Contents

Introduction..................................................................................................................................................................................... 3
Growth.............................................................................................................................................................................................. 3
Student growth percentiles........................................................................................................................................................ 5
Applying SGP to Star Assessments......................................................................................................................................... 6
Reliable and valid results ............................................................................................................................................................ 9
Reporting SGPs.............................................................................................................................................................................. 10
Sample characteristics .............................................................................................................................................................. 12
Frequently asked questions...................................................................................................................................................... 13
References..................................................................................................................................................................................... 21

Figures

Figure 1. Growth is better understood when performance history and peer group are considered......................... 4
Figure 2. Decision rules for score selection in SGP model................................................................................................. 7
Figure 3. Sample Dashboard screen........................................................................................................................................ 10
Figure 4. Sample Star Math Growth Report ........................................................................................................................ 11
Figure 5. Sample Star Early Literacy Growth Proficiency Chart ....................................................................................... 11
Figure 6. Sample view of Goal-Setting Wizard .................................................................................................................... 12
Introduction

Student achievement typically is gleaned from one score at a single point in time. When you consider growth in addition to achievement, it can greatly enrich your 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, one of your students may be performing at a low level, yet experiencing high rates of growth. Conversely, a high-performing student’s growth could be stagnating. It is important to know how a student is performing, but this information must have context—such as, how remarkable is this growth given a student’s achievement history?

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 (Domaelski & Perie, 2012). Let’s examine the student growth percentile (SGP) score—a widely accepted indicator of student progress used by states for purposes including instructional decisions and accountability reports—that is used in the Renaissance Star Reading®, Renaissance Star Math®, and Renaissance 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 you can capture and accurately interpret growth information, you are able to 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).

Figure 1 (next page) illustrates why we need context to fully understand differing rates of 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 jump by 1 inch. How should we interpret these gains? Whose growth was more impressive? At first glance, it may seem Athlete A has made greater improvement; however, to determine the significance of the athletes’ achievements, we must also consider their performance history and peer groups.

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. Having background information helps us know that the growth achieved by the expert Olympian was more impressive than the novice’s improvement. Without data about the growth that would be expected for each type of athlete, it is difficult to draw these conclusions.
Likewise 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.

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 now been adopted by a number of states for instructional and accountability purposes.

Star Early Literacy, Star Reading, and Star Math were the first interim tests to report SGPs. Growth models like SGP require an enormous amount of data to generate reliable results (Castellano & Ho, 2013a). Fortunately, widespread national use of Renaissance assessments provides a wealth of data, allowing SGPs to be reported for nearly all students in every grade, 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. 11.

1 Two exceptions: (1) first graders do not receive SGPs reflecting spring-to-spring, spring-to-fall, or fall-to-fall growth—these require at least one test score from kindergarten, which are not included in the SGP model for Star Reading or Star Math, and (2) scores for Star Early Literacy are included in the model only through third grade.
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 another 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 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 considered expected growth for one student may not be for another. An SGP of 50 can be thought of as typical growth for a particular student, given his/her grade and prior score history; however, state and local policy makers may define typical growth as a less precise range, such as 35 to 65 or 40 to 60 SGP.

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.

Distribution of scores

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, 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 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 this just means that 81 percent of the student’s academic peers from the same grade with a similar score history experienced more growth. SGP compares each student’s performance to that of a group of unique academic peers, which 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. Although we reference PR scores to illustrate points about achievement and growth, PRs are not used in the SGP calculation.
About Renaissance Star Assessments®

Star Assessments are reliable, valid, and time-efficient assessments of early literacy (Renaissance Star Early Literacy®), reading (Renaissance Star Reading®), and mathematics (Renaissance Star Math®) skills. Quick and accurate results from these assessments provide specific benchmarking, screening, progress-monitoring, and diagnostic information to help you 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.

Applying SGP to Star Assessments®

During the 2011–2012 school year, we 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 tests 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 assumed 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 Assessments as screening or benchmarking tests 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.

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. Our 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).

We specified date ranges for the windows while 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, and using data
Recent enhancements to the SGP model prioritize available data points to make the best use of information across time, by factoring a student’s current test score (the posttest) and up to two prior test scores (the pretest and, if available, an additional prior test) into the calculation:

- **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 SGP score a student receives (i.e., the period of growth it covers) is prioritized from top to bottom in the table, depending on available data. As the graph shows, when more than one test has been taken in an SGP window, the model selects the following prior 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 score selection in SGP model**

![Table showing decision rules for score selection in SGP model](image-url)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Fall–Spring</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Fall–Winter</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Winter–Spring</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Spring–Fall</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Spring–Spring</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Fall–Fall</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Fall–Spring</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Fall–Winter</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Winter–Spring</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Spring–Fall</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Spring–Spring</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Fall–Fall</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Test window dates are fixed, and may not correspond to the beginning/ending dates of your school year. Students will only have SGPs calculated if they have taken at least two tests, and the date of the most recent test has to be within the past 18 months.

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 possible: 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, scores from the posttest (current test) and pretest are used to determine growth, while an additional prior score serves to stabilize the student’s pretest score, minimize the impact of measurement error, 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 another pretest score, the SGP calculation uses this additional reference point only 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 the 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 now provides an even more complete picture of individual student growth.

For example, suppose two students have very similar posttest (current) and pretest scores. You might expect their resulting SGP scores to also be very similar. The scores may 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 adds context and stabilizes each student’s pretest score. In examining this additional data point, you 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 are academic peers, using an additional data point provides a more accurate picture of each students’ individual score histories.

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 to inform the usability of the SGP score and frequent updating of the SGP norming samples (a common practice for any norm-referenced score).

Adjusting for time

Our goal is to provide the best possible indication of how a student is growing, given the available data and research. Because ongoing research has demonstrated that adjustments to the SGP calculation will improve this

---

2 Standard error of measurement (SEM) is unavoidable and is present to some degree in all assessments. Assessment developers can seek only to minimize the impact of SEM. Tests with sound technical characteristics, such as Star Assessments, should reliably generate consistent and accurate estimates of student achievement. (For more on the value of adding an additional prior score to the SGP model, see the technical paper by Betebenner, 2016.)
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.\(^3\)

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 taken 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 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. 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 higher.

### Reliable and valid results

Each year since its initial development, we have reviewed the SGP model and made minor improvements to increase reliability and validity. Within Star tests, these advances yield results that are highly correlated across years, meaning you can confidently use all SGP results to inform both student goal setting and educator evaluation purposes.

In early 2016, Renaissance conducted an analysis of Star scores to understand how well 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.

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 its previous iterations, you will find that students’ SGP scores tend to fluctuate from test period to test period. Why might SGPs vary across time? You 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.

---

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

\(^4\) As expected, the results did not perfectly correlate, which would call into question the efficacy of model enhancements that produce precisely the same results.
We advise that you consider expert recommendations (e.g., Hamilton et al., 2009) regarding using multiple sources of data 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 allow it to provide an SGP for every student at the start the school year after the first fall test is taken (as long as data exists from the previous year). For example, if a student has a Star score from the prior spring and another test score from fall of the current school year (from the same assessment—e.g., Star Math), an SGP will generate indicating spring to fall growth. Having SGP scores available in fall allows you 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, and Growth Proficiency Chart.

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

The Star Growth Report (see figure 4, next page) 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 the Growth Report presents only the current SGP derived from the most recent Star administration, we advise that you generate and save this report on a periodic basis to have a record of SGP data.  

You may also view some historical SGPs within the Reading and Math Dashboards under the All-Time view. For more information about the Growth Report, see *Frequently asked questions*, p. 11.

---

5 *Note: Test dates on the Growth Report (figure 4, next page) apply to all scores shown except SGP. For SGP, dates displayed may not be the parameters used to determine the reported score. Thus, we set off SGP from other scores on the report and included a footnote explaining the information provided. Tests dates used to determine a student’s current SGP are shown on the Growth Proficiency Chart (see figure 5, also on next page).*
The Growth Proficiency Chart (see figure 5) is an interactive Star tool that displays data on the relationship between estimated proficiency and growth (expressed with SGPs). The chart shows whether students, classes, or schools are experiencing low proficiency and low growth, low proficiency and high growth, high proficiency and low growth, or high proficiency and high growth.

Figure 5. Sample Star Assessment® Growth Proficiency Chart

6 Star Reading and Star Math are statistically linked to most states’ summative assessments 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 begin testing students in grade 3, when students would typically take a Star Reading assessment.
The Goal-Setting Wizard (see figure 6, next page) helps you calculate and set goals for student performance, including scaled score and SGP. The wizard does not recommend goals for students; instead, it provides a tool to inform goal creation and monitor student progress against interventions and other initiatives.

**Figure 6. Sample view of Goal-Setting Wizard**

<table>
<thead>
<tr>
<th>Manage Goals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Define an intervention and set a goal</td>
</tr>
<tr>
<td>School: Oakwood Elementary School</td>
</tr>
<tr>
<td>Student: Farrell, Hope</td>
</tr>
<tr>
<td>Latest Test</td>
</tr>
<tr>
<td>12/9/2015</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Intervention Details</th>
<th>Interpretation &amp; Guidance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intervention Name</td>
<td>Guided Reading</td>
</tr>
<tr>
<td>Appears in report details</td>
<td></td>
</tr>
<tr>
<td>Goal End Date</td>
<td>3/25/2016</td>
</tr>
<tr>
<td>Used for SS/week calculation</td>
<td></td>
</tr>
</tbody>
</table>

**Goal**

Expected growth rate and score

Starting test: 12/9/2015 - 226 SS / 13 PR
(Set intervention line; starts trend and goal lines)

Select a goal type (based on students who scored similarly*)
- Moderate: 50 SGP = (press Calculate Goal button)
- Moderately Ambitious: 66 SGP = (press Calculate Goal button)
- Catch Up/Keep Up: (press Calculate Goal button)
- Stay Up/Move Up: (press Calculate Goal button)

Or define a custom goal:
- Scaled Score

*National growth data indicate that about 50% of this student’s academic peers (those with a similar score history in the same grade) were able to achieve or exceed the amount of growth indicated by the “Moderate” rate, and about 34% were able to achieve or exceed the “Moderately Ambitious” rate.

“Catch Up/Keep Up” refers to the amount of growth necessary for the student to reach the estimated threshold for Proficient by the time of the spring state test. “Stay Up/Move Up” refers to the amount of growth necessary for the student to reach the estimated threshold above Proficient.

**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. We update this information throughout the year, and 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 score information using a statistical model of growth and achievement called quantile regression. The model is designed to prioritize data points to make the best use of data across time. The 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 a prior test window, the following test scores are selected as priors:
    - 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 a student’s scores with those of his/her peers (by taking into account an additional prior score for the student from the previous school year), the SGP model now provides 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 (from the previous school year) in the SGP calculation helps to precisely identify 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 seek only to minimize the impact of SEM. Tests with sound technical characteristics, such as Star Assessments, should reliably generate consistent and accurate estimates of student 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 Assessments 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, one prior-year test score). Some historical SGPs can be viewed within the Reading and Math Dashboards under the All-Time view. We also advise that you 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. For SGP, dates displayed 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 SGP model each year, scores can be compared over time. In 2016, to study comparability of the scores, Renaissance 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 that produce precisely the same results.)

Because of the important role SGP scores play in instructional and accountability decisions, both Renaissance 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 improve the model’s 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 have 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?
You may expect to see highly consistent SGPs for a given student within or across school 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, we advise that you consider expert recommendations (e.g., Hamilton et al., 2009) regarding using multiple sources of data 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.

Does whether a test is adaptive versus fixed form make a difference in the interpretation of growth with the SGP score?
SGP scores have the same basic interpretation regardless of the type of test taken. The SGP calculation for Star Assessments (which are adaptive) was developed by Dr. Damian Betebenner, who also worked with numerous state summative assessment programs (which are mostly fixed form) to report SGP. Therefore, interpretation of SGP is the same across grade, subject, and assessment, assuming Dr. Betebenner’s model is followed.

One consideration when interpreting SGPs with Star Assessments is that the timeframe of growth characterized by the SGP score may differ between Star tests and summative assessments. Star Assessments can be given throughout the year, whereas state summative tests are usually restricted to once per year, often in the spring. Therefore, Star SGP may reflect growth from any number of time periods (e.g., Fall to Spring, Spring to Fall, Fall to Winter), while state summative SGPs generally reflect Spring to Spring growth. If questions arise about the period of growth represented by a student’s SGP, check the Star Growth Proficiency Chart for dates of the test scores used to calculate SGP.
I have heard it can be difficult to differentiate growth in older (e.g., high school) student populations. Does that make SGP less useful or valid for secondary students?

No, the process to calculate SGP and the interpretation of its scores is consistent across all grades. It is well documented that growth over time is not perfectly linear; the rate of student achievement growth tends to slow as students get older. However, this phenomenon does not impact the interpretation of SGP nor threaten its validity. Although high school students tend to grow less than elementary students, in absolute scaled score change from beginning to end of year, there is enough variability in growth in the upper grades for a model such as SGP to work just as well as it does in earlier grades. The SGP calculation takes into consideration grade level, meaning students are compared only with academic peers within their current grade.

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, and Growth Proficiency Chart. Some 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. For SGP, dates displayed may not be the parameters used to determine the reported score. Actual test dates are shown on the Growth Proficiency Chart; see p. 13.)

Will I see an updated SGP immediately after testing?

Student data will populate overnight and reporting will reflect an updated SGP the day following testing.

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 may result in no SGP score being reported.) This also applies to other cross-year SGPs for Fall-Fall and Spring-Spring time periods.

• Student data populates overnight, and a minimum of one assessment in at least two consecutive SGP windows (from the same assessment—e.g., two test events in Star Reading) must be taken in order to calculate an SGP score. Thus, if the report is being viewed the same day the test was taken, an SGP score may not be visible.

• SGP score data is reported based on the decision rules displayed in the table (next page), with the type of score received prioritized from top to bottom, depending on available data. Score history that deviates from what is shown here may result in the program being unable to report an SGP score.
**Decision rules for score selection in SGP model**

<table>
<thead>
<tr>
<th>Most Recent Test Is</th>
<th>Type of SGP Calculated</th>
<th>Test Windows in Prior School Years</th>
<th>Test Windows in Current School Year*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fall</td>
<td>Fall-Spring</td>
<td>Fall 8/1-11/30, Winter 12/1-3/31, Spring 4/1-7/31</td>
<td>Fall 8/1-11/30, Winter 12/1-3/31, Spring 4/1-7/31</td>
</tr>
<tr>
<td>Fall</td>
<td>Fall-Winter</td>
<td>Winter 8/1-11/30, Spring 12/1-3/31, Fall 4/1-7/31</td>
<td>Winter 8/1-11/30, Spring 12/1-3/31, Fall 4/1-7/31</td>
</tr>
<tr>
<td>Winter</td>
<td>Winter-Spring</td>
<td>Winter 8/1-11/30, Spring 12/1-3/31, Fall 4/1-7/31</td>
<td>Winter 8/1-11/30, Spring 12/1-3/31, Fall 4/1-7/31</td>
</tr>
<tr>
<td>Spring</td>
<td>Spring-Fall</td>
<td>Spring 8/1-11/30, Winter 12/1-3/31, Fall 4/1-7/31</td>
<td>Spring 8/1-11/30, Winter 12/1-3/31, Fall 4/1-7/31</td>
</tr>
<tr>
<td>Spring</td>
<td>Spring-Spring</td>
<td>Spring 8/1-11/30, Winter 12/1-3/31, Fall 4/1-7/31</td>
<td>Spring 8/1-11/30, Winter 12/1-3/31, Fall 4/1-7/31</td>
</tr>
<tr>
<td>Fall</td>
<td>Fall-Fall</td>
<td>Fall 8/1-11/30, Winter 12/1-3/31, Spring 4/1-7/31</td>
<td>Fall 8/1-11/30, Winter 12/1-3/31, Spring 4/1-7/31</td>
</tr>
</tbody>
</table>

*Test window dates are fixed, and may not correspond to the beginning/ending dates of your school year. Students will only have SGP calculated if they have taken at least two tests, and the date of the most recent test has to be within the past 18 months.

---

**Why do I see an SGP in the fall?**

Recent improvements to the SGP model allow teachers to start the school year with an SGP for all students (who have data from the previous school year). For example, if the student has a Star score from the prior spring and another test score from fall of the current school year (from the same assessment—e.g., Star Math), an SGP will generate indicating spring to fall growth. This feature provides teachers with immediate information about students’ growth history to provide insight and assist with initial instructional decisions.

**How do Star Assessments taken in the summer months affect a student’s fall SGP score?**

If you test students with Star Assessments during summer school, and that summer testing takes place on or after August 1, SGP projections and calculations for the new school year may be based on a test which was not intended. So that SGP information is based on the first test of a new school year, we recommend the following options:

**Option 1:** Do not use Star Assessments after August 1, until the first screening window of the new school year. This will ensure that (1) the first test of the year is linked to the student in their new grade, (2) SGP calculations will anchor to this test, and (3) SGP growth projections are used correctly in multiple reports/functions.

**Option 2:** Start your new school year in Star on August 1, and promote students at that point. Keep in mind that the first test of this new school year will be used as the Fall test for both SGP calculation and growth projections on Star reports.
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 and Star Math, as sufficient sample size must be established in order to compute SGP scores. Renaissance recommends that educators test students with Enterprise versions of 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, exercise caution when interpreting results.

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

To get an SGP score, 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 a final time to obtain an SGP.

How can my student be at the 95th percentile and have a 19 SGP? Is SGP accurate for high-achieving kids?

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 to have an SGP that is relatively low, typical, or relatively high.

Let’s examine a student with a fall PR 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 this SGP indicates that 81 percent of the student’s academic peers (from the same grade with a similar scaled score (SS) history) experienced more growth. Again, no matter how high, low, or middle-of-the-road a PR score, each student has the same chance of receiving any SGP score from 1–99. (Note: Although we reference PR scores to illustrate points about achievement and growth, PRs are not used in the SGP calculation. No assumptions can or should be made about a student’s SGP based on PR performance.)

What is typical growth?

Renaissance does not provide a benchmark for typical growth; however, many states that have adopted SGP consider 35–65 SGP as typical growth. 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?

Percentile rank (PR) and student growth percentile (SGP) are very different metrics (see table, next page). PR is an achievement (i.e., performance) score that describes a single point in time while SGP is a growth measure that explains student growth between points in time. Both measures are norm-referenced but have different norming groups—for PR, this group includes all students in a particular grade level, and for SGP, each student has a unique academic peer group.

Because both scales span 1–99, score interpretation is similar between PR and SGP: lower numbers indicate lower relative achievement/growth and higher numbers indicate higher relative achievement/growth (e.g., an SGP of 75 means the student’s growth exceeds that of 75 percent of students, just as a PR of 75 means a student’s achievement exceeds that of 75 percent of students).

Although these scores can be interpreted similarly, that does not mean a student with a high PR score will likely receive a high SGP score. A high PR score means a student is performing well at a certain point in time, but no
matter how high or low a student’s PR, there is an equal chance for that student to receive any SGP score from 1–99. SGP compares a student’s performance to students in the same grade with a similar scaled score history, and this unique group of academic peers is recalculated each time an assessment is taken. No assumptions can or should be made about a student’s SGP based on PR performance, and PRs are not used in the SGP calculation.

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

**What does it mean when a student has a high PR and a low SGP?**

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 how students are performing at a single point in time but are only a piece of the puzzle. It is also important to know how students fare over time in relation to their peers, a question comparative growth data from SGP can help answer. For example, consider a student who begins and ends the year at the 50th PR with an SGP of 30. This student has consistently performed better than 50 percent of students in the same grade nationwide; however, the SGP score shows that the student has grown more than just 30 percent of students in the same grade with a similar score history.

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

**In 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. We advise that you consider expert recommendations (e.g., Hamilton et al., 2009) regarding using multiple sources of data 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 do I get SGP scores based only on students in my state?**

Growth models like SGP require an enormous amount of data to generate reliable results (Castellano & Ho, 2013a). We examine data for students nationwide so that our sample size is adequate 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 that report SGPs on their summative tests, Renaissance reports median SGP. We recognize, however, recent research on this topic (Castellano & Ho, 2013b) concludes it may be appropriate to use mean or median. You may want to consult your state’s position 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 may provide misleading estimates of central tendency, depending on distribution of scores and group size; thus, some states do not 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 as participating in special education. However, much remains to be learned about how these students grow and whether it is reasonable to expect their growth to be comparable to other students. So that Renaissance can study this topic and better assist educators with goal setting, we are 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 Assessments 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. 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


Independent technical reviews of Star Assessments®


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.