Student Growth Percentile by Content Area for 2014

Content Area	Median SGP	Mean SGP
Earth Science	50	50.0
Biology	50	49.9
Chemistry	50	49.7
Physics	50	50.2
Sec Math I	51	50.8
Sec Math II	51	50.7
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 3 and 4 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 Utah SOE, 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<sup>th</sup> 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 Goodness of Fit

Examination of goodness-of-fit was conducted by comparing the estimated conditional density with the theoretical uniform density of the SGPs. Despite the use of B-splines to accommodate heteroscedasticity and skewness of the conditional density, assumptions are made concerning the number and position of spline knots that impact the percentile curves that are fit. 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 would be indicative of lack of model fit.

### 5.1 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). Percentages of student growth percentiles between the 10<sup>th</sup>, 20<sup>th</sup>, 30<sup>th</sup>, 40<sup>th</sup>, 50<sup>th</sup>, 60<sup>th</sup>, 70<sup>th</sup>, 80<sup>th</sup>, and 90<sup>th</sup> percentiles were calculated based upon the decile of the prior year's scaled score (the total students in each for the analyses varies depending on grade and subject). Results for the B-spline parameterizations for all analyses are given in Appendix A.

The results in all subjects are excellent with a 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. As an example, Figure 1 shows the results for 8<sup>th</sup> grade ELA.

## Student Growth Percentile Goodness-of-Fit Descriptives

2014 ELA SGP, Grade 8 (N = 39,653)

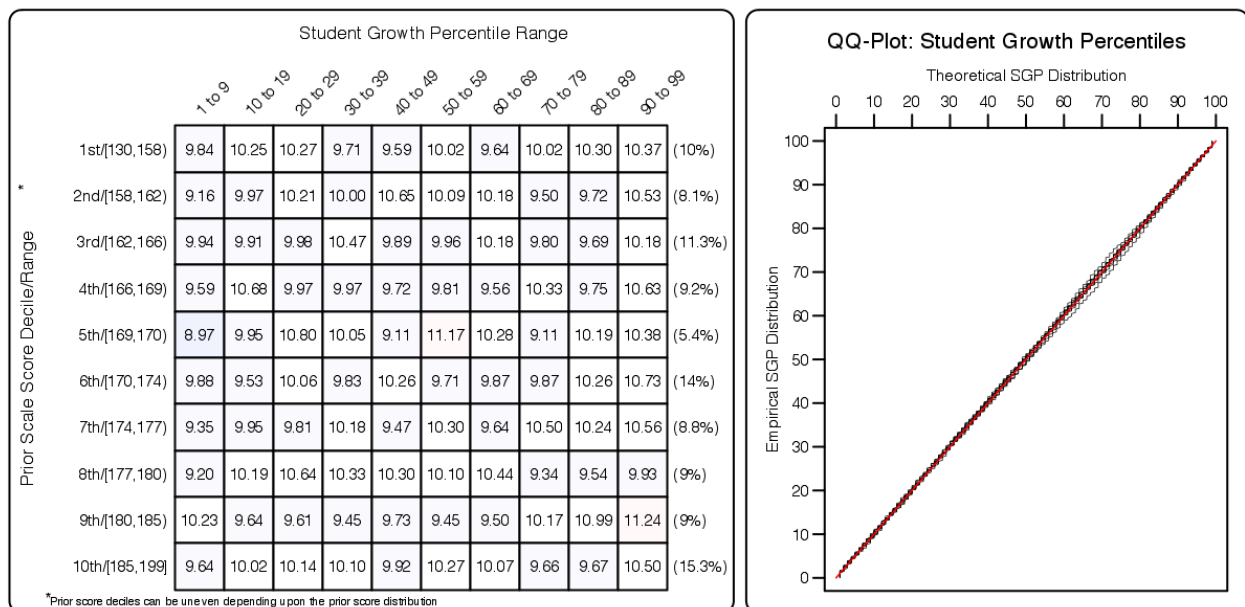


Figure 1: Goodness of Fit Plot for 2014 8<sup>th</sup> Grade ELA.



## 5.2 Student Level Results

To investigate the possibility that individual level misfit might impact summary level results, student growth percentile analyses were run on all students and the 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 prior student scale scores and student growth percentiles were calculated.<sup>2</sup>

Student level correlations between the SGP and prior achievement are presented here. The results are generally as expected. With cohort-referenced percentiles, when the model is perfectly fit to the data, the correlation between students' most recent prior achievement scores and their student growth percentiles is zero (i.e., there is a uniform distribution of percentiles across all previous scale score levels). Correlations for Utah cohort-referenced SGPs are all essentially zero. This provides assurance that the models have fit the data well, and indicate that students can demonstrate high (or low) growth regardless of prior achievement using cohort-referenced SGPs.

**Table 5:** EOGT Student Level Correlations between Prior Standardized Scale Score and SGP

Content Area	Grade	Correlation	N Size
ELA	4	0.00	42,288
	5	0.00	42,366
	6	0.00	41,105
	7	0.00	40,202
	8	0.00	39,653
	9	0.00	37,960
	10	0.01	35,078
	11	0.01	31,517
Mathematics	4	0.00	42,317
	5	0.00	42,354
	6	0.00	41,038
	7	0.01	39,100
	8	0.00	37,446
Science	5	0.00	42,572
	6	0.00	41,058
	7	0.00	36,818
	8	0.00	38,655

<sup>2</sup>In addition to providing information about model fit, these student level correlations can assess potential impact of test ceiling effects.

**Table 6:** EOCT Student Level Correlations between Prior Standardized Scale Score and SGP

Content Area	Correlation	N Size
Earth Science	0.01	21,493
Biology	0.01	32,848
Chemistry	0.01	15,599
Physics	0.01	11,091
Sec Math I	0.00	34,096
Sec Math II	0.04	30,462
Sec Math III	0.00	11,078

### 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 (McCaffrey, Han, & Lockwood, 2008).

#### 5.3.1 School Level Results

To illustrate these relationships visually, the bubble charts in Figures 2, 3 and 4 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<sup>3</sup>. 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.

<sup>3</sup>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.

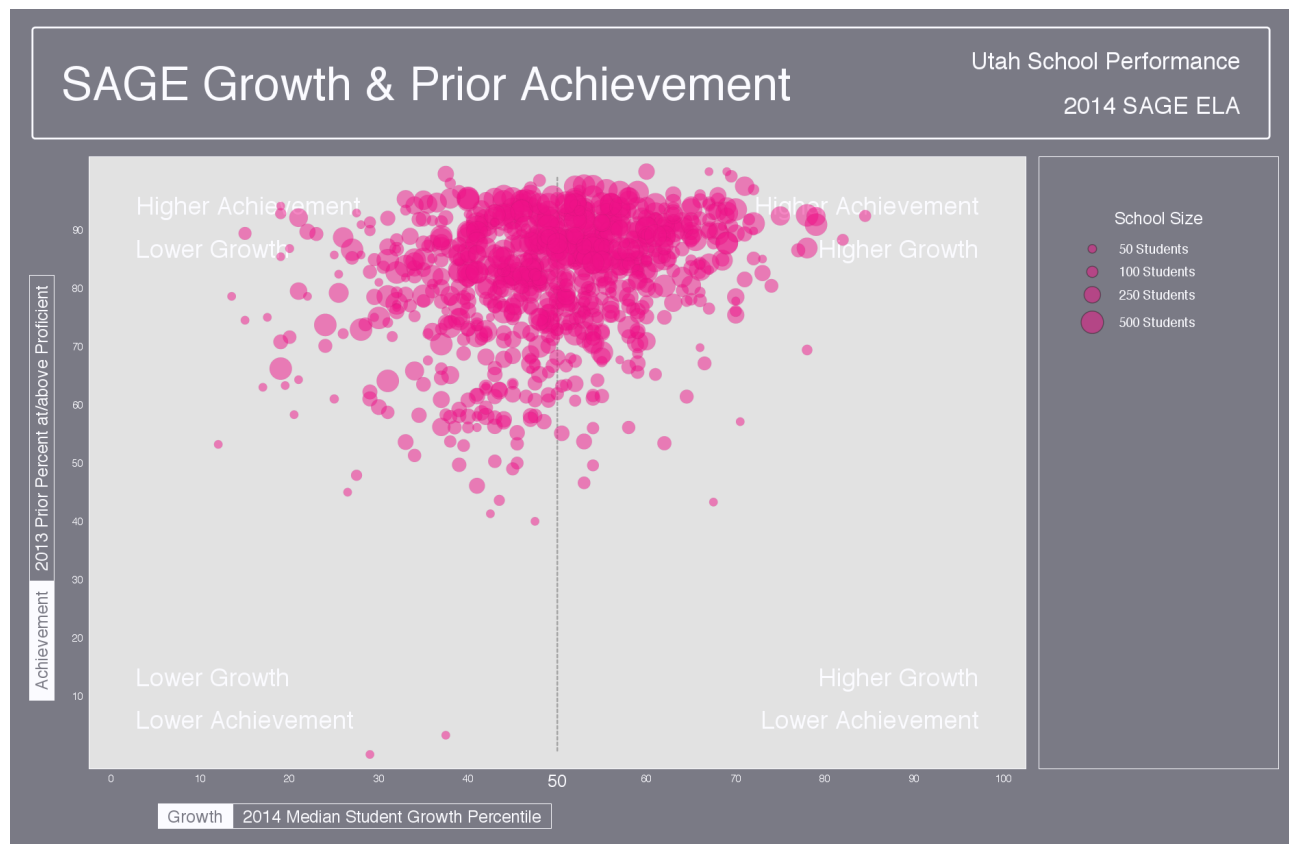


Figure 2: School Level Bubble Plots for Utah: ELA, 2013-2014.

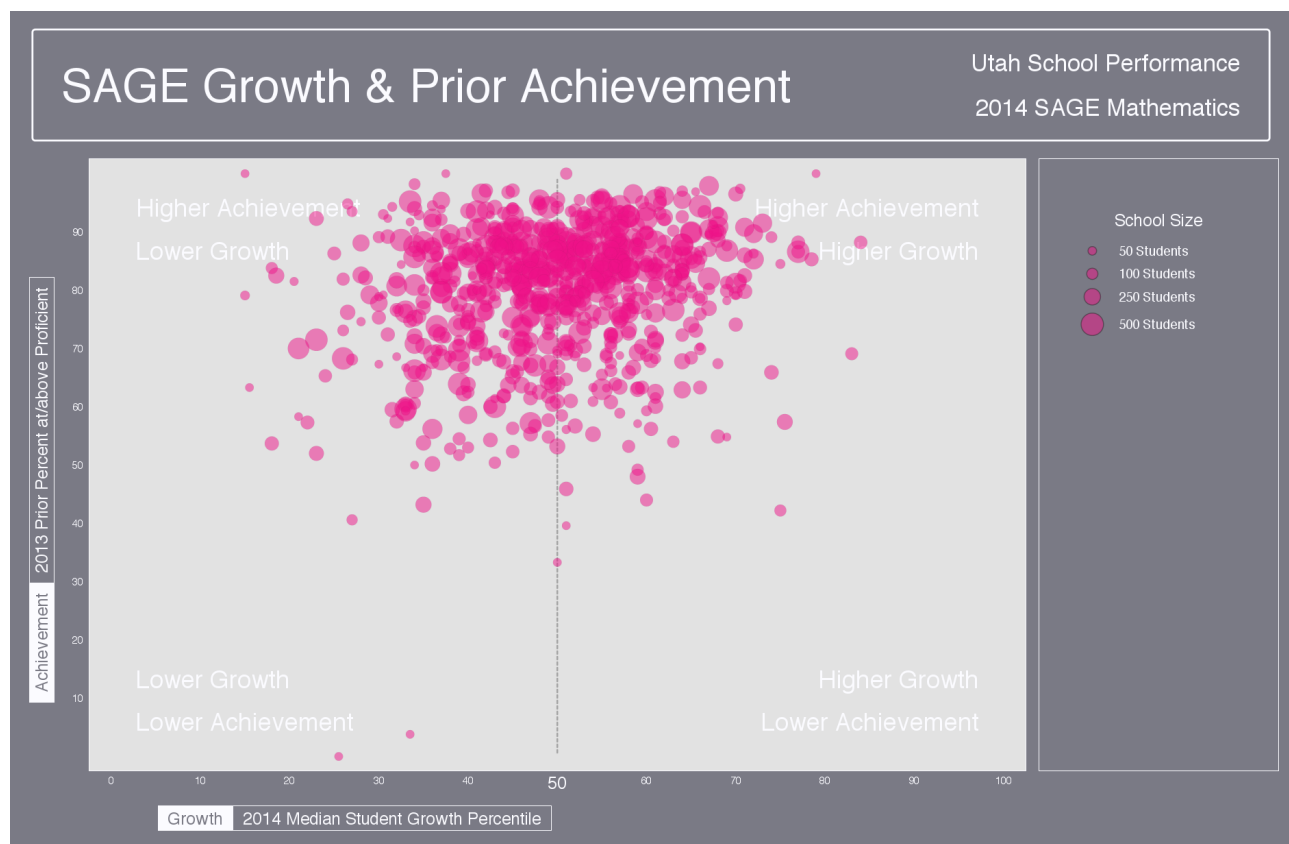


Figure 3: School Level Bubble Plots for Utah: Mathematics, 2013-2014.

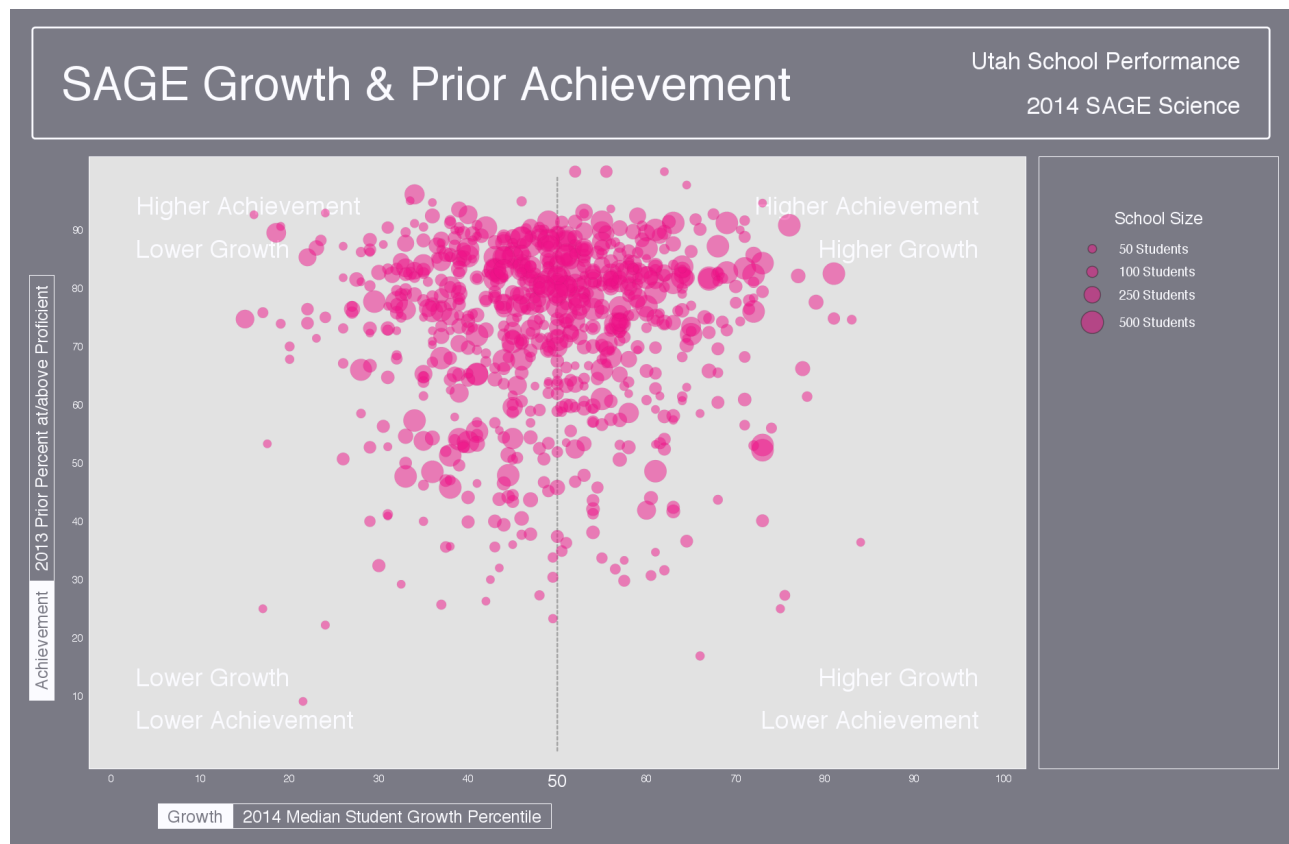


Figure 4: School Level Bubble Plots for Utah: Science, 2013-2014.

The relationship between average prior student achievement and median SGP observed for Georgia 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<sup>4</sup>). All results shown here are for schools with 10 or more students.

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

Year	Median SGP	Mean SGP
2011	0.47	0.47
2012	0.29	0.30
2013	0.30	0.32
2014	0.36	0.37

Correlation tables describing the relationship between prior achievement (again defined as mean prior standardized scale score) and aggregate growth percentiles are presented below in separate subsections for grade level and EOCT 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 dis-aggregated by content area, and content area and grade to provide more detail.

**Table 8:** School Level EOGT Correlations between Mean Prior Standardized Scale Score and SGP Aggregations.

Content Area	Year	Median SGP	Mean SGP
ELA	2012	0.25	0.27
	2013	0.36	0.40
	2014	0.37	0.38
Mathematics	2012	0.26	0.28
	2013	0.27	0.28
	2014	0.21	0.21
Science	2012	0.22	0.25
	2013	0.21	0.21
	2014	0.03	0.05

<sup>4</sup>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 9:** 2014 School Level EOGT Correlations between Mean Prior Standardized Scale Score and Aggregate SGPs by Grade.

Content Area	Grade	Median SGP	Mean SGP
ELA	4	0.22	0.23
	5	0.25	0.24
	6	0.15	0.16
	7	0.24	0.26
	8	0.24	0.25
	9	0.24	0.22
	10	0.20	0.23
	11	0.27	0.29
Mathematics	4	0.09	0.11
	5	0.02	0.04
	6	-0.05	-0.06
	7	0.18	0.22
	8	0.22	0.24
Science	5	-0.06	-0.05
	6	-0.11	-0.12
	7	0.14	0.14
	8	0.02	0.02

**Table 10:** 2012 to 2014 School Level EOCT Correlations between Mean Prior Standardized Scale Score and Aggregate SGPs.

Content Area	Year	Median SGP	Mean SGP
Earth Science	2012	0.03	0.02
	2013	0.11	0.13
	2014	-0.17	-0.18
Biology	2012	0.10	0.11
	2013	0.26	0.26
	2014	0.04	0.07
Chemistry	2012	-0.03	-0.07
	2013	0.15	0.17
	2014	0.07	0.08
Physics	2012	0.19	0.17
	2013	0.42	0.40
	2014	0.22	0.21
Sec Math I	2014	0.18	0.21
Sec Math II	2014	0.41	0.44
Sec Math III	2014	0.18	0.17



## References

Betebenner, D. W., VanIwaarden, A., Domingue, B., & Shang, Y. (2014). *SGP: An r package for the calculation and visualization of student growth percentiles & percentile growth trajectories*. Retrieved from <https://github.com/CenterForAssessment/SGP>

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 *National center for performance incentives conference performance incentives: Their growing impact on american k-12 education*, vanderbilt university, nashville, tN.

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>