RENAISSANCE LEARNING

May 2016

Special Report

Student Growth Percentile in STAR Assessments™

STAR™
Assessments

Reports and software screens may vary from those shown as enhancements are made.

RENAISSANCE LEARNING
P.O. Box 8036
Wisconsin Rapids, WI 54495-8036
(800) 338-4204
www.renaissance.com
educatordevelopment@renaissance.com

05/16

Contents

Figures

Introduction

Student achievement typically is gleaned from one score at a single point in time. However, considering growth in addition to achievement greatly enriches an educator's understanding of how well a student is performing (Betebenner, 2009; Thurlow, Lazarus, Quenemoen, & Moen, 2010). While *achievement* indicates whether performance is below, above, or on par with grade-level expectations, *growth* explains the type of progress the student is making over time. For example, a student may be performing at a low level, yet experiencing high rates of growth. Conversely, a high-performing student's growth could be stagnating. In other words, it is important to know how a student is performing, but this information must have context—how remarkable is this growth given a student's achievement history?

> While *achievement* indicates whether performance is below, above, or on par with grade-level expectations, *growth* explains the kind of progress the student is making over time.

Many state accountability systems incorporate a plan for measuring growth over time, reflecting broad agreement that such systems must go beyond reporting the percentage of students obtaining proficiency status by the end of the school year (Domaleski & Perie, 2012). This paper describes student growth percentiles (SGP), an increasingly popular method of characterizing student growth that is used in Renaissance Learning's STAR Reading, STAR Math, and STAR Early Literacy assessments.

Growth

Growth over time, which is sometimes called slope or rate of improvement, is of central importance in evidence-based instructional models such as Response to Intervention and Multi-Tiered Systems of Support. When educators are able to capture and accurately interpret growth information, they can make informed, data-based decisions regarding the extent to which students are benefiting from intervention or regular classroom instruction, or whether changes are warranted (Fox, Carta, Strain, Dunlap, & Hemmeter, 2009).

To illustrate why interpreting different rates of growth can be more complex than it may seem, consider the following example. Figure 1 highlights the importance of understanding growth by depicting the performance of two high jumpers. Over a four-month period, Athlete A increased her high jump by 4 inches, while Athlete B increased his by 1 inch. At first glance, Athlete A seems to have made greater improvement. However, to determine the significance of these increases in jump height, we must also consider the athletes' performance history and peer groups.

Figure 1. Growth is better understood when performance history and peer group are considered

Athlete A, a novice, increased her high jump by 4 inches over four months.

Athlete B, an Olympian, improved his high jump by 1 inch over four months.

Athlete A is a novice who had room for improvement, while Athlete B is an Olympian who, even while performing at his peak, was able to improve. How should we interpret these gains? Whose growth was more impressive? Having background information helps us know that the growth achieved by the expert Olympian was more impressive than

the novice's improvement. Absent information about the growth that would be expected for each type of athlete, it is difficult to draw these conclusions.

In education, knowing absolute change in achievement—in scaled score, for example—is not helpful for making meaning from data. Without context, we do not know if the growth was expected, below what was expected, or extraordinary. The amount each student grows can vary by test/subject, grade, and prior achievement, so simply knowing that a student's scores increased is only half the story.

The amount each student grows can vary by test/subject, grade, and prior achievement, so simply knowing that a student's scores increased is only half the story.

A number of statistical models have been designed to measure student growth. Castellano and Ho (2013a) provide an overview of seven such models. One of the most widely used is student growth percentile, which was developed by Dr. Damian Betebenner of the National Center for the Improvement of Educational Assessment and piloted in partnership with various state departments of education (Betebenner, Vanlwaarden, Domingue, & Shang, 2016). SGPs have been adopted by a number of states for instructional and accountability purposes.

Renaissance's STAR Assessments (STAR Early Literacy, STAR Reading, and STAR Math) were the first interim tests to report student growth percentiles. Growth models like SGP require an enormous amount of data to generate reliable results (Castellano & Ho, 2013a). Fortunately, widespread national use of STAR Assessments provides ample data, enabling SGPs to be reported for nearly every student in every grade,[1] no matter how high or low their initial achievement level. To learn more about the sample used in creating the SGP model, see *Sample characteristics*, p. 8.

Student growth percentiles

SGPs are a norm-referenced quantification of individual student growth derived using quantile regression techniques (Betebenner, 2011). The SGP score compares a student's growth from one period to the next with that of his or her academic peers nationwide—defined as students in the same grade with a similar scaled score history. SGPs range from 1–99 and interpretation is similar to percentile rank (PR) scores: lower numbers indicate lower relative growth and higher numbers indicate higher relative growth. For example, an SGP of 75 means that the student's growth exceeds the growth of 75 percent of students with a similar score history.

SGPs help us understand, given where a student started, to what extent the growth achieved was as expected. Without an SGP, a teacher may not know if a scaled score increase of 100 is good, not-so-good, or average because what is expected growth for one student may not be for another. An SGP of 50 is typical growth for a particular student, given his/her grade and prior score history; however, state and local policy makers may define typical SGP as a less precise range, such as 35 to 65 or 40 to 60.

SGPs can be aggregated to describe growth for groups of students—such as for a whole class, grade, or school—by calculating the group's mean or median (middle) growth percentile. No matter how SGPs are aggregated, the statistic and its interpretation remain the same. For example, a median SGP of 62 for a class means the middle student in that group achieved higher growth than 62 percent of his or her academic peers.

All students, regardless of their score history, have as good a chance of demonstrating high growth as low growth (i.e., scoring at any of the 99 SGPs).

A common misunderstanding regarding SGP scores is that their statistical distribution is normal, like a bell curve. This would mean that there are more SGPs reported in the middle (near 50) than there are at the tails, near 1 and near 99. This is not true. While it is possible for SGP scores at local (e.g., class) levels to have any type of distribution,

1 A few exceptions: (1) first graders do not receive SGPs reflecting spring-to-spring, spring-to-fall, or fall-to-fall growth because each requires at least one test score from the kindergarten year, and kindergarten scores are not included in the SGP model for STAR Reading or STAR Math, and (2) for STAR Early Literacy, scores are only included in the model through third grade.

nationally the distribution is approximately flat for all grades and subjects. Thus, within any subject/grade, the number of reported scores at every point between 1 and 99 will be about the same (each score is reported for about 1 percent of students). There will be approximately the same number of students with an SGP of 50 as 6 as 92 as 37, and so on. Because of this uniform distribution, all students, regardless of score history, have as good a chance of demonstrating high growth as low growth (i.e., scoring at any of the 99 SGPs).

It is important to remember that no matter how high, low, or middle of the road a student's PR score, the student has an equal chance of receiving any SGP score ranging from 1–99. Take, for example, a student with a fall percentile rank of 95 who receives an SGP of 19 at the end of the year. It may not seem reasonable that such a high-performing student would receive a relatively low growth score, but what this indicates is that 81 percent of this student's academic peers from the same grade with a similar score history experienced more growth. SGP compares the student's performance to that of a group of unique academic peers—students with a similar scaled score history—that is recalculated each time the student takes an assessment. No assumptions can or should be made about a student's SGP based on PR performance. (Note: Although we reference PR scores to illustrate points about achievement and growth, PRs are not used in the SGP calculation.)

Applying SGP to STAR Assessments™

During the 2011–2012 school year, Renaissance first reported SGPs in STAR Reading and STAR Math for grades 1–12 and in STAR Early Literacy for grades K–3. To apply the SGP approach to STAR Assessment data, Renaissance researchers worked closely with SGP creator Dr. Betebenner.

Testing windows

Because SGP was initially developed for measuring growth on state tests across school years, applying the SGP approach to interim assessment data involved a number of technical challenges, primarily regarding differences in the timing of STAR versus state test administrations.

State summative tests are typically administered once a year, at approximately the same time, to nearly all students. Thus, score comparisons from one state test administration to another speak to growth across school years. Consequently, the original SGP model first developed by Dr. Betebenner for state use assumes fairly constrained administration parameters with approximately the same amount of time in between tests. In stark contrast, STAR Assessments can be considered "on-demand" tests and are far more flexible. Administration decisions (when and to which students) are left to local educators based on their purposes and needs for assessment. Most commonly, schools choose to use STAR as a screening or benchmarking test for all, or nearly all, students 2–4 times per year. Students requiring progress monitoring may take the assessments more frequently to inform instructional decisions, such as whether a student is responding adequately to an intervention.

About the STAR Assessments

STAR Assessments are reliable, valid, and time-efficient assessments of early literacy (STAR Early Literacy), reading (STAR Reading), and mathematics (STAR Math) skills. Quick and accurate results from these assessments provide teachers with specific benchmarking, screening, progress-monitoring, and diagnostic information to help tailor instruction, monitor growth, and improve achievement for all students.

STAR Assessments are highly rated for progress monitoring and screening by the National Center on Intensive Intervention (2016a, 2016b, 2016c) and the National Center on Response to Intervention (2010a, 2010b, 2010c, 2011a, 2011b, 2011c). For more information on the reliability, validity, and other technical aspects of STAR Assessments, see the STAR technical manuals, available by request to research@renaissance.com.

Given that not all students take STAR Assessments at the same time, and that the number and dates of test administrations may vary from one student to the next, it was necessary to make two adaptations for STAR SGP: (1) identify testing windows and, (2) adjust for variable time between tests. Analysis of STAR data revealed a clear pattern for the majority of tests taken during the school year, which corresponded closely with the timing of district screening or benchmarking: Fall (August 1–November 30), Winter (December 1–March 31), and Spring (April 1–July 31).

Specific date ranges for the windows were identified when defining the data sets used to determine SGPs. Establishing testing windows allowed STAR SGPs to be reported within-year in a manner consistent with most district testing calendars.

Calculating SGPs

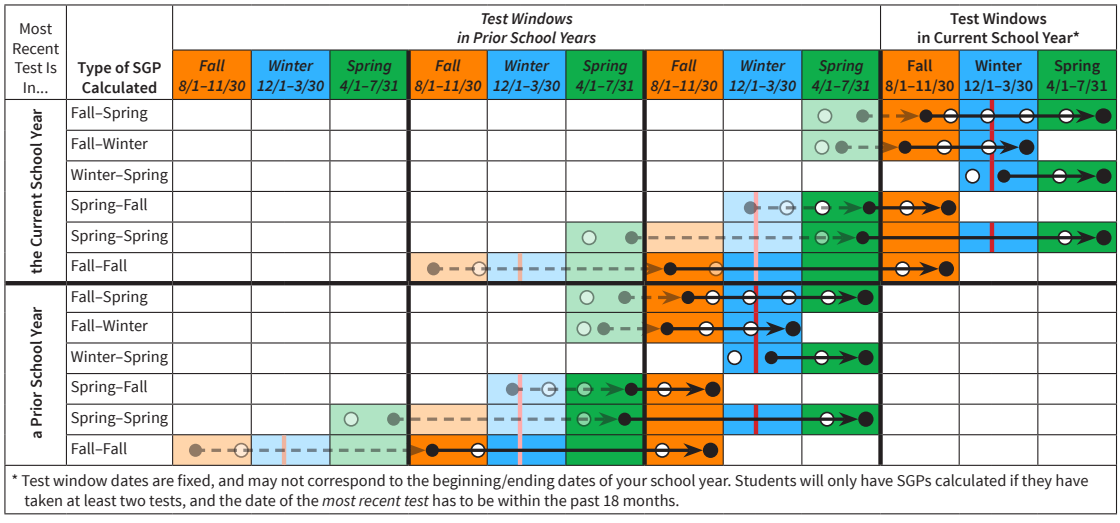
Quantile regression is a statistical process used in SGP models to estimate the conditional distribution of an outcome variable (a test score) given prior information (a student's prior scores). An SGP reflects the likelihood of a specific outcome (an amount of growth over a period of time) given a student's prior score history, using data available from all students from recent years that characterize how different students grow. In general, this method can be viewed as a type of smoothing, in which information from neighboring score values can be used to inform percentiles for hypothetical score combinations not yet observed (Betebenner, 2016).

Recent enhancements to the SGP model prioritize available data points to make the best use of information across time, by using a student's current test score (the posttest) and up to two prior test scores (the pretest and, if available, an additional prior test):

- Posttest: A score from the most recent test taken within the last 18 months.
- Pretest: A score from a test in an SGP window prior to the window the posttest falls within.
- Additional prior test: A score, if available, from a window in the previous school year. Empirical evidence (Betebenner, 2016) shows that using a student's prior-year score, when available, ensures the most accurate representation of growth within an academic year.

Each time a student takes a STAR Assessment, he/she receives a current SGP score. The score is reported based on the available STAR test score history for that student. Figure 2 shows the decision rules that guide how an SGP score is reported. The type of score a student receives is prioritized from top to bottom in the table, depending on available data. When more than one test has been taken in an SGP window, the model uses the following scores: the first test taken in fall, the test taken closest to January 15 in winter, and the last test taken in spring.

Figure 2. Decision rules for SGP model score selection



Note: The type of SGP score a student receives is prioritized from top to bottom in this table, depending on available test data.

Getting the most accurate SGP: The purpose of the additional prior score

Academic peer groups are key to calculating SGPs. But how can the model ensure the best possible peer-group selection? Considering an additional prior score, along with the pretest and posttest scores, helps to identify each student's ideal academic peer group (Betebenner, 2016).

In the SGP calculation, the posttest (current test) and pretest scores are used to determine growth, while the additional prior score serves to stabilize the student's pretest score, minimize the impact of measurement error,[2] and ensure the most accurate picture of the student's optimal academic peer group. While it may appear the model is considering data from a prior school year as a pretest, it is actually just using this additional reference point to further inform each student's unique academic peer group. Disregarding this additional data point from a student's prior performance would be to knowingly ignore valuable baseline information.

Using a prior-year score to better pinpoint a student's unique academic peer group does not mean that estimates of student growth within a current school year are any less useful or appropriate on their own. Rather, Dr. Betebenner's ongoing research has shown convincing evidence that by improving the association of students' scores with those of their peers, the SGP model can now provide an even more complete picture of individual student growth. Because of the important role SGP scores play in instructional and accountability decisions, Renaissance and Dr. Betebenner are committed to a continuous improvement cycle. Enhancements include conducting research that informs the usability of the SGP score, as well as frequent updating of the SGP score norming samples , a common practice for any norm-referenced score. For more information on how scores generated by the SGP model correlate well from year to year, see *Reliable and Valid Results*, p. 6.)

For example, suppose two students have very similar posttest and pretest scores. One might expect their resulting SGP scores to also be very similar. The scores may very well turn out to be the same or close, but simply looking at similar growth between a posttest and pretest does not provide as complete a picture of the students' growth as is possible. Incorporating an additional prior score into the calculation provides added context and stabilizes each student's pretest score. In examining this additional data point, we may find, for example, that the timing of the prior test events differed for the students, thereby giving them varying levels of exposure to skills and learning time. Even more importantly, one student's prior score might have been higher than his/her pretest score, while the other student's prior score might have been much lower than the pretest. This would mean the students' academic peer groups were different, which would result in varying SGPs. In other words, although the most recent test scores make it seem that these two students would be academic peers, using an additional data point provides a more accurate picture of each students' individual score histories.

Adjusting for time

At Renaissance, our goal is to provide the best possible indication of how a student is growing, given the available data and research. As ongoing research has demonstrated that adjustments to the SGP calculation will improve this growth measure, we believe in utilizing that research to ensure fair and accurate comparisons of data. Thus, the STAR SGP model has evolved to use time in two ways:[3]

> Considering an additional prior score, along with the pretest and posttest scores, helps to identify each student's ideal academic peer group.

(1) The amount of days between the posttest and the pretest. The testing windows alone do not address the fact that students in the same window may have spans of time between tests that vary greatly—and, consequently, different opportunities to learn and grow. For instance, a student with tests on the first day of the fall window and the last day of the spring window would have 364 days between test events, while another student testing on the last day of the fall window and the first day of the spring window would have 122 days between tests. The more days between two testing events, the more growth that can be expected.

2 Standard error of measurement (SEM) is unavoidable and is present to some degree in all assessments. Assessment developers can only seek to minimize the impact of SEM. Tests with good technical characteristics, such as the STAR Assessments, should reliably generate consistent and accurate estimates of a student's achievement. (For more information on the value of adding an additional prior score to the SGP model, see the technical paper by Betebenner, 2016.)

3 For more information on the time-sensitive calculation implemented in the SGP model, see the technical paper by Betebenner (2016).

(2) When in the window a student took the current test (which indicates how close or far the student is from the start of the testing window). Students at the end of the testing window have had more exposure to content and, thus, their scaled scores are likely to be higher.

Reliable and valid results

Each year since its initial development, the SGP model has been reviewed, with minor improvements made to increase its reliability and validity. Within STAR, these advances yield results that are highly correlated across years, meaning educators can use all SGP results with confidence to inform both goal setting for students and educator evaluation purposes.

In early 2016, Renaissance conducted an analysis of STAR scores to understand the extent to which the most recent enhancements to the SGP model for the 2015–16 school year (which consider an additional prior score with pre/post scores and an adjustment to how time is handled) correlate with the previous calculation (used in 2014–15). Researchers ran the same set of student scores through both iterations of the calculation and compared the resulting SGPs.

> Changes in instruction, the school environment, and the students' aptitude may explain why students do not receive the same SGP score over time.

The sample included STAR Early Literacy scores for 639,425 students in grades K–3, STAR Math scores for 3,499,359 students in grades 1–12, and STAR Reading scores for 6,352,572 students in grades 1–12. Most records included three scores (posttest, pretest, and additional prior), but some included only two scores (posttest and pretest). Results revealed high average correlations in the mid .9s, with a range of coefficients from .82 to .99 when looking at specific grade/subject combinations. Overall, the analysis showed that although recent changes provide meaningful improvement in the accuracy of the SGP score, both calculations sort students in a consistent manner and provide reliable estimates of student growth.[4]

Even though the SGP calculation correlates closely with previous iterations, teachers will find that their students' SGP scores tend to fluctuate from test period to test period. Why might SGPs vary across time? Educators may expect to see highly consistent SGPs for a given student or group of students within year or across years, but this is highly unlikely for several reasons. Changes in instruction, the school environment, and the students' aptitude, as well as the impact of measurement error (common in all educational tests) may explain why students do not receive the same SGP score over time.

Educators are advised to consider expert recommendations (e.g., Hamilton et al., 2009) regarding the use of multiple source of information to inform instructional decisions. Although STAR SGP is a robust growth measure on its own, it should be used in combination with other reliable and valid sources of information about student achievement and growth.

Reporting SGPs

Recent improvements to the model also provide educators with an SGP for every student at the start the school year (as long as data exists from the previous year). The availability of an SGP in fall allows teachers to begin the year understanding students' recent growth history, which can provide immediate insight and assist with initial instructional decisions. As the year progresses and additional assessments are taken, STAR Assessments then report each student's current SGP in the District Dashboard, Reading Dashboard and/or Math Dashboard, Growth Report, Growth Expectations Extract, Growth Proficiency Chart, and Goal-Setting Wizard.

As figure 3 shows, the Dashboard displays data on student performance, charting a student's current score and a prism representing future growth possibilities. This tool addresses questions such as, *How is a student performing over time and relative to state proficiency benchmarks? What are the likely growth possibilities for this student?*

4 As expected, the results did not perfectly correlate, which would call into question the efficacy of model enhancements if they produce precisely the same results.

The STAR Growth Report (see figure 4) summarizes growth between two testing periods in the same school year as soon as a student has both a pretest and posttest score. Teachers can run the report for a class or specific groups of students and administrators can see growth for each class or grade in their schools.

Because this report presents the current SGP from the most recent STAR administration, educators are advised to generate and save the Growth Report on a periodic basis in order to have a record of SGP data.[5]

Historical SGPs can also be viewed within the Reading and Math Dashboards under the All-Time view. For more information about the Growth Report, see *Frequently asked questions*, p. 11.

Figure 3. Sample Dashboard screen

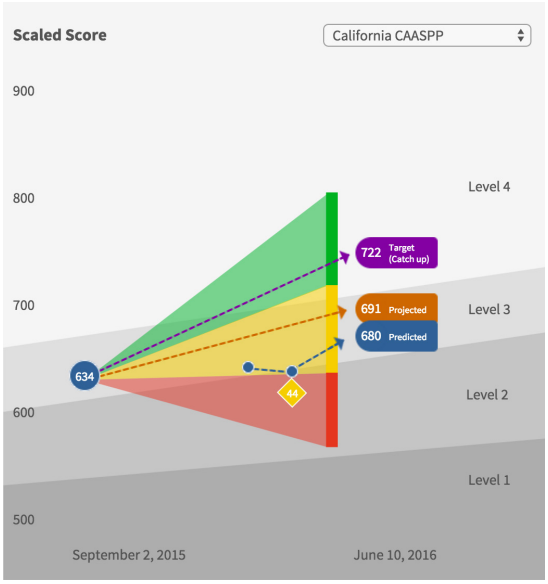
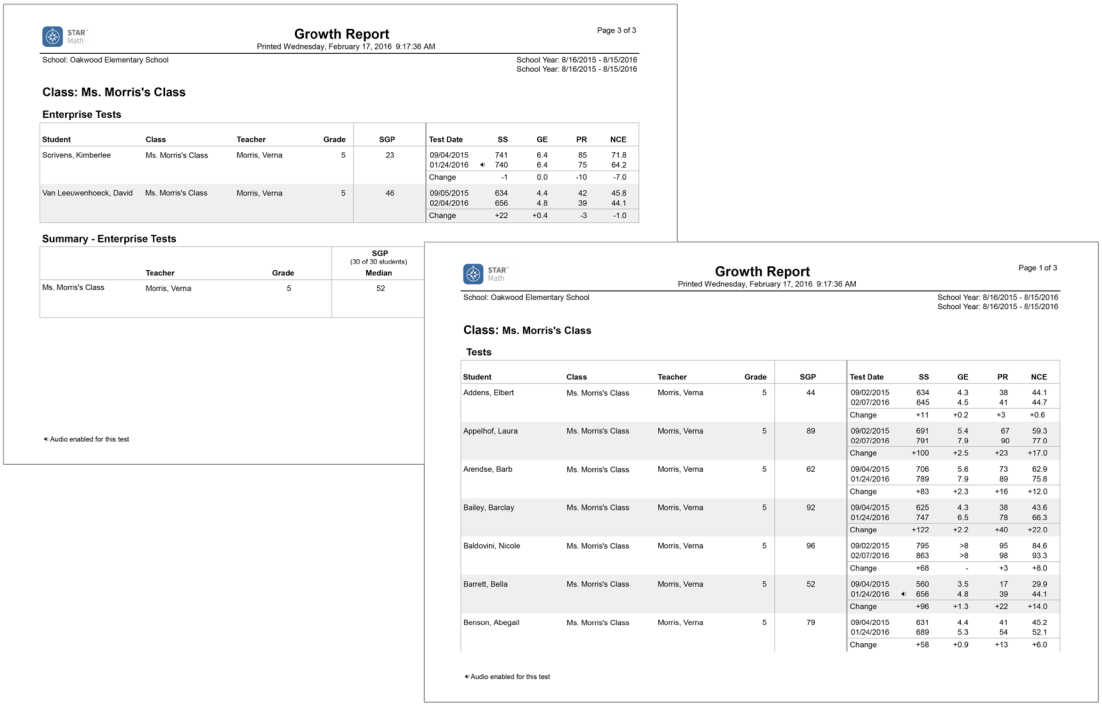


Figure 4. Sample STAR Math Growth Report



STAR Math

Growth Report

Printed Wednesday, February 17, 2016 9:17:36 AM

Page 3 of 3

School: Oakwood Elementary School

School Year: 8/16/2015 - 8/15/2016
School Year: 8/16/2015 - 8/15/2016

Class: Ms. Morris's Class

Enterprise Tests

Student	Class	Teacher	Grade	SGP	Test Date	SS	GE	PR	NCE
Scrivens, Kimberlee	Ms. Morris's Class	Morris, Verna	5	23	09/04/2015	741	6.4	85	71.8
					01/24/2016	740	6.4	75	64.2
					Change	-1	0.0	-10	-7.0
Van Leeuwenhoeck, David	Ms. Morris's Class	Morris, Verna	5	46	09/05/2015	634	4.4	42	45.8
					02/04/2016	656	4.8	39	44.1
					Change	+22	+0.4	-3	-1.0

Summary - Enterprise Tests

	Teacher	Grade	SGP (30 of 30 students) Median
Ms. Morris's Class	Morris, Verna	5	52

◂ Audio enabled for this test

STAR Math

Growth Report

Printed Wednesday, February 17, 2016 9:17:36 AM

Page 1 of 3

School: Oakwood Elementary School

School Year: 8/16/2015 - 8/15/2016
School Year: 8/16/2015 - 8/15/2016

Class: Ms. Morris's Class

Tests

Student	Class	Teacher	Grade	SGP	Test Date	SS	GE	PR	NCE
Addens, Elbert	Ms. Morris's Class	Morris, Verna	5	44	09/02/2015	634	4.3	38	44.1
					02/07/2016	645	4.5	41	44.7
					Change	+11	+0.2	+3	+0.6
Appelhof, Laura	Ms. Morris's Class	Morris, Verna	5	89	09/02/2015	691	5.4	67	59.3
					02/07/2016	791	7.9	90	77.0
					Change	+100	+2.5	+23	+17.0
Arendse, Barb	Ms. Morris's Class	Morris, Verna	5	62	09/04/2015	706	5.6	73	62.9
					01/24/2016	789	7.9	89	75.8
					Change	+83	+2.3	+16	+12.0
Bailey, Barclay	Ms. Morris's Class	Morris, Verna	5	92	09/04/2015	625	4.3	38	43.6
					01/24/2016	747	6.5	78	66.3
					Change	+122	+2.2	+40	+22.0
Baldovini, Nicole	Ms. Morris's Class	Morris, Verna	5	96	09/02/2015	795	>8	95	84.6
					02/07/2016	863	>8	98	93.3
					Change	+68	-	+3	+8.0
Barrett, Bella	Ms. Morris's Class	Morris, Verna	5	52	09/04/2015	560	3.5	17	29.9
					01/24/2016	656	4.8	39	44.1
					Change	+96	+1.3	+22	+14.0
Benson, Abegail	Ms. Morris's Class	Morris, Verna	5	79	09/04/2015	631	4.4	41	45.2
					01/24/2016	689	5.3	54	52.1
					Change	+58	+0.9	+13	+6.0

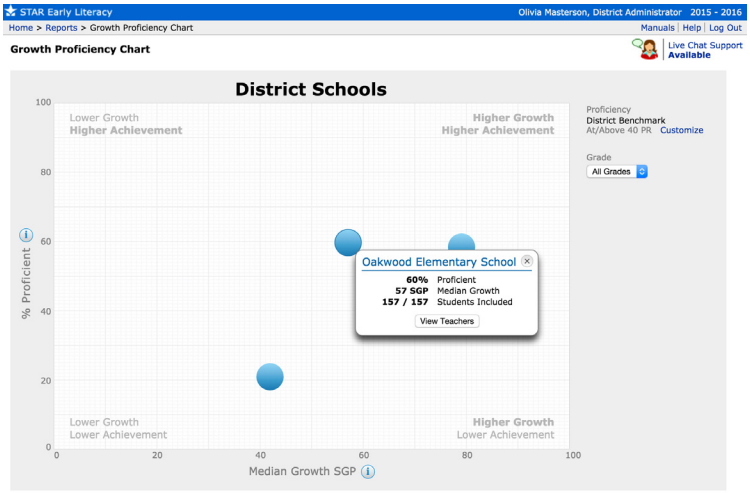
◂ Audio enabled for this test

The Growth Proficiency Chart (see figure 5, next page) is an interactive STAR tool that displays data on the relationship between estimated proficiency[6] and growth (expressed with SGPs). The chart shows wheither students, classes, or schools are experiecing low proficiency and low growth, low proficiency and high growth, high proficiency and low growth, or high proficiency and high growth.

5 *Note:* Test dates on the Growth Report apply to all scores shown, except SGP. In the case of the SGP score, the dates may not be the parameters used to determine the reported score. Thus, the report displays the SGP score set off from other scores and includes a footnote explaining the information provided. Tests used to determine a student's current SGP are shown on the Growth Proficiency Chart (see figure 5, next page).

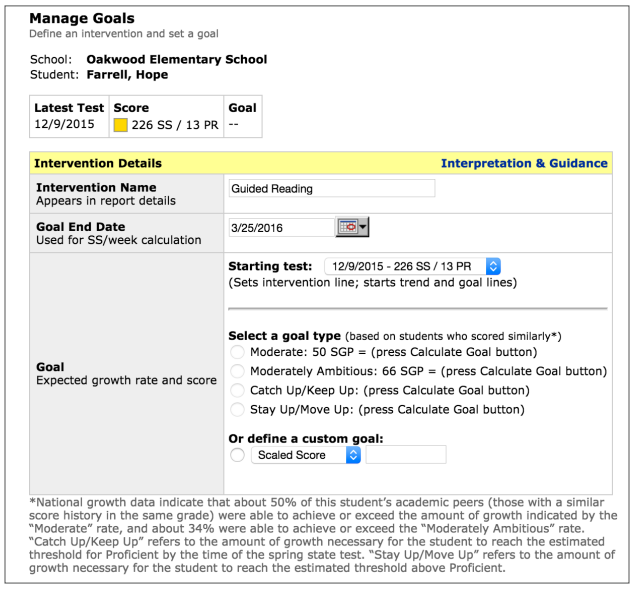
6 STAR Reading and STAR Math are statistically linked to summative assessments used by the majority of states to help answer the question, *How will my students perform on the state test?* To populate the Growth Proficiency Chart, linking study data is combined with expected weekly scaled score growth (from decile-based growth norms, which take into account grade and observed starting score). STAR Early Literacy scores are not linked to state tests because most states do not test students until grade 3.

Figure 5. Sample STAR Early Literacy Growth Proficiency Chart



The Goal-Setting Wizard (see figure 6) allows teachers to set goals for student performance. Using this tool, teachers can calculate scaled score and SGP goals for students. The wizard does not set or recommend goals for students; rather, it provides educators with a tool to both inform educators' creation of goals and monitor student progress against interventions and other initiatives.

Figure 6. Sample view of Goal-Setting Wizard



Sample characteristics

Each year, approximately 60 to 70 million STAR tests are taken, and nearly all of these scores are included in the sample used to report SGPs. Renaissance updates this information throughout the year. While we do not report SGPs for specific subgroups of students, all students—regardless of special education or English learner status—are retained in the sample. We do, however, limit the sample to STAR tests administered in typical school settings; tests administered by tutoring centers or virtual schools are excluded from the analysis.

Frequently asked questions

What is a student growth percentile (SGP)?
A student growth percentile, or SGP, compares a student's growth to that of his or her academic peers nationwide. Academic peers are students in the same grade with similar achievement history on STAR Assessments. SGP is reported on a 1–99 scale, with lower numbers indicating lower relative growth and higher numbers indicating higher relative growth. For example, an SGP score of 90 means the student has shown more growth than 90 percent of his/her academic peers.

How are SGPs determined?
SGPs are based upon the best available information using a statistical model of growth and achievement called a quantile regression. The way the model prioritizes data points is designed to make the best use of data across time. The SGP calculation uses test scores from at least two SGP windows, and a third SGP window when available:

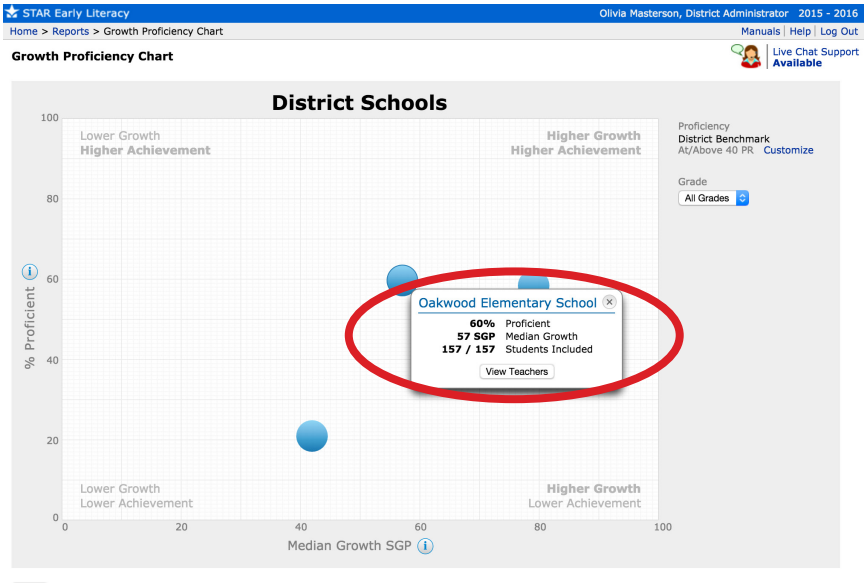
- **Posttest**: A score from the most recent test taken within the last 18 months.
- **Pretest**: A score from a test in an SGP window prior to the window the posttest falls within.
- **Additional prior test**: A score, if available, from a window in the previous school year. Empirical evidence (Betebenner, 2016) shows that using a student's prior-year score, when available, ensures the most accurate representation of growth within an academic year.

- **SGP windows**
  - Fall (August 1–November 30)
  - Winter (December 1–March 31)
  - Spring (April 1–July 31)
  - When there is more than one test taken in the SGP window, the following test scores are used:
    - Fall: first test taken
    - Winter: test closest to January 15
    - Spring: last test taken

I tested a student in the fall, winter, and spring. Why is the previous year's test score used to determine the SGP score?
SGP creator Damian Betebenner's ongoing research (2016) has shown convincing evidence that by improving the association of students' scores with those of their peers (by taking into account an additional prior score for the student from the previous school year), the SGP model can now provide an even more complete picture of individual student growth. Including this additional data point helps to better pinpoint a student's optimal academic peer group, which results in a more accurate measurement of growth between fall and spring.

What is the purpose of the additional prior test score in the SGP calculation?
Using the posttest (current) score, the pretest score, and an additional prior score in the SGP calculation helps to identify the most accurate picture of a student's academic peer group (Betebenner, 2016). The posttest and pretest scores are used to determine growth, while the additional prior serves to stabilize the pretest score, aid in the selection of the student's ideal peer group, and minimize the impact of measurement error.* Disregarding this additional data point from a student's prior-year performance would be to knowingly ignore valuable baseline information. *(*Note: Standard error of measurement [SEM] is unavoidable and is present to some degree in all assessments. Assessment developers can only seek to minimize the impact of SEM. Tests with good technical characteristics, such as the STAR Assessments, should reliably generate consistent and accurate estimates of a student's achievement. For more information on the value of adding* an *additional prior score to the SGP model, see the technical paper by Betebenner, 2016.)*

How do I know which test scores were used to determine the SGP score?
The tests used to determine a student's SGP are listed on the STAR Growth Proficiency Chart (as circled below).



How do I get an SGP for an earlier time in the year or a prior school year?
STAR will report a student's current SGP by using a score from the current SGP window (the posttest) and up to two test scores from prior SGP windows (the pretest and, if available, an additional prior test). Historical SGPs can be viewed within the Reading and Math Dashboards under the All-Time view. Educators are also advised to generate and save the Growth Report* on a periodic basis to have a record of SGP data. *(*Note: Test dates on the Growth Report apply to all scores shown, except SGP. In the case of the SGP score, the dates may not be the parameters used to determine the reported score. Actual test dates are shown on the Growth Proficiency Chart; see figure above.)*

Can I get a Winter-to-Spring SGP?
If STAR scores are available for the Fall, Winter, and Spring windows, the student will receive an SGP reflecting fall-to-spring growth. The model defaults to reporting fall-to-spring growth because historically this has been the period of greatest interest to educators using STAR Assessments. If no fall score exists, but STAR tests were taken in both the Winter and Spring windows, the reported SGP will reflect winter-to-spring growth.

Can SGP scores be compared from year to year?
Yes. Although adjustments are made to the model each year, the scores can be compared over time. To study the comparability of the scores, in 2016, researchers examined SGP scores from a large set of STAR student records from the 2014–15 school year. The scores were run through both the SGP model used in 2014–15 and the recently enhanced SGP model used in 2015–16. Results revealed high average correlations in the mid .9s, with a range of coefficients from .82 to .99 when looking at specific grade/subject combinations. Overall, the analysis showed that although recent changes provide meaningful improvement in the accuracy of the SGP score, both calculations sort students in a consistent manner and provide reliable estimates of student growth. (As expected, the results did not perfectly correlate, which would call into question the efficacy of model enhancements if they produce precisely the same results.)

Because of the important role SGP scores play in instructional and accountability decisions, Renaissance Learning and SGP creator Dr. Betebenner are committed to a continuous improvement cycle. Since SGPs were first reported during the 2011–12 school year, yearly refinements have been made to the model to improve functionality and accuracy. These changes, although meaningful, impact neither the interpretation nor the general distribution of the SGP score. All students, whether low, average, or high performing, always has an equal opportunity to achieve any of the 99 SGP scores.

Why might SGPs for the same student vary across time or between different assessments?
Educators may expect to see highly consistent SGPs for a given student within year or across years, but this is highly unlikely for several reasons. Changes in instruction, the school environment, and the student's aptitude, as well as the impact of measurement error (common in all educational tests), may explain why a student does not receive the same SGP every time.

In the case of varying SGPs on different assessments, educators are advised to consider expert recommendations (e.g., Hamilton et al., 2009) regarding the use of multiple source of information to inform instructional decisions. Although STAR SGP is a robust growth measure on its own, it should be used in combination with other reliable and valid sources of information about student achievement and growth.

How are SGP scores distributed nationally?
A common misunderstanding regarding SGP scores is that their statistical distribution is normal, like a bell curve. This would indicate that there are more SGPs reported in the middle (near 50) than there are at the tails, near 1 and near 99. This is not true. While it is possible for SGP scores at local (e.g., class) levels to have any type of distribution, nationally the distribution is approximately flat for all grades and subjects. Thus, within any subject/grade, the number of reported scores at every point between 1 and 99 will be about the same (each score is reported for about 1 percent of students). There will be approximately the same number of students with an SGP of 50 as 6 as 92 as 37, and so on.

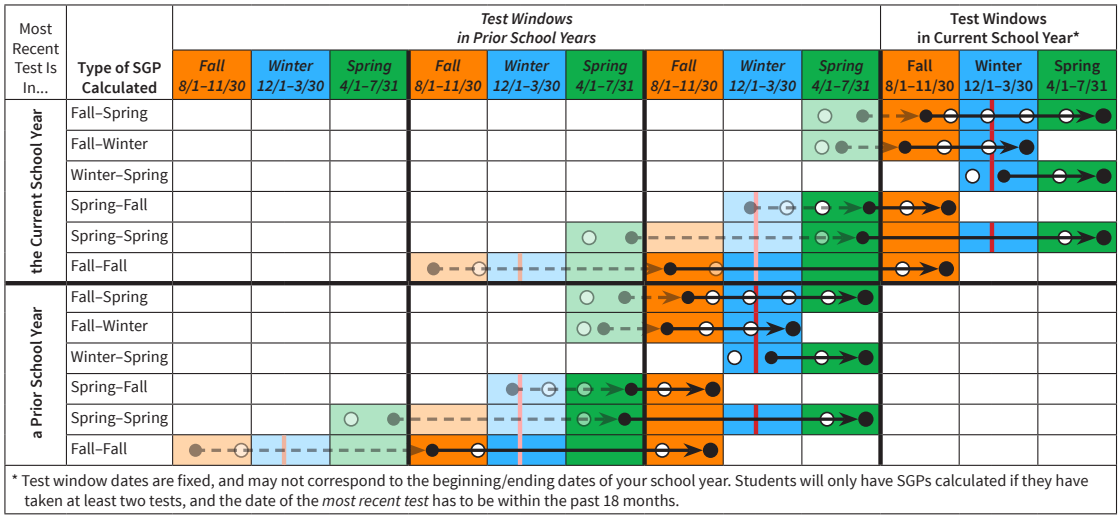
Where are SGPs reported?
STAR Assessments report a student's current SGP in the District Dashboard, Reading Dashboard and/or Math Dashboard, Growth Report*, Growth Expectations Extract, Growth Proficiency Chart, and Goal-Setting Wizard. Historical SGPs can be viewed within the Reading and Math Dashboards under the All-Time view. *(*Note: Test dates on the Growth Report apply to all scores shown, except SGP. In the case of the SGP score, the dates may not be the parameters used to determine the reported SGP score. Actual test dates are shown on the Growth Proficiency Chart; see figure, p. 10.)*

Will I see an updated SGP immediately after testing?
Student data will populate overnight and reporting will reflect an updated SGP the following day.

What does the dash mean on the Growth Report?
There are certain circumstances when a dash (–) will appear in lieu of an SGP on the Growth Report:

- A kindergarten student taking a STAR Reading or STAR Math assessment will not receive an SGP score, as these tests were designed for grades 1–12. (SGPs are reported for kindergarten students who take STAR Early Literacy tests.)
- For a student to receive an SGP based on prior-year spring to current-year fall growth, advancement of one grade is necessary between test administrations. (In addition, mid-year promotion or demotion of grades can result in no SGP score being reported.) This also applies to other cross-year SGPs that would be reported for Fall-Fall and Spring-Spring time periods.
- Student data populates overnight, and a minimum of two assessments must be taken in order to calculate an SGP score. Thus, if it is a student's first test in the second testing window, and the report is being viewed the same day as the test was taken, an SGP score will not appear.
- SGP score data is reported based on the decision rules displayed in the table* shown on the next page. Score history that deviates from what is shown may result in the program being unable to report an SGP score. *(*Note: In the table, the type of score received is prioritized from top to bottom, depending on available data.)*

Decision rules for SGP model score selection



Test Window	If more than one test was taken in a prior test window, which is used to calculate SGP?
Fall Window	First test taken
Winter Window	Test closest to 1/15 (red line)
Spring Window	Last test taken

Note: The type of SGP score a student receives is prioritized from top to bottom in this table, depending on available test data.

Why do I see an SGP in the fall?

Recent improvements to the SGP model allow teachers to begin the school year with an SGP for all students (who have data from the previous school year). For example, if the student had a STAR score in the prior spring and another in the fall of the current school year, an SGP will be generated indicating spring to fall growth. This provides teachers with immediate information about their students' growth history, which can provide insight and assist with initial instructional decisions.

Which STAR Assessments provide SGPs?

SGPs can be calculated for the Enterprise and non-Enterprise versions of STAR Reading, STAR Math, and STAR Early Literacy; however, tests must be taken within the same STAR Assessment (i.e., only STAR Reading, STAR Math, or STAR Early Literacy) in order to obtain an SGP. SGPs cannot be calculated for STAR Reading Spanish, STAR Math Enterprise Geometry, STAR Math Enterprise Algebra, and kindergarten students in STAR Reading, as sufficient sample size must be established for these additional tests in order to compute SGP scores. Renaissance recommends that educators test students with the Enterprise versions of the STAR Assessments. If data for two (or more) tests of the same Enterprise/non-Enterprise version are not available, the software will still calculate an SGP; however, please exercise caution when interpreting results.

How do I obtain an SGP for students who begin the school year taking STAR Early Literacy and transition to STAR Reading during the same school year?

In order to obtain an SGP, tests must be taken within the same STAR Assessment (i.e., only STAR Reading, STAR Math, or STAR Early Literacy). If a student has transitioned from STAR Early Literacy to STAR Reading, consider administering STAR Early Literacy an additional time to obtain an SGP.

Is SGP accurate for high-achieving kids? How can my student be at the 95th percentile and have a 19 SGP?
With SGP, all students, no matter their score history, have an equal chance to demonstrate growth at any of the 99 percentiles. High-achieving students are compared against a national sample of other high-achieving students with similar achievement history (i.e., their academic peers). Thus, it is possible for a student who is scoring well above average at the beginning of the year to have an SGP that is relatively low, typical, or relatively high.

Take, for example, a student with a fall percentile rank of 95 who receives an SGP of 19 at the end of the year. It may not seem reasonable that such a high-performing student would receive a relatively low growth score, but what this indicates is that 81 percent of this student's academic peers from the same grade with a similar score history experienced more growth. No matter how high, low, or middle of the road a student's PR score, the student has an equal chance of receiving any SGP score ranging from 1–99. SGP compares the student's performance to that of a group of unique academic peers—students with a similar scaled score history—that is precisely recalculated each time the student takes an assessment. No assumptions can or should be made about a student's SGP based on PR performance. (Note: Although we reference PR scores to illustrate points about achievement and growth, PRs are not used in the SGP calculation.)

What is typical growth?
Renaissance does not provide benchmarks for typical growth. However, many states that have adopted SGP consider 35–65 SGP as the benchmark for typical growth. For more information, we recommend educators look to states that have adopted SGP to learn more about how they use data from this metric (see *Typical Growth Defined by States,* http://doc.renlearn.com/KMNet/R00585975038A824.pdf).

What is the difference between PR and SGP?
Although they both use a 1–99 scale, percentile rank (PR) and SGP are very different metrics (see table below). PR is an achievement (performance) score that describes a single point in time. SGP is a growth measure that explains student growth between points in time. Both measures are norm-referenced, but they have different norming groups. The norming group for PR is all students in a particular grade level. The norming group for SGP is each student's own academic peer group.

Because they use a similar 1–99 scale, score interpretation is similar between these two scores: lower numbers indicate lower relative growth and higher numbers indicate higher relative growth (e.g., an SGP of 75 means that the student's growth exceeds the growth of 75 percent of students). However, it should be noted that although these scores can be interpreted similarly, that does not mean that a student with a high PR score will likely receive a high SGP score. A high PR means a student is performing well at a certain point in time. No matter how high, low, or middle of the road a student's PR score, the student has an equal chance of receiving any SGP score ranging from 1–99. SGP compares the student's performance to that of a group of unique academic peers—students with a similar scaled score history—that is precisely recalculated each time the student takes an assessment. No assumptions can or should be made about a student's SGP based on PR performance. (Note: Although we reference PR scores to illustrate points about achievement and growth, PRs are not used in the SGP calculation.)

Percentile rank (PR)	Student growth percentile (SGP)
Based on scale of 1–99	Based on a scale of 1–99
Performance score	Growth score
PR reported after one test	At least two tests are needed to report an SGP
Describes a student's achievement at single point in time	Measures a student's growth
Norm-referenced—compares students in the same grade	Norm-referenced—compares students in same grade with similar achievement history
Scaled score is compared to national norm group of grade-level peers	Scaled scores are compared to national norm group of grade-level academic peers

What does it mean when a student has a high PR and a low SGP?
This critical question pinpoints why it is important to look at both achievement (percentile rank/scaled score (SS)) and growth (SGP). Achievement scores, like PR and/or SS, tell us at what level students are performing at a single point in time; however, this is only a piece of the puzzle. It is also important to know how students perform over time in relation to their peers, a question that can be answered using comparative growth data from SGP. For example, consider a student who starts and ends the year at the 50th PR with an SGP of 30. This student is consistently performing better than 50 percent of students in the same grade nationwide. However, when examining growth over time in relation to academic peers (SGP), this student is growing more than only 30 percent of his/her academic peers with similar score histories.

Keep in mind that no matter how high, low, or middle of the road a student's PR score, the student has an equal chance of receiving any SGP score ranging from 1–99. SGP compares the student's performance to that of a group of unique academic peers—students with a similar scaled score history—that is recalculated each time the student takes an assessment. No assumptions can or should be made about a student's SGP based on PR performance. (Note: Although we reference PR scores to illustrate points about achievement and growth, PRs are not used in the SGP calculation.)

In the states that have adopted SGP, how will a student's SGP from the state test compare to a STAR SGP?
A student's SGP on any assessment can vary from a STAR SGP because of differences in test content, blueprint, and delivery, as well as the amount of time between test administration and the norming groups used. Educators are advised to consider expert recommendations (e.g., Hamilton et al., 2009) regarding the use of multiple source of information to inform instructional decisions. Although STAR SGP is a robust growth measure on its own, it should be used in combination with other reliable and valid sources of information about student achievement and growth.

Why can't I get SGP based only on students in my state?
Growth models like SGP require an enormous amount of data to generate reliable results (Castellano & Ho, 2013a). Examining data for students nationwide provides an adequate sample size to calculate reliable and valid SGPs that compare students accurately to their academic peers throughout the U.S.

Mean or median?
In keeping with the vast majority of states who report SGPs on their state summative tests, Renaissance reports median SGP. However, we recognize recent research on this topic (Castellano & Ho, 2013b) concludes it may be appropriate to use mean or median. Educators in states that report SGP on state summative tests may want to consult their state's position on this matter and use the preferred statistic. All educators should exercise caution when aggregating SGP results for small classes/groups (fewer than 20 students). Both mean and median are subject to providing misleading estimates of central tendency, depending on distribution of scores and group size. For this reason, some states have chosen not to report SGP results for small groups.

Can SGP be used with English learners or students receiving special education services?
Yes. The SGP norming sample includes students categorized as English learners and participating in special education. However, much remains to be learned regarding how these students grow and whether it is reasonable to expect the same amount of growth as other students. To study this topic and better assist educators with goal setting, Renaissance is collecting special education categorical data with the assistance of Dr. James Ysseldyke (University of Minnesota). Future data-collection efforts will focus on English learners. If your district uses STAR and would like to contribute data to this research project, contact research@renaissance.com to learn more.

Are there other ways besides SGP to understand student growth?
Yes, there are many ways to understand student growth. Castellano & Ho (2013a) provide a fairly exhaustive list of methods. One approach is to calculate the change in a normative score such as a normal curve equivalent (NCE). NCEs provide a way of representing PR scores so they can be accurately averaged and compared with each other. Because NCEs are derived from percentiles, they measure growth in comparison to national norms. Positive NCE change means student achievement grew at a faster rate than the national average (an NCE gain of zero). Another widely used model is value-added. Scores from STAR Assessments can be used in such models.

References

Betebenner, D. W. (2009). Norm- and criterion-referenced student growth. *Educational Measurement: Issues and Practice, 28*(4), 42–51.

Betebenner, D. W. (2011). *A technical overview of the student growth percentile methodology: Student growth percentiles and percentile growth projections/trajectories.* Dover, NH: The National Center for the Improvement of Educational Assessment.

Betebenner, D. W. (2016). *An overview of time-dependent student growth percentiles (SGPt)*. Dover, NH: The National Center for the Improvement of Educational Assessment.

Betebenner, D. W., VanIwaarden, A., Domingue, B., & Shang, Y. (2016). *SGP: Student growth percentiles and percentile growth trajectories.* (R package version 1.5-0.0).

Castellano, K. E. & Ho, A. D. (2013a). *A practitioner's guide to growth models.* A paper commissioned by the Technical Issues in Large-Scale Assessment (TILSA) and Accountability Systems & Reporting (ASR) State Collaboratives on Assessment and Student Standards, Council of Chief State School Officers.

Castellano, K. E., & Ho, A. D. (2013b). Contrasting OLS and quantile regression approaches to student "growth" percentiles. *Journal of Educational and Behavioral Statistics, 38*(2), 190–215.

Domaleski, C., & Perie, M. (2012). *Promoting equity in state education accountability systems*. Dover, NH: The National Center for the Improvement of Educational Assessment.

Fox, L., Carta, J., Strain, P., Dunlap, G., & Hemmeter, M. L. (2009). *Response to intervention and the pyramid model*. Tampa, Florida: University of South Florida, Technical Assistance Center on Social Emotional Intervention for Young Children.

Hamilton, L., Halverson, R., Jackson, S., Mandinach, E., Supovitz, J., & Wayman, J. (2009). *Using student achievement data to support instructional decision making* (NCEE 2009-4067). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. Retrieved from http://ies.ed.gov/ncee/wwc/pdf/practice_guides/dddm_pg_092909.pdf

Thurlow, M., Lazarus, S., Quenemoen, R., & Moen, R. (2010). *Using growth for accountability: Considerations for students with disabilities (Policy Directions 21)*. Minneapolis: University of Minnesota, National Center on Educational Outcomes. Retrieved from http://education.umn.edu/NCEO/OnlinePubs/Policy21

Independent technical reviews of STAR Assessments™

U.S. Department of Education: National Center on Intensive Intervention. (2016a). *Review of progress monitoring tools* [Review of STAR Early Literacy]. Washington, DC: Author. Available online from http://www.intensiveintervention.org/chart/progress-monitoring

U.S. Department of Education: National Center on Intensive Intervention. (2016b). *Review of progress monitoring tool*s [Review of STAR Math]. Washington, DC: Author. Available online from http://www.intensiveintervention.org/chart/progress-monitoring

U.S. Department of Education: National Center on Intensive Intervention. (2016c). *Review of progress monitoring tools* [Review of STAR Reading]. Washington, DC: Author. Available online from http://www.intensiveintervention.org/chart/progress-monitoring

U.S. Department of Education: National Center on Response to Intervention. (2010a). *Review of progress monitoring tools* [Review of STAR Early Literacy]. Washington, DC: Author. Available online from https://web.archive.org/web/20120813035500/http://www.rti4success.org/pdf/progressMonitoringGOM.pdf

U.S. Department of Education: National Center on Response to Intervention. (2010b). *Review of progress monitoring tools* [Review of STAR Math]. Washington, DC: Author. Available online from https://web.archive.org/web/20120813035500/http://www.rti4success.org/pdf/progressMonitoringGOM.pdf

U.S. Department of Education: National Center on Response to Intervention. (2010c). *Review of progress monitoring tools* [Review of STAR Reading]. Washington, DC: Author. Available online from https://web.archive.org/web/20120813035500/http://www.rti4success.org/pdf/progressMonitoringGOM.pdf

U.S. Department of Education: National Center on Response to Intervention. (2011a). *Review of screening tools* [Review of STAR Early Literacy]. Washington, DC: Author. Available online from http://www.rti4success.org/resources/tools-charts/screening-tools-chart

U.S. Department of Education: National Center on Response to Intervention. (2011b). *Review of screening tools* [Review of STAR Math]. Washington, DC: Author. Available online from http://www.rti4success.org/resources/tools-charts/screening-tools-chart

U.S. Department of Education: National Center on Response to Intervention. (2011c). *Review of screening tools* [Review of STAR Reading]. Washington, DC: Author. Available online from http://www.rti4success.org/resources/tools-charts/screening-tools-chart

Advisor

Dr. Damian Betebenner is senior associate at The National Center for the Improvement of Educational Assessment in Dover, New Hampshire. He specializes in applied statistics, and his current research focuses on longitudinal data analysis—specifically with regard to state and federal performance mandates.

Renaissance Learning™
P.O. Box 8036 | Wisconsin Rapids, WI 54495-8036
(800) 338-4204 | www.renaissance.com
R57137.60276.050516