

The Hawaii 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)

December 2015

Abstract

DRAFT REPORT - DO NOT CITE! This report provides details about the Hawaii 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 Hawaii. The National Center for the Improvement of Educational Assessment (NCIEA) contracted with the Hawaii State Office of Education (HIDOE) to implement the SGP methodology using data derived from the Smarter Balanced Assessment (SBA) program to create the Hawaii Student Growth Model. The goal of the engagement with HIDOE is to create a set of open source analytic techniques and conduct a set of initial analyses that will eventually be conducted by HIDOE 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 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 2015 analyses.
- *Goodness of Fit* describes how well the statistical models used to produce SGPs fit Hawaii 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 more detailed investigation of potential ceiling and/or floor effects present in Hawaii's student assessments and growth analyses.

2 Data

Data for the Hawaii Smarter Balanced Assessment (SBA) were supplied by the Hawaii SOE to the NCIEA for analysis in late summer 2015 for use in the SGP analyses. 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 HDOE to maintain a comprehensive longitudinal data set for all students taking the SBA assessments.

Student Growth Percentiles have been produced for students that have a current score and at least one prior score in Mathematics or Reading tests.

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. All valid case selection business rules were implemented by HDOE prior to submission of the 2015 student level data.

Table 1 shows the number of valid grade level student records available for analysis¹.

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

Content Area	Grades						
	3	4	5	6	7	8	11
Mathematics	14,812	14,596	14,397	13,727	13,066	12,063	10,215
Reading	14,755	14,586	14,368	13,679	13,082	12,092	10,257

¹This number does not represent the number of SGPs produced, however, because students are required to have at least one prior score available as well.

3 Analytics

This section provides basic details about the calculation of student growth percentiles and percentile growth trajectories (‘projections’) from Hawaii SBA 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 HIDOE the staff through Github.

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

1. Data Preparation
2. Data Analysis

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

3.1 Data Preparation

The data preparation step involves taking data provided by the HIDOE and producing a `.Rdata` file that will subsequently be analyzed in Step 2. This process is carried out annually as new data becomes available from the state assessment program. The data supplied by the HIDOE Information Technology department and subsequently cleaned and processed using R.

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 HIDOE 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 an appropriate longitudinal data prepared, 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 conducted 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 calculations for each of Hawaii’s annual analyses. For a more in-depth discussion of the calculation and estimation of SGPs, see Betebenner ((2009)) and Appendix B of this report for further information on the SGP methodology.

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

1. Update the `SGPstateData` object in the `SGP` package. For 2015 this included **a)** adding meta-data used in the calculation of student growth percentiles and projections, including the SBA test knots and boundaries for the cubic basis splines and the test specific proficiency cutscores, and **b)** adding new SBA related meta-data for the production of individual student reports (ISRs).
2. Conduct SGP Analyses.
3. Export data, and produce summaries and visualizations from the `Hawaii_SGP` data object (including ISRs).

3.2.1 Update Hawaii 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)** adding meta-data used in the calculation of student growth percentiles and projections, including the SBA test knots and boundaries for the cubic basis splines and the test specific proficiency cutscores, and **b)** adding new SBA related meta-data for the production of individual student reports (ISRs).

Calculation and addition of knots and boundaries

Calculation of SGPs includes the use of cubic B-spline basis functions to more adequately model the heteroscedasticity and non-linearity found in assessment data.² These functions require the selection of boundary and interior knots. Boundary knots are end-points outside of the scale score distribution that anchor the B-spline basis. These are generally selected by extending the entire range of scale scores by 10%. That is, they are defined as lying 10% below the lowest obtainable (or observed) scale score (LOSS) and 10% above the highest obtainable scale score (HOSS). The interior knots are the *internal* breakpoints that define the spline. The default choice in the `SGP` package (Damian W. Betebenner et al., 2015) is to select the 20th, 40th, 60th and 80th 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 Hawaii State Assessment (HSA) in previous years, and were also used in 2015 to compute the SBA assessments' knots and boundaries using SBA data from Hawaii and other SBA Consortium states. These new knots and boundaries were used in 2015 as part of the equated student growth projections, and will be required for Hawaii SBA assessments beginning in 2016 as they are now used as the dependent variables in the quantile regressions.

²It 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 students' prior score histories) 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).

SBA proficiency level cutscores

Cutscores, which are set externally by the SBA Consortium 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 SBA cutscore data was added to the `SGPstateData` by the Center for Assessment.

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 Hawaii entry of the `SGPstateData` can be [viewed here](#).

3.2.2 Conduct SGP Analyses.

As in previous years, 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) produce SGPs for ELA/Reading³ and Mathematics (`analyzeSGP` step).

Running this analysis will internally run the six steps of a typical SGP analysis, including the following:

1. `prepareSGP` Combines the 2015 LONG data file with the existing longitudinal LONG file in `*Hawaii_SGP@Data*` and run checks associated with `prepareSGP` on merge result.
2. `analyzeSGP` Performs 2015 SGP analyses including student growth percentiles, student growth projections, and lagged student growth projections.
3. `combineSGP` Merges results from 2015 analyses back into the master longitudinal record, `*Hawaii_SGP@Data*`, and calculates/creates scale scores associated with SGP targets.
4. `summarizeSGP` Saves previous summaries to hard disk, `*Hawaii_SGP@Summary*`, as `Hawaii_SGP_Summ` and calculates new summaries for the current year.
5. `visualizeSGP` Creates state level summary bubble plots, growth achievement plots (wave charts), and demonstration student growth plots.
6. `outputSGP` Exports data in multiple formats:
 - `Hawaii_SGP_Data_LONG.txt.zip` a pipe delimited copy of `*Hawaii_SGP@Data*`, the complete master longitudinal record.
 - `Hawaii_SGP_Data_LONG_2015.txt.zip` a pipe delimited copy of the 2015 longitudinal record subsetted from `*Hawaii_SGP@Data*`.
 - `Hawaii_SGPData_WIDE.txt.zip` a pipe delimited WIDE version of the master longitudinal record, `*Hawaii_SGP@Data*`.
 - `Hawaii_SGP_INSTRUCTOR_Data.txt.zip` a pipe delimited LONG version of the INSTRUCTOR_NUMBER file merged with the master longitudinal data record.

³Note that the official name of the SBA assessment is “English Language Arts/Literacy” (ELA), but is referred to here as “Reading” to maintain consistency with the HSA assessments.

4 SGP Results

The following sections provide basic descriptive statistics from the 2015 analyses, including the state level mean and median growth percentiles. Currently Hawaii 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 2 presents the cohort-referenced growth percentile medians and means by grade and content area.

Table 2: Grade Level SBA Median (Mean) Student Growth Percentile by Grade and Content Area for 2015

Content Area	Grades					
	4	5	6	7	8	11
Mathematics	50 (50)	51 (50.4)	52 (51.4)	51 (51)	51 (50.7)	51 (50.8)
Reading	50 (50)	50 (50.3)	51 (50.8)	51 (50.8)	51 (50.6)	50 (50.1)

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 1 are close to perfect, with almost all values equal to 50. XXX 6th grade Math??? XXX

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 Hawaii 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 50th 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 test 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). A goodness of fit plot is produced for each unique analysis run in 2015. Each analysis is defined by the grade and content area for the grade-level analyses.

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 it 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 to typical, and therefore SGP estimates will reflect typical growth (e.g. 50th percentile). In some cases small deviations from these extreme score values might even yield low growth estimates. Although these score patterns can legitimately be estimated as a low growth percentile, it is potentially an unfair description of student growth (and by proxy teacher or school, etc. performance or “value added”). Ultimately this is an artifact of the inability of the assessment 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 SGPs⁴. 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/LOSS⁵, 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 score⁶, 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 score⁷. In either case, a more

⁴Note that the prior year scale scores are not represented here, but are critical in the SGP calculation of all students

⁵This is particularly true when IRT Theta ($\hat{\theta}$) estimates are used rather than scaled scores, which often apply artificial LOSS/HOSS cut points.

⁶This can make the interpretation of the SGP distribution somewhat harder because current year scores not directly at the extremes do not necessarily preclude maximum SGP estimates.

⁷This also leads to potential difficulties in interpretation because with a higher number of students comes a greater distribution of prior scale scores and therefore a greater distribution of the expected SGPs

fine grained analysis of the relationship between score histories and SGPs and the associated potential for ceiling or floor effects in the models/model estimates may be necessary. These plots are meant to serve more as a “canary in the coal mine” than as a detailed, conclusive indicator of an effect.

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 10th, 20th, 30th, 40th, 50th, 60th, 70th, 80th, and 90th percentiles were calculated based upon the empirical decile of the cohort’s prior year scaled score distribution⁸. 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. A more detailed discussion about the reasons for and implications of model misfit for the various SGP analysis types can be found in the “Goodness of Fit” section of the 2015 Hawaii Student Growth Model report.

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.

As an example, Figure 1 shows the results for 4th grade ELA/Reading as an example of good model fit. The model fit also appears to be excellent in terms of lack of floor effects, although there is some evidence of a minor ceiling effect despite all other indicators suggesting good model fit. Figure 2 is the fit plot for 7th grade Mathematics, and demonstrates minor model misfit and evidence of both ceiling and floor effects.

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 Hawaii data.

⁸The total students in each for the analyses varies depending on grade and subject.

Student Growth Percentile Goodness-of-Fit Descriptives

2015 Reading SGP, Grade 4 (N = 13,596)

Ceiling / Floor Effects Test

		Student Growth Percentile Range													
Current Observed Score Extremes (Range)		1 to 5 5 to 9 10 to 19 20 to 29 30 to 39 40 to 49 50 to 59 60 to 69 70 to 79 80 to 89 90 to 94 95 to 98													
		Lowest OSS:	[2131,2208]	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	N = 24
			[2208,2230]	96.2	3.8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	N = 26
		Highest OSS:	[2663]	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	6.9	19.3	17.9	N = 145

		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/[195,267]	10.41	9.80	10.33	10.71	10.18	9.73	9.27	9.65	9.57	10.33	(9.7%)
		2nd/[267,282]	8.60	9.93	9.85	9.41	10.00	10.29	10.00	10.88	10.96	10.07	(10%)
		3rd/[282,294]	9.50	10.24	9.50	9.73	9.50	9.95	10.99	9.80	9.35	11.43	(9.9%)
		4th/[294,305]	9.02	10.21	9.99	10.58	10.66	10.28	9.61	9.46	10.95	9.24	(9.9%)
		5th/[305,314]	10.62	9.54	10.46	9.38	10.15	10.23	9.54	9.77	9.77	10.54	(9.6%)
		6th/[314,323]	9.10	10.30	10.52	9.85	9.70	9.48	11.64	10.07	8.73	10.60	(9.9%)
		7th/[323,333]	9.19	10.43	9.33	10.15	9.81	9.67	9.26	10.08	10.84	11.25	(10.7%)
		8th/[333,344]	9.93	8.90	9.93	10.53	10.97	10.30	9.56	9.86	10.67	9.34	(9.9%)
		9th/[344,360]	9.95	10.38	10.31	10.17	9.52	10.09	9.95	10.31	9.30	10.02	(10.2%)
		10th/[360,446]	9.45	9.88	9.81	9.52	9.95	9.81	10.38	9.59	10.45	11.17	(10.3%)

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

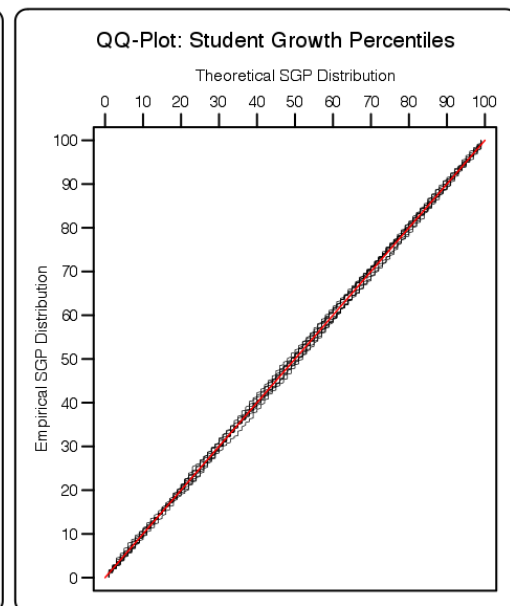


Figure 1: Goodness of Fit Plot for 2015 4th Grade Reading: Example of good model fit.

Student Growth Percentile Goodness-of-Fit Descriptives

2015 Mathematics SGP, Grade 7 (N = 11,928)

Ceiling / Floor Effects Test

		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		
Current Observed Score Extremes (Range)	Lowest OSS:	[2250]	75.0	9.4	10.4	2.1	1.0	1.0	1.0	0.0	0.0	0.0	0.0	N = 96	
	Highest OSS:	[2778]	0.0	0.0	0.0	0.0	0.0	0.0	3.2	6.5	9.7	24.2	21.0	N = 62	

		Student Growth Percentile 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		
Prior Scale Score Decile/Range *	1st/[137,256]	10.35	10.00	9.56	8.32	10.27	9.65	10.44	8.76	10.44	12.21	(9.5%)	
	2nd/[256,272]	8.52	9.25	10.48	10.57	10.32	10.24	10.16	9.01	10.32	11.14	(10.2%)	
	3rd/[272,284]	7.77	8.02	9.17	8.93	9.59	9.42	10.17	12.15	10.99	13.80	(10.1%)	
	4th/[284,294]	9.88	9.35	9.71	10.15	9.53	9.44	9.88	9.17	11.58	11.31	(9.4%)	
	5th/[294,305]	8.24	11.06	10.57	9.32	10.07	10.73	9.82	8.74	9.98	11.48	(10.1%)	
	6th/[305,316]	9.95	8.54	9.37	9.62	10.28	9.45	10.70	11.19	9.95	10.95	(10.1%)	
	7th/[316,328]	9.71	9.79	10.04	9.71	10.12	9.47	9.30	9.88	10.29	11.69	(10.2%)	
	8th/[328,343]	9.11	9.94	9.86	10.69	8.78	10.11	10.02	9.61	9.53	12.34	(10.1%)	
	9th/[343,365]	9.34	9.83	9.17	9.17	10.97	9.58	10.32	10.40	10.57	10.65	(10.2%)	
	10th/[365,480]	8.97	10.65	9.30	10.56	9.56	10.39	9.39	9.97	10.90	10.31	(10%)	

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

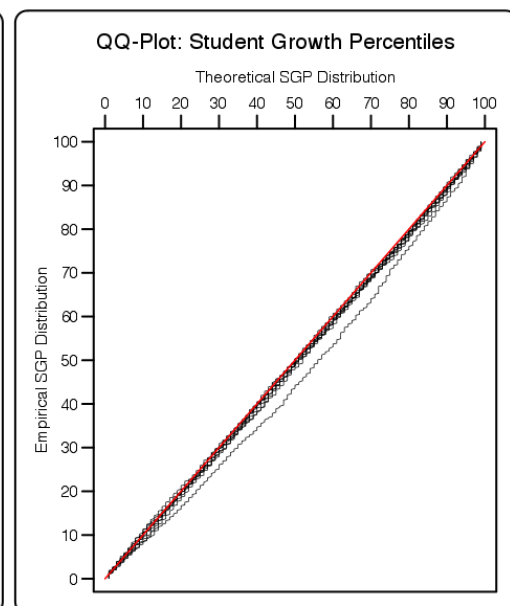


Figure 2: Goodness of Fit Plot for 2015 7th Grade Mathematics: Example of slight model misfit.

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.⁹ Additionally, the correlation between prior and current scale scores (achievement) are calculated to provide a contrast. For evidence of good model fit, the desired observed relationships would be no relationship (zero correlation) between prior achievement and growth, and a positive relationship between prior and current achievement, which suggests that growth is in fact detectable and growth modeling is therefore reasonable to begin with.

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 Hawaii 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. Furthermore, a strong relationship exists between prior and current scale scores.

⁹In addition to providing information about model fit, these student level correlations can assess potential impact of test ceiling effects.

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

Content Area	Grade	r_{SGP}	$r_{ScaleScore}$	N Size
Mathematics	4	0.00	0.84	13,538
	5	-0.01	0.83	13,329
	6	0.00	0.82	12,697
	7	-0.01	0.83	11,928
	8	-0.01	0.82	11,154
	11	-0.01	0.80	9,434
Reading	4	0.00	0.80	13,596
	5	-0.01	0.82	13,357
	6	0.00	0.80	12,702
	7	-0.01	0.78	11,985
	8	-0.01	0.79	11,212
	11	-0.01	0.74	9,460

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 3 and 4 depict growth as quantified by the median SGP of students at the school against achievement/status, quan-

tified by percentage of student at/above proficient at the school¹⁰. 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.

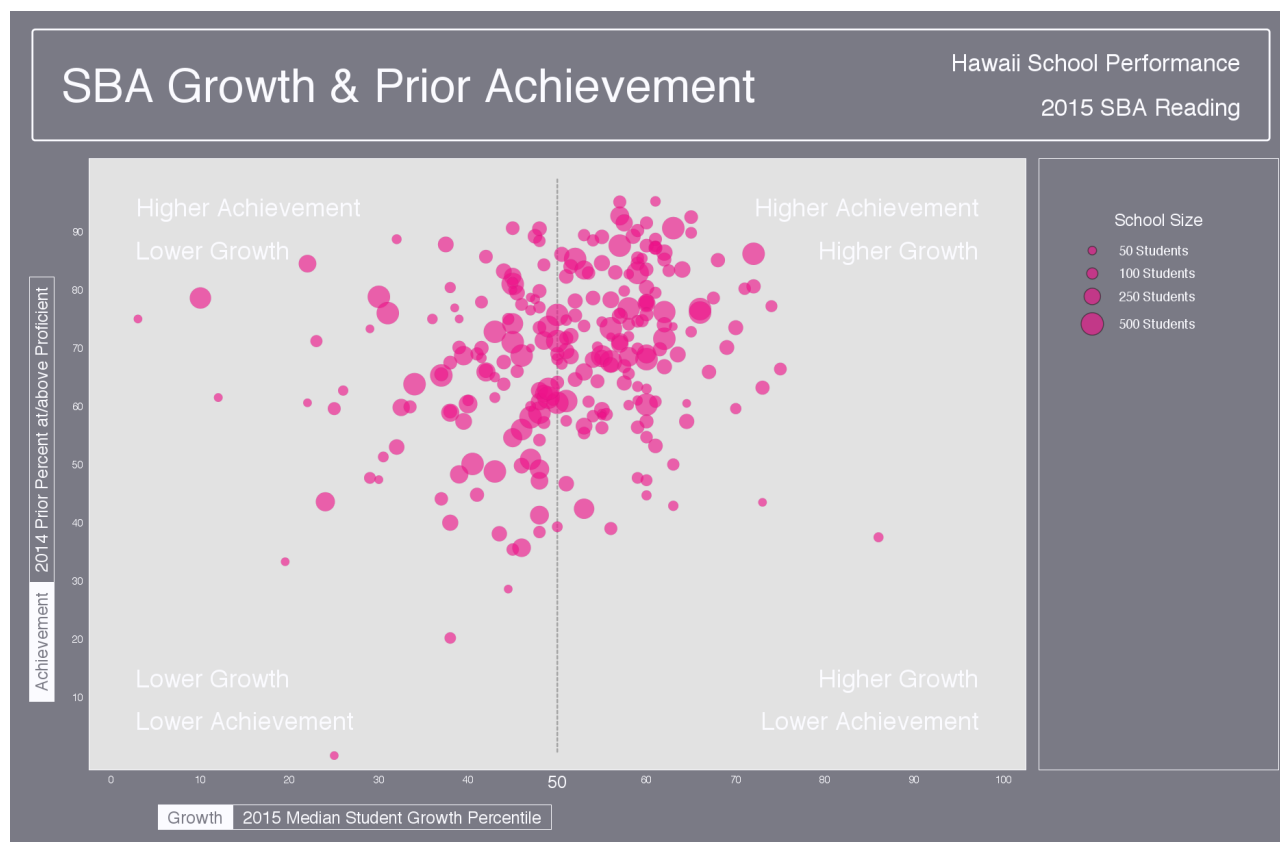


Figure 3: School Level Bubble Plots for Hawaii: Reading, 2014-2015.

The relationship between average prior student achievement and median SGP observed for Hawaii is typical to that in other states for which the Center has done SGP analyses. Table 4 shows correlations between prior achievement (measured as the mean prior standardized scale score as well as the percent at/above proficient at the school¹¹). All results shown here are for schools with 10 or more students.

¹⁰Percent 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.

¹¹Percent 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.

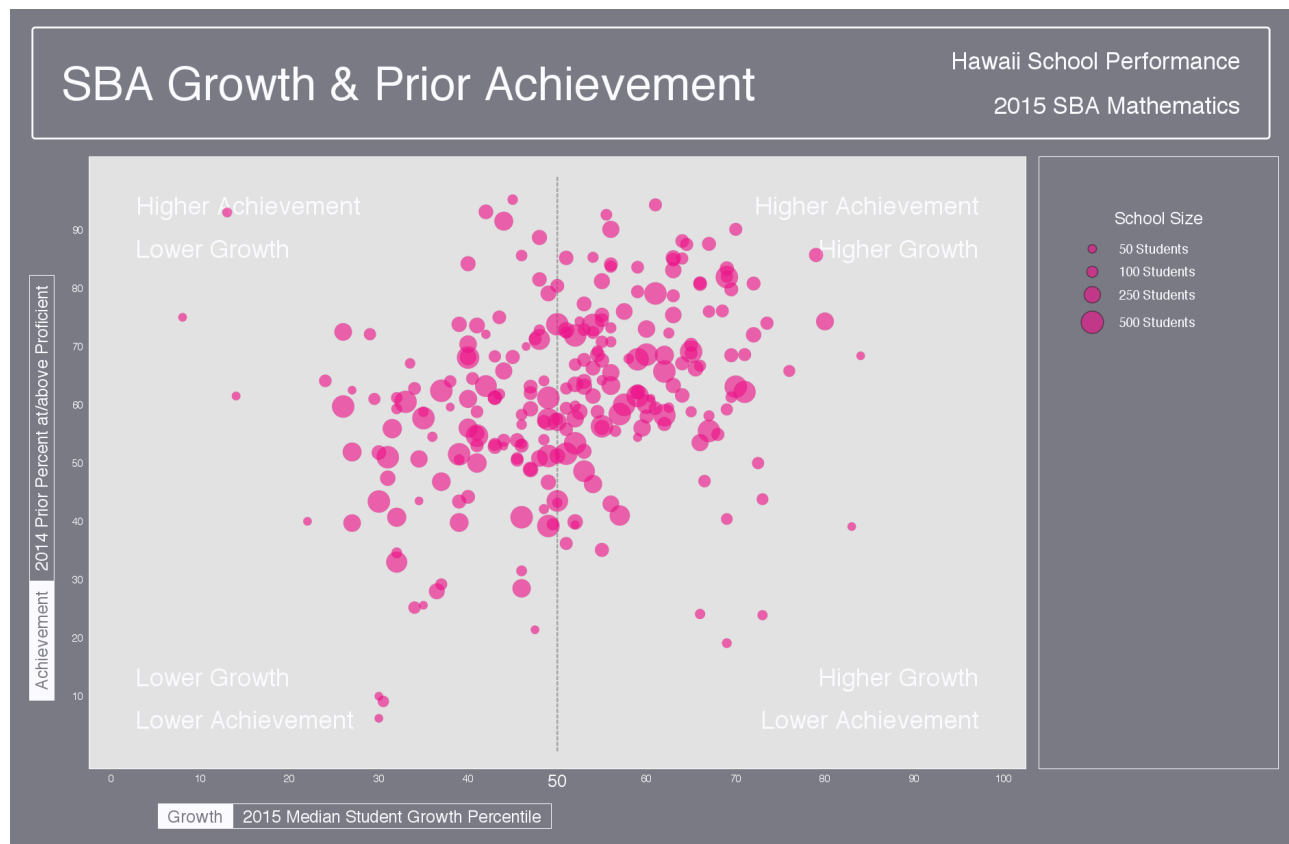


Figure 4: School Level Bubble Plots for Hawaii: Mathematics, 2014-2015.

Table 4: School Level Correlations between Mean Prior Standardized Scale Score and 1) Aggregate SGPs or 2) Aggregate Current Scale Score - (Combined Subjects)

Year	Median SGP	Mean SGP	Pct Proficient Or Above
2013	0.20	0.21	0.80
2014	0.16	0.18	0.81
2015	0.29	0.32	0.81

Correlation tables describing the relationship between aggregate growth percentiles and prior achievement (again defined as mean prior standardized scale score) are presented below in separate subsections for grade level and EOCT subjects. Additionally, the correlation between the growups 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.

Table 5: School Level EOCT Correlations between Mean Prior Standardized Scale Score and 1) Aggregate SGPs or 2) Aggregate Current Scale Score by Content Area.

Content Area	Year	Median SGP	Mean SGP	Pct Proficient Or Above
Mathematics	2013	0.18	0.21	0.78
	2014	0.15	0.18	0.78
	2015	0.24	0.27	0.82
Reading	2013	0.09	0.13	0.85
	2014	0.21	0.24	0.88
	2015	0.28	0.31	0.83

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

Table 6: 2015 School Level EOGT Correlations between Mean Prior Standardized Scale Score and 1) Aggregate SGPs or 2) Aggregate Current Scale Score by Grade.

Content Area	Grade	Median SGP	Mean SGP	Pct Proficient Or Above
Mathematics	4	0.12	0.13	0.84
	5	0.12	0.11	0.83
	6	0.18	0.20	0.77
	7	-0.09	-0.09	0.69
	8	0.31	0.31	0.77
	11	0.09	0.11	0.72
Reading	4	0.23	0.24	0.86
	5	0.13	0.13	0.88
	6	0.23	0.25	0.83
	7	0.15	0.13	0.76
	8	0.19	0.19	0.79
	11	0.12	0.11	0.70

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., Fernandez-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 *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>