Using STAR™ for Educator Evaluation: Frequently Asked Questions (FAQ)

Background Questions

What is the background behind the current, nationwide conversations regarding Teacher Evaluations?
Teachers have always been evaluated, usually by way of observation (a qualitative metric). How often, how rigorously, and with what consequences, has varied considerably from state to state and district to district. With the Race to the Top competitions, and the Waiver Requests associated with the delayed reauthorization of the Every Student Succeeds Act (ESSA), states are requiring that teacher evaluations follow explicit formulas incorporating student achievement and/or growth (a quantitative metric), and that evaluation results are included in personnel decisions.

Are Educator Effectiveness and Teacher Evaluation the same thing?
We believe that the purpose of educator evaluation is to support effective teaching and improve student learning. Educator effectiveness refers to the degree of quality in specific dimensions of teaching, such as classroom management and a deep understanding of how learning progresses in a discipline. Educator evaluation, generally a summative measure determined from an examination of quantitative as well as qualitative data, refers to a degree of quality in overall performance and achievement for a particular school year or span of years. Effectiveness for all educators is the goal. Evaluation, as required by local, state, and federal legislation, is one pathway to achieve that goal, and is the focus of this document.

As an interim assessment, STAR provides educators with reliable data during the year so they can see the path ahead in time to impact it. This document will show how STAR can provide teachers with critical data for documenting instructional practice and building a body of evidence of student growth and achievement as part of their district’s educator evaluation process. STAR assessments can be administered multiple times throughout the school year, creating a trustworthy trend line that tells a story. Teachers and principals can demonstrate student progress toward incremental benchmarks, mid-course corrections in instruction and the resulting effects, efforts to screen and identify students in need of (and in response to) intervention, trends toward state proficiency, and patterns in learning. For purposes of this FAQ, we are generally talking about how measures of student growth are being used in teacher evaluation.

Why is Renaissance Learning involved in Educator Effectiveness/Teacher Evaluation?
Our customers know that the STAR assessments provide data that has been proven reliable and valid for many purposes, including measuring student achievement and growth. STAR assessments are approved by many states and districts for high-stakes purposes including serving as an indicator of student growth in educator evaluation formulas, grade promotion policies, as well as Gifted & Talented identification. For any assessment purpose, our goal is to help educators use their STAR data effectively and responsibly.

In what states has STAR been approved for use as a student growth measurement?
STAR assessments have been included in “approved lists” for use in educator effectiveness and/or as student growth measurements in New Mexico, New York, Ohio, Tennessee, and Virginia. STAR data is also being used in many other states where there is local choice in assessments.

What if my district tells me something different than what Renaissance Learning suggests?
This information is not intended as a recommendation or to invalidate district policy. Educators should always consult and follow district guidelines.
Questions Related to the use of STAR™ in Educator Evaluation

How can STAR data be used in Teacher Evaluation?
In this and other support documents we provide examples of ways that educators may choose to use STAR data in the process of measuring student growth for the purposes of teacher evaluation. Oftentimes this “impact on student learning” score is a component of the “quantitative” side of the evaluation formula, and is merged in a prescribed way with other scores. In many states, this component of the formula takes the form of a student learning objective (SLO), particularly for teachers of grades and subjects not included in state tests. See Considerations for Using STAR Data with Educator Evaluation for additional information.

Please note that Renaissance Learning makes suggestions only (not recommendations) for how data can be used based upon research, best practices, consultation with education experts, and work with schools nationwide. Educators are always cautioned against placing too much emphasis on any one evidence source, particularly in high-stakes decision making (e.g. teacher evaluation).

What is the purpose of STAR? How does this relate to high-stakes decision-making?
Educators use assessments for different purposes. Some assessments can only be used for a single purpose, while others, like STAR, can meet many different needs. Many of the uses of STAR are instructional—helping teachers understand what students know and what they’re ready to learn next, how much they’re growing, or whether they are responding adequately to instruction. But as educators are fully aware, states and districts have been using assessment results for other, higher stakes decisions (e.g. indicators of growth in educator evaluation formulas, grade-promotion decisions, and Gifted & Talented identification). That STAR is often approved for these uses provides further reinforcement that the assessment meets high technical standards.

When is it appropriate to use STAR as an assessment for Educator Teacher Evaluation?
STAR assessments may be appropriate to use for teacher evaluation when they are a good match for the content students are supposed to have learned and when they align with specific state and district policies for demonstrating such learning. For example:

- If a teacher wants to measure early literacy skills development, STAR Early Literacy can be an appropriate assessment for grades K-3.
- If a teacher wants to measure reading comprehension development in grades 1–12, STAR Reading can be an appropriate assessment.
- If a teacher wants to measure math computation and/or math reasoning in grades 1–12, STAR Math can be an appropriate assessment.
Growth-Measurement Questions for Teacher Evaluation

**What kind of growth measurements are built into STAR?**
Renaissance Learning suggests that for purposes of high-stakes decision making such as teacher evaluation, student growth should be measured with psychometrically sound growth scores such as these. Student Growth Percentile (SGP) is an acceptable source of evidence for student growth for student learning objectives (SLOs) and other elective sources of data. With the updated SGP model, STAR is now even better at measuring within-year growth for educator evaluation purposes. STAR assessments continue to fully meet the requirements for measuring within-year growth for educator evaluation, including in states where it is approved for this purpose.

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 more information on SGP, see the Educator Effectiveness/Educator Evaluation FAQ and the SGP FAQ.

For SGP to be reported in STAR, students must be tested within at least two of the following date ranges:

- Fall: August 1 – November 30
- Winter: December 1 – March 31
- Spring: April 1 – July 31

**Student Growth Percentile (SGP)**

- Student growth percentiles are reported at the student level so they can be used for student learning objectives (SLOs) where each student (or a percentage of students) must meet a growth target. SGP growth targets for SLOs using STAR data are set via a local decision-making process. Renaissance Learning offers general considerations and reference points to help frame decision making in this area. See Using STAR Data for SLOs (and other questions in this document) for additional information.
- The growth target feature in STAR includes the student growth percentile metric. District-set student growth targets are reported on the Growth Report.
- SGPs are also reported at the group level via ‘median SGP’ on the Growth Report. Median SGP is sometimes used in states that don’t require individual SLOs as a component in their teacher evaluation formula.

**My state uses value-added. Can I still use STAR SGP in my evaluation?**
Yes, STAR SGP can be used to report student growth for teacher evaluation in states that report value added on the state test. Student growth percentile and value-added are both widely accepted growth models that facilitate looking at student growth from different angles. For more information on the similarities and differences between these models, see the SGP FAQ.

**My state uses SGP. Can I still use STAR data in my evaluation?**
Yes, STAR SGP can be used to report student growth for teacher evaluation in states that report SGP on the state test. STAR SGP is very similar to the SGP model typically used in state reporting. Both are based on the work of the lead developer of SGP, Dr. Damian Betebenner.

These are some of the differences between state-reported SGP (summative) and STAR SGP (interim):

- The state test and the STAR assessments are different tests.
- The amount of time between tests is different. SGP is a “within-year” measure (fall to winter and fall to spring); state-test SGP is an “across year” measure, typically spring to spring.
- The norming groups are different. With STAR SGP, students are compared to other students in the same grade nationwide who take STAR tests. State-test SGP compares students in the same grade in the same state.

For more information on STAR SGP, please see the SGP FAQ.
Is there a difference between growth and change? Which does STAR report?
There are important distinctions between growth and change. For example, scaled score change can be negative or positive, while SGP values which represent growth, will always be positive (1-99).

STAR reports both change and growth. Change is reported in a number of ways in STAR (e.g. change in scaled score, change in PR score, change in GE score, etc.). STAR reports growth in student growth percentiles (SGP) and growth norms, which are both norm-referenced growth measures. SGPs quantify growth for a student compared with his or her academic peers, whereas the STAR growth norms quantify growth with reference to the typical growth of similar students in a nationally representative sample. For more information about change and growth scores, please see the SGP FAQ. For more information about STAR scores, we suggest the technical manuals for STAR Reading, STAR Early Literacy, and STAR Math.

Can/Should I use the Screening Report to set growth goals for Teacher Evaluation?
For setting growth goals, we suggest that educators use a statistical growth model such as SGP. Because the STAR assessments report SGP (via the Growth Report), the Screening Report (i.e. the pretest score) is not necessary for setting a growth target.

At the same time, Fall Screening does help establish a baseline, or starting point, for the learning that will take place during the school year. The Screening Report indicates where students are at that point in time in relation to relevant benchmarks. Teachers and administrators use this information for instructional planning and resource allocation.

What do I use for baseline when using SGP for my growth target?
Typically, educators have looked to the pretest score found on Screening Report as a starting point for setting growth targets. With SGP however, the growth target is not contingent on a student’s starting score. SGP is a research-based growth metric which simplifies the target-setting process and provides a better and more meaningful baseline. With an SGP growth target, students are compared not just to their own starting scores, but to a nationwide, normed sample of students who started with the same pretest or starting scores.

How do I know how much growth to expect from my students?
The setting of SGP growth targets using STAR data is a local decision-making process. Renaissance Learning offers the following general considerations and reference points to help frame decision making in this area when there is not a preset target.

- Educators first consult with their district for any district-determined growth targets.
- STAR SGP data from previous years, when available, should also be consulted, noting any trend information.
- If your state reports SGP on the annual summative test, you may want to consult that report for state-level guidelines on SGP growth expectations. (See the related question above for differences between summative and interim SGP.)
- If you have no guidelines from your state or district on setting SGP targets, refer to Typical Growth Defined by States for a listing of what is considered “typical growth” in many states. Note that many states that report SGP consider the 35-to-65 SGP range to be typical growth.
- Though there is no fixed definition of what a year’s growth is, many states that use SGP as part of their accountability system consider an SGP of 50 to denote a year’s growth. When using SGP for SLO’s, you may consider a range around 50 SGP such as 35 to 65 SGP.
- While educators want all students to learn and grow, for purposes of teacher evaluation, typical growth guidelines can be helpful in setting reasonable and attainable goals.
- Whenever possible, educators may want to consider other sources of information when setting “reasonable and attainable” growth targets, such as how much students grew in the recent past, extenuating circumstances in recent years including changes to curriculum and instruction, professional development initiatives, and personalized professional growth plans. In other words, while the assessment administered and the growth metric used for target setting and growth measurement should be of national caliber, baseline data is appropriately a more localized concern. For more information, see Viewing and Setting Growth Targets.
How will I know if students are “on track” to meet their SGP growth targets?

Midyear analysis focused on Growth Report data provides an excellent opportunity for checking to see if students are on track to meet their growth targets. After students test with STAR during the SGP Fall and Winter windows, a mid-year SGP score will be found on the Growth Report. This mid-year SGP is especially valuable because it offers insight into students’ growth during the year, while there is still time to impact it.

The STAR Growth Proficiency Chart is another valuable tool for mid-year analysis and checking to see if students are on track. This chart is made up of four quadrants, with the lower left hand quadrant indicating students who are below benchmark and are also demonstrating low growth.

Administrators will also want to run a Student Growth Extract, another feature in STAR, to see more detailed information about how students are growing. This report provides student growth targets and the corresponding scaled scores students will reach if they hit their targets.

Can I use domain scores to measure growth? How do I set growth targets with domain scores?

Renaissance Learning recommends against the use of domain and/or skill-set scores, or other scores that indicate absolute change for measuring growth. Although these scores—which indicate degrees of mastery of certain content—may seem like appealing measures of students’ knowledge and skills, they are far from ideal for use as growth measures, for the following reasons:

- It is important to keep in mind that each domain score is derived from the scaled score. Using multiple domain scores, or using a domain score and a scaled score, is double counting the assessment, which is a highly questionable practice.
- There are a multitude of domain and skill-set scores to choose from, which can create confusion. In addition, the reported values of domain and skill scores will vary both within and across grades, and the content of the skill areas varies from grade to grade. To use any of these as the basis for measuring change, it would be necessary to choose which one or ones to use; if a decision were made to use more than one within a grade, a method of combining them would be needed, but there is no clear rationale for doing so.
- When measuring growth for educator effectiveness or other purposes, it seems far more appropriate for educators to use growth metrics that come from measures designed for measuring growth, such as student growth percentile and the weekly growth rates.

Can I use STAR Early Literacy to measure growth in early numeracy skills?

Renaissance Learning recommends against the use of STAR Early Literacy as an assessment of growth in early numeracy skills. STAR Early Literacy reports an early numeracy (EN) score on several reports. This score is derived from the scaled score and represents a student’s expected ability to identify and name numbers; understand number-object correspondence; complete sequences; compose and decompose groups of up to ten; and compare sizes, weights, and volumes.

While the early numeracy score does provide the teacher with some information about the student’s skill development, it is not intended to be used as a growth score. STAR Early Literacy measures a variety of skills expressed as domain scores. While early numeracy is one of the domains that is assessed, domain scores themselves are not intended to be used for growth measurement, as explained in response to the “Can I use domain scores to measure growth?” question above.

When could data from district created assessments be used for educator evaluation?

Data from district created assessments that are delivered through STAR Custom could be used to gauge growth for educator evaluation in states that allow districts to develop assessments for use on a district-wide basis by all teachers in a given grade or subject area. In other words, when a state’s educator evaluation policy allows, a district could deliver assessments created for this purpose district-wide via.
Because of the design of STAR Custom, the tool allows for the delivery of assessments that produce criterion-reference scores. The criterion-referenced data from districtwide, carefully-constructed assessments could be used for educator evaluation. Below are guidelines for creating districtwide educator evaluation assessments that are delivered through STAR Custom. For more information, see Using District Created Assessments for Educator Evaluation Delivered via STAR Custom™

Looking Ahead

**Can other Renaissance Learning programs be used in Educator Effectiveness and/or Teacher Evaluation?**

At this point in time, we are providing support for educators who wish to use STAR data as a quantitative measure of student growth and achievement within their states’ teacher evaluation formulas.

Certainly, other Renaissance Learning programs also provide qualitative evidence of effective teaching and student learning. We have a long history of helping educators measure the occurrence of learning and use data to raise achievement. We are currently investigating opportunities to support educators with additional quantitative and qualitative information that will further the very intention of teacher evaluation—to increase both educator effectiveness and student learning.

**Disclaimer of Liability**

The information contained in this document is solely for educational purposes and marketing. Renaissance Learning, Inc., its affiliated companies and their respective owners, officers, employees, agents (collectively, “Renaissance”) make no, and disclaim all, express and implied representations or warranties, including, without limitation, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE, concerning STAR, the information or materials described herein or any uses made thereof (including evaluation or assessment), or the results or consequences derived or resulting from the foregoing. Although this document has been produced from sources believed to be reliable at the time, no representation or warranty, express or implied, is made regarding the accuracy, adequacy, timeliness, completeness, legality, reliability, or usefulness of any information contained herein.

Each user of STAR must determine for itself the appropriateness and advisability of using STAR, the information or materials described herein, or any output, work product, analysis, results, or conclusions derived or resulting, in whole or in part, therefrom, and assumes sole liability therefore. Each such user waives to the fullest extent permitted by law any and all claims, demands, suits, causes of action, and other rights against Renaissance with respect thereto (regardless of whether any such claim, etc. is based upon contract, tort, negligence, strict liability, common law, statutory, or other theory of recovery). Each such user agrees that it (and not Renaissance) shall be responsible for any claims, demands, suits, causes of action, losses, damages, and other costs and expenses asserted by any third party affected by user’s use of the foregoing.