GEORGIA STUDENT GROWTH MODEL FOR ENGLISH LANGUAGE PROFICIENCY

A TECHNICAL OVERVIEW OF THE 2023 WIDA ACCESS FOR ELLS STUDENT GROWTH PERCENTILE CALCULATIONS

JUNE 2023

DAMIAN W. BETEBENNER
ADAM R. VAN IWAARDEN

Center for Assessment

National Center for the Improvement
of Educational Assessment
Dover, NH

Center for Assessment

GEORGIA STUDENT GROWTH MODEL FOR ENGLISH LANGUAGE PROFICIENCY

A TECHNICAL OVERVIEW OF THE 2023 WIDA ACCESS FOR ELLS STUDENT GROWTH PERCENTILE CALCULATIONS

Submitted to:

Superintendent Richard Woods
Georgia Department of Education
June 2023

Author(s):

Damian W. Betebenner and
Adam R. Van Iwaarden

Acknowledgements:

We thank the staff of the GaDOE Office of Assessment & Accountability and English to Speakers of Other Languages (ESOL) program, particularly Adrienne Walker, Ph.D. and Margaret Baker, Ed.D., for their help in the growth analysis and student reporting processes, as well as their feedback and editing of this report.

Suggested Citation:

The National Center for the Improvement of Educational Assessment. 2023. Georgia Student Growth Model for English Language Proficiency. Submitted to Georgia Department of Education, Atlanta, GA

REPORT CONTENTS

TABLES AND FIGURES

INTRODUCTION

This report contains details on the 2023 implementation of the student growth percentile (SGP) model for the state of Georgia. The Center for Assessment (The Center) contracted with the Georgia Department of Education (GaDOE) to apply the SGP methodology using data derived from the WIDA Consortium (WIDA ACCESS for ELLs) assessments to create the Georgia Student Growth Model for English Language Proficiency (GSGM for ELP) . The goal of the engagement with GaDOE is to create a set of open source analytics techniques and conduct analyses, which may eventually be conducted by GaDOE staff in following years.

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

This report includes four sections:

- ***Data -*** describes the student level data requirements for inclusion in SGP analyses and provides valid record counts for the past three years.
- ***Analytics -*** includes details on the decision rules used in the raw data preparation and student record validation process and introduces some of the basic statistical methods and the computational process implemented in the 2023 analyses.[1].
- ***Goodness of Fit -*** investigates how well the statistical models used to produce SGPs fit Georgia students' data. This includes discussion of goodness of fit plots and the student-level correlations between SGP and prior English language proficiency.
- ***SGP Results -*** provides basic descriptive statistics from the 2023 analyses at both the state and school levels. The findings from these analysis augment model fit analyses and provide information that supports the interpretation and use of SGPs.

Multiple appendices to the report are provided. Appendix A, *SGP Model Fit Plots*, provides the model goodness of fit plots for all 2023 SGP analyses. Appendix B, *SGP Methodology*, describes the motivation, concept and approach of SGP methods, and technical aspects of SGP calculation are covered in detail.

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

DATA

GaDOE supplied the WIDA ACCESS for ELLs data used in the SGP analyses to The Center for Assessment in the summer of 2023. These test records were added to existing Georgia WIDA ACCESS for ELLs assessment data to create the longitudinal data set from which the 2023 SGPs were calculated. Subsequent years' analyses will augment this multi-year data set allowing GaDOE to maintain comprehensive longitudinal data for all students taking the WIDA ACCESS for ELLs assessments.

For the 2023 academic year, Student Growth Percentiles were produced by grade level for students that have a current Overall Composite scale score and at least one prior Overall Composite scale score.

LONGITUDINAL DATA

Growth analyses on assessment data require data that are linked to individual students over time. Student growth percentile analyses require a minimum of two (but preferably three or more) years of assessment data for analysis of student progress. To this end, it is necessary that a unique student identifier be available so that student data records across years can be merged with one another and subsequently examined.

Because some records in the assessment data set may contain students with more than one test score in a given year, a data cleaning process is required to create unique student records in each year in order to carry out subsequent growth analyses. Furthermore, student records may be excluded from the growth analyses for other reasons. See the Data Preparation section for details on the business rules used in this process for 2023.

Table 1 shows the number of valid student records available in the past three years after applying the data preparation business rules. Note that these counts do not represent the number of SGPs produced in any year, however, because students are required to have at least one prior score available as well.

Table 1: Valid Overall Composite Student Records by Grade for 2023, 2022 and 2021 (Elementary Grades)

	Grades					
Year	**K**	**1**	**2**	**3**	**4**	**5**
2023	15,264	16,078	15,309	15,281	14,873	11,302
2022	14,959	14,597	15,150	15,047	14,454	12,230
2021	12,727	13,719	13,996	13,773	13,386	10,249

Table 2: Valid Overall Composite Student Records by Grade for 2023, 2022 and 2021 (Middle/High Grades)

	Grades						
Year	**6**	**7**	**8**	**9**	**10**	**11**	**12**
2023	9,293	10,182	9,549	10,321	7,222	4,576	3,459
2022	9,525	9,040	8,334	8,973	5,570	4,086	3,168
2021	7,594	6,858	5,682	5,304	4,204	2,918	2,227

ANALYTICS

This section provides basic details about the calculation of student growth percentiles from assessment data using the R Software Environment (R Core Team 2023) in conjunction with the SGP package (Damian W. Betebenner et al. 2023).

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

1. Data Preparation
2. Data Analysis

The majority 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.

DATA PREPARATION

The data preparation step involves taking data provided by the GaDOE and producing a `.Rdata` file that will subsequently be analyzed using the SGP software. This process is carried out annually as new data becomes available from the WIDA ACCESS for ELLs assessment program.

For the 2023 Georgia WIDA ACCESS for ELLs data preparation and cleaning, we first subset the raw data to include only the variables that are relevant to the SGP analyses. These variables were then renamed to conform to the SGP package conventions.

Invalid records were identified based on the following criteria:

- Cases with incorrect student ID characteristics (e.g., fewer than 10 characters)
- Students with duplicate records. In these instances, a student's highest scale score is retained as the "valid" case in the analyses.

DATA ANALYSIS

The objective of the student growth percentile (SGP) analysis is to describe how (a)typical a student's growth is by examining their current English language proficiency relative to students with similar test score histories; i.e their *academic peers*.[2] This norm-referenced growth quantity is estimated using quantile regression (Koenker 2005) to model curvilinear functional relationships between students' prior and current scores. One hundred such regression models are calculated for each separate analysis (defined as a unique ***year*** *by* ***grade*** *by* ***prior order*** combination). The end product of these 100 separate regression models is a single coefficient matrix, which serves as a look-up table to relate prior student English language proficiency to current proficiency for each percentile. This

[2] See this presentation for a description of academic peers.

process ultimately leads to thousands of model calculations during each of Georgia's annual round of WIDA ACCESS for ELLs analyses. For a more in-depth discussion of SGP calculation, see Betebenner (2009) and B of this report.

SGP analyses follow a work flow established that includes the following steps:

1. Update the Georgia WIDA ACCESS for ELLs assessment meta-data required for SGP calculations using the `SGP` package.
2. Conduct all SGP analyses.
3. Combine results into the master longitudinal data set and output data.

CREATE GEORGIA WIDA ACCESS FOR ELLS 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 2023 WIDA ACCESS for ELLs SGP growth analyses utilized the `SGPstateData` information established for Georgia in 2022. Although the Georgia WIDA ACCESS for ELLs meta-data is based in part on the WIDA Consortium and member states' meta-data, extensive customization was added for the analyses and rendering of individual student reports for growth and English language proficiency. Some of the more important elements are detailed below.

Knots and boundaries

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

In general the knots and boundaries are computed from a distribution comprised of several years of test data (i.e. multiple cohorts combined) so that any irregularities in a single year are smoothed out. This is important because subsequent annual analyses use these same knots and boundaries.

The knots and boundaries used in the Georgia WIDA ACCESS for ELLs analyses are a standard set that were constructed from multiple WIDA Consortium members' data using the default knot locations described above. These values are used by the majority of WIDA member states with which the Center works.

Proficiency level cutscores

Cutscores for the WIDA ACCESS for ELLs assessment system are set by WIDA. Details on the proficiency levels and standard-setting process can be found in their online resources.

Although WIDA establishes the proficiency levels and cutscores, Georgia, like most other states, selects the levels at which specific proficiency criteria are considered to have been met. For example, Georgia has identified Proficiency Level 4.3 as the minimum proficiency level for schools and/or school systems/charter schools to allow students to exit their English Learner (EL) programs.

In regards to the SGP analyses, local proficiency level criteria are mainly required for reporting additional growth metrics such as student growth projections and growth targets.

Student report configurations

The `SGPstateData` also houses information used to create individual student reports that were customized to meet Georgia's design and use requirements.

CONDUCT SGP ANALYSES

Georgia currently uses cohort-referenced SGPs as the official student-level English language proficiency growth metric. All SGPs were calculated concurrently using the `R` Software Environment in conjunction with the `SGP` package. Broadly, the Georgia WIDA ACCESS for ELLs analyses were completed in 5 steps.

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

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

We use the `updateSGP` function to ***a)*** do the final preparation and addition of the cleaned and formatted new annual data, (`prepareSGP` step), ***b)*** calculate SGP estimates (`analyzeSGP` step), ***c)*** merge the results into the master longitudinal data set (`combineSGP` step) and ***d)*** output a pipe delimited version of the complete long data (`outputSGP` step).

Visualize results

Once all analyses were completed via `updateSGP`, individual student growth and English language proficiency reports were produced using the `visualizeSGP` function and a custom template

designed for Georgia. English and Spanish language versions of these reports were created, and individual reports and school level catalogs were bundled according to Georgia's specifications.

Custom data formatting and district output

The 2023 WIDA ACCESS for ELLs SGP results data were submitted to GaDOE with additional formatting to add fields including students' prior language proficiency level.

System level student data sets with fields used to create individual student reports were also submitted to GaDOE in Microsoft Excel format.

GOODNESS OF FIT

Assessment data are generally imperfect and require sophisticated statistical methods to deal with the various issues they present. Cubic B-spline basis functions are used in the calculation of SGPs to more adequately model issues such as heteroscedasticity, non-linearity, skewness and ceilings/floors in the data. Assumptions that are made in the statistical modeling process can impact how well the percentile curves fit the data.[3] Accordingly, a thorough examination of the fit of the SGP model to the assessment data was performed.

Examination of the Georgia Student Growth Model for English Language Proficiency 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 English language proficiency. In the next section we discuss the model fit plots in general and provide examples from the 2023 analyses. We then provide tables of student level correlation results as further evidence of model goodness of fit.

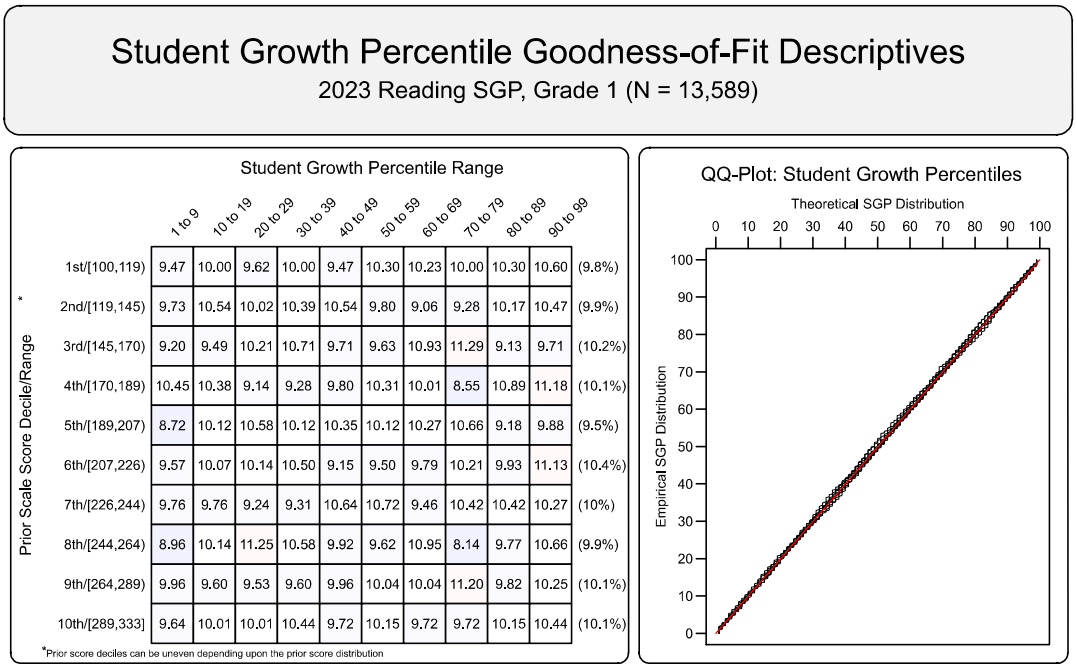
MODEL FIT PLOTS

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). Each analysis is defined by the grade and content area for the grade-level analyses. A goodness of fit plot is produced for each unique analysis run in 2023 and the complete portfolio of each fit plot is provided in Appendix A of this report.

The fit plots for the 2023 WIDA ACCESS for ELLs SGP analyses are excellent with few exceptions. Figure 1 displays the 1st grade model as an exemplar of model fit.

[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* estimated growth percentiles for *higher* observed scale scores at the extremes (given 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).

Figure 1: Goodness of Fit Plot for 2023 1st grade : Example of good model fit.



The two panels compare the observed conditional density of the SGP estimates with the theoretical (uniform) density. The bottom left panel is a 10 by 10 cell grid that shows the empirical distribution of SGPs given prior scale score deciles. The cells display the percentages of SGPs between the 10th, 20th, 30th, 40th, 50th, 60th, 70th, 80th, and 90th percentiles for each of the ten empirical decile groups based upon the cohort's prior year scaled score distribution[4]. With an infinite population of test takers and perfect model fit, the expectation would be to have 10 percent of the estimated growth percentiles in each cell. Deviations from 10 percent, indicated by red and blue shading, suggests lack of model fit. Cells further *above* 10% are shaded darker red, and those further *below* 10% are shaded darker blue.

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.

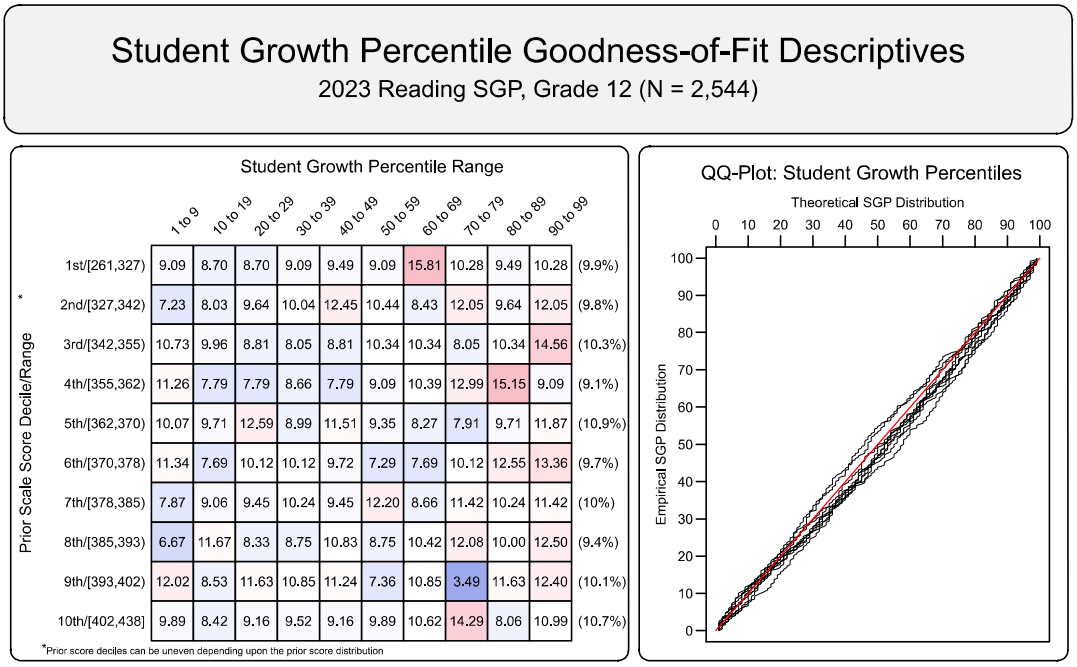
Misfit in SGP models can be caused by several factors, such as the use of a single prior in the model, relatively small cohort size (e.g. fewer than 2,000 students) and non-representative or homogeneous cohorts (e.g., all advanced learner cohorts). The magnitude of misfit these factors cause is usually minor, often resulting in clustering of SGPs that is visible in some conditional distribution grid cells

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

(dark red cells adjacent to dark blue cells). The issue common to these factors is that the norm group data does not provide sufficient information to differentiate between students.

Figure 3 is from the 12th grade model, which exemplifies the extent to which model misfit is present in the 2023 Georgia WIDA ACCESS for ELLs SGP analyses.

Figure 2: Goodness of Fit Plot for 2023 12th grade : Example of model misfit.



STUDENT LEVEL GROWTH AND PRIOR ENGLISH LANGUAGE PROFICIENCY

To investigate the possibility that individual level misfit might impact summary level results, student level SGP results were examined relative to prior English language proficiency. With perfect fit to data, the correlation between students' most recent prior test score 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 and **b)** prior score and student growth percentiles were calculated.

Evidence of good model fit begins with a strong positive relationship between prior and current test scores, which suggests that growth is detectable and modeling it is reasonable to begin with. Conversely, a lack of relationship (zero correlation) between prior status and growth confirms that the model has fit the data well and produced a uniform distribution of percentiles across all previous scale score levels. This indicates that students can demonstrate high (or low) growth regardless of prior English language proficiency.

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

Grade	$r_{TestScore}$	r_{SGP}	N Size	Calculation Rate
1	0.71	0.00	13,589	84.5%
2	0.77	-0.01	13,086	85.5%
3	0.79	-0.01	13,396	87.7%
4	0.78	-0.02	13,031	87.6%
5	0.79	-0.02	9,686	85.7%
6	0.72	-0.02	7,775	83.7%
7	0.79	0.00	8,492	83.4%
8	0.82	-0.01	7,900	82.7%
9	0.79	-0.01	6,698	64.9%
10	0.81	0.00	5,376	74.4%
11	0.81	-0.03	3,304	72.2%
12	0.77	-0.01	2,544	73.5%

Student-level correlations presented in Table 3 are generally as expected. Strong positive relationships exist between prior and current test scores for the grade level analyses (shown in column 2). The observed correlations between Georgia's WIDA ACCESS for ELLs SGPs and prior status are all essentially zero (column 3).

SGP RESULTS

In the following sections basic descriptive statistics from the 2023 analyses are provided, including the state-level mean and median growth percentiles. Currently Georgia uses cohort-referenced SGPs as the official student-level growth metric. Descriptive statistics from the WIDA ACCESS for ELLs SGP results are presented here. The interested reader can find more in depth discussions of the SGP methodology in the available literature and in Appendix B of this report.

STATE LEVEL RESULTS

Growth percentiles, being quantities associated with each individual student, can be easily summarized across groups to provide aggregate information regarding the growth results. The median and mean are used as measures of central tendency to summarize the distribution of growth percentiles as a single number. With perfect model fit, we expect the state-wide median of all student growth percentiles in any grade to be 50 because the data are norm-referenced across all students in Georgia. 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 4 presents Georgia growth percentile medians and means by grade level.

Table 4: 2023 Median (Mean) Student Growth Percentiles by Grade

Grades											
1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
50 (49.9)	50 (50.4)	51 (50.7)	51 (50.8)	51 (50.5)	51 (51.0)	51 (50.8)	51 (50.7)	51 (50.8)	50 (49.9)	51 (50.6)	52 (51.3)

Based upon perfect model fit to the data, the median of all the state's growth percentiles should be 50 for each *grade by year* combination. 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 4 are close to perfect, with all values equal to or near 50.

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

It is important to note how, at the entire state level, the *norm-referenced* growth information returns little information on annual trends due to its norm-reference nature. The results indicate that an 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.). When examining subgroups in terms of the mean or median of their student growth percentiles, it is 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.

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 English language proficiency and aggregate growth is rarely zero. The correlation between prior student status and growth at the school level is a compelling descriptive statistic because it indicates whether students attending schools serving higher scoring students grow faster (on average) than those students attending schools serving lower scoring students. Results from previous analyses conducted using state achievement assessments show correlations between prior status 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[5]. When observed, these positive correlations indicate that students attending schools serving students with lower status (English language proficiency in the context of this report) tend to, on average, demonstrate less exemplary growth than those attending schools serving students with higher status. Equivalently, based upon ordinary least squares (OLS) regression assumptions, the prior status 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 studies on value-added models show similar results (McCaffrey, Han, and Lockwood 2008).

SCHOOL LEVEL RESULTS

The relationship between growth (quantified by the median SGP of students at the school across all grades) and prior status (quantified by the schools' mean standardized scale scores across all grades) is depicted visually in the bubble chart in Figure 3. This type of chart has been successful in motivating discussions of the two qualities: student English language proficiency and student growth. Though the figure is not detailed enough to indicate strength of relationship between growth and status, they are suggestive and valuable for discussions with stakeholders who are being introduced to the growth model for the first time.

[5] These relationships vary greatly between states and even between subjects within states. Stronger relationships have been observed in some states as well, but generally stay consistent within any given state and subject from one year to the next.

Figure 3: 2023 WIDA ACCESS for ELLs School-level Growth and Prior English Language Proficiency

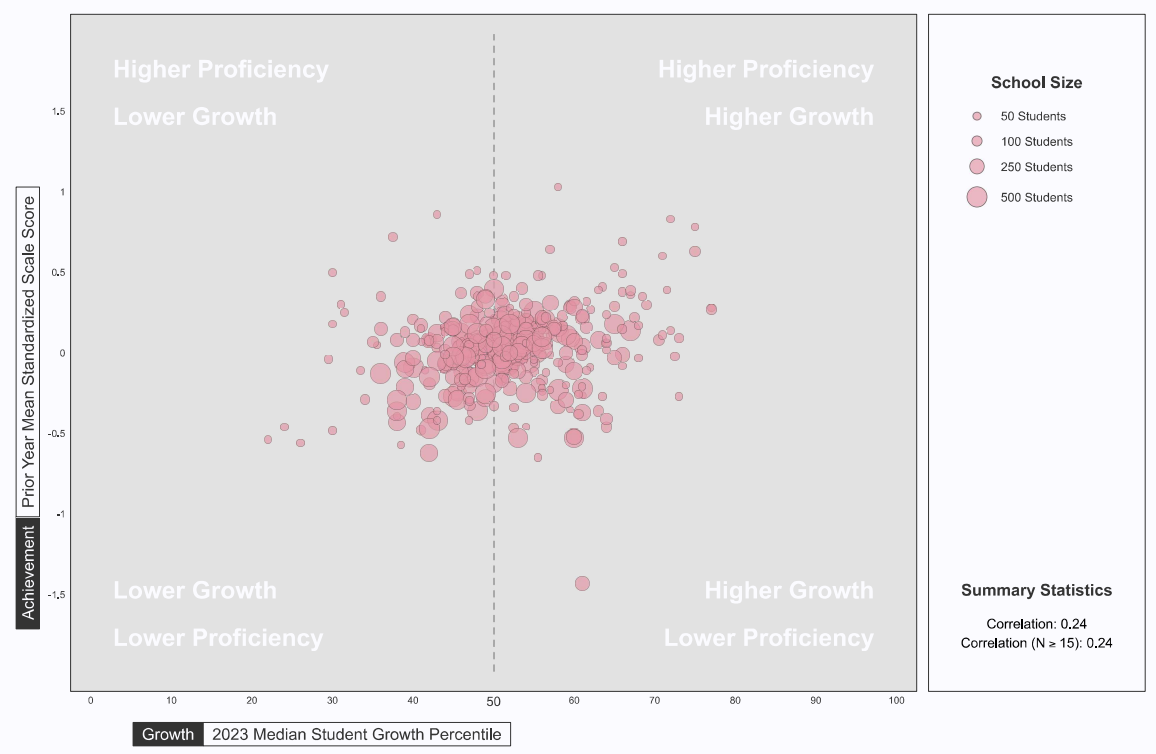


Table 5 shows the correlations between overall prior English language proficiency and overall school growth (both median and mean SGP) for the previous three years. All results shown here include only schools with 15 or more students with valid SGPs.

Table 5: School Level Correlations Between Mean Prior Standardized Scale Score and Median/Mean SGPs by Year

Year	Median SGP	Mean SGP	N
2023	0.24	0.25	361
2022	0.31	0.31	353
2021	0.35	0.36	350

The context Table 5 provides for the relationship between school level growth and English language proficiency combined across grades is important because school accountability measures typically use this level of aggregate measures to produce a single school accountability metric. Although information about the relationship between these high-level aggregates may be helpful and/or familiar when these use cases are relevant, they may well present an "ecological fallacy". That is, aggregating across lower-level groups (e.g., the cohorts upon which the individual SGP analyses are based) tends to inflate resulting correlation statistics.

In order to provide a more detailed look at the relationship between school level growth and prior English language proficiency, Table 6 disaggregates the school level correlations for 2023 by grade. This provides detail at the level at which the SGP analyses are performed. Note that the correlations gradually decrease as the aggregation levels become more granular.

Table 6: School Level Correlations Between Mean Prior Standardized Scale Score and Median/Mean SGPs for 2023 by Grade

Grade	Median SGP	Mean SGP	N
1	0.30	0.32	193
2	0.11	0.12	195
3	0.09	0.08	189
4	0.06	0.04	195
5	0.11	0.09	170
6	-0.03	-0.02	129
7	0.30	0.31	127
8	0.00	0.04	128
9	0.12	0.08	100
10	-0.10	-0.12	89
11	0.00	-0.02	69
12	0.00	0.01	58

REFERENCES

Betebenner, Damian W. 2009. "Norm- and Criterion-Referenced Student Growth."*Educational Measurement: Issues and Practice* 28 (4): 42–51.

Betebenner, Damian W., Adam VanIwaarden, Ben Domingue, and Yi Shang. 2023. *SGP: Student Growth Percentiles & Percentile Growth Trajectories*.sgp.io.

Chernozhukov, Victor, Ivan Fernandez-Val, and Alfred Galichon. 2010. "Quantile and Probability Curves Without Crossing."*Econometrica* 78 (3): 1093–1125.

Koenker, Roger. 2005. *Quantile Regression*. Cambridge: Cambridge University Press.

McCaffrey, Daniel, Bing Han, and J Lockwood. 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 Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. http://www.R-project.org.