Offeror Name: Renaissance Learning™, Inc.
Proposed Assessment Name: STAR Math Enterprise™
Content Area(s) and Grade Level(s) Assessed: Mathematics, grades 1–12

Section 1: Overview of Tests

Requirement: 1.1
1.1 Describe the specific grade(s) and subject area(s) covered by each assessment and provide an overview of the content and skills measured. Include the types of test items used, the mode(s) of delivery, the availability of equivalent forms, including short forms or screeners (if available) and a test blueprint for each test being proposed.

STAR Math Enterprise is an integrated, multi-faceted assessment that consists of the following: 1) an online computer-adaptive skills-based interim test that measures math achievement for students in grades 1–12, and 2) a classroom-based assessment system called Core Progress™ for Math, an empirically validated learning progression that helps teachers bridge assessment and instruction by guiding them to appropriate instructional resources—including performance tasks and constructed response items—that are designed to assess students' deeper understanding of key concepts. Both assessment components are described in more detail below.

STAR Math Enterprise. Students taking the computer-adaptive STAR Math Enterprise test receive 34 questions in four broad domains: Numbers and Operations; Algebra; Geometry and Measurement; and Data Analysis, Probability, and Statistics. These four domains form the basis of the Core Progress™ learning progression for math described below. Within each domain, skills are organized by sets of closely related skills. The STAR Math Enterprise item bank includes more than 5,000 carefully calibrated items that test a total of 556 skills, with multiple items available to measure each skill.

Core Progress for Math. Within the STAR Math Enterprise software, educators have access to the Core Progress learning progression for math. This learning progression provides descriptions of the incremental learning that needs to take place for students to make progress as well as instructional resources that foster this growth. Once students complete a STAR Math Enterprise test, teachers have valuable information about how students are progressing toward a multitude of outcomes. With Core Progress, teachers are able to identify specific materials that support differentiated instruction for students, based on what they are ready to learn next. Core Progress provides a classroom-based assessment system that not only facilitates student practice of these skills, but also allows educators to assess each student's deeper, conceptual understanding of them.

The Core Progress for math skills list provided in appendix 2 contains a complete list of the skills assessed by STAR Math Enterprise and Core Progress resources. The skills are organized by both domain and grade level. The domain view shows the progression of grade-level skill statements from grade to grade within the domain, and it is designed to provide a better understanding of the progression of skills within each domain. The grade-level view shows all of the grade-level skill statements for a particular grade, organized by domain. Use this view to understand which skills are associated with a

particular grade.

Types of Items

The computer-adaptive STAR Math Enterprise assessment provides a comprehensive assessment solution that utilizes highly efficient selected-response (multiple-choice) items. The decision to use the multiple-choice assessment item format was based on several considerations, including its research-proven validity and effectiveness; its accessibility and ease of use by a broad range of students, including ELLs and students with various types of disabilities; its ease of scoring, particularly in the context of computer-based tests; and especially its high degree of efficiency in delivering a high level of information about student skills and abilities in a minimum amount of time, and with minimal disruption of instructional and learning time.

The Core Progress interface includes multiple item types, including performance tasks that are designed to gauge students' deeper understanding of key concepts. Through Core Progress, educators have access to performance tasks for each grade level. Each of these tasks is designed to help students transition from learning and practicing skills in isolation to using them in an integrated way that reflects the highest Depth of Knowledge (DOK) levels. In addition to the performance tasks, Core Progress also include sample items designed to probe individual skills, and teacher activities that educators can use to teach specific math skills. Content is available for all grades so that teachers can assess students working at any grade level. This could be especially beneficial for students in Virginia schools who are working above or below grade-level benchmarks.

Modes of Delivery

The STAR Math Enterprise assessment is web based and easy to use. Each test runs on the Renaissance Place Real Time™ platform, a web-hosted system-wide management program that consolidates all Renaissance Learning software.[1] Accessing the Real Time platform requires a user name and password. To take a STAR assessment, students simply log in on an Internet-connected computer at the school and take the test. Afterward, the teacher or administrator logs in on an Internet-connected computer and runs easy-to-read reports.

Equivalent Forms

STAR Math Enterprise is a computer-adaptive assessment based on item response theory (IRT). Thus, every administration of STAR is considered an "equivalent form" because each administration is equally as difficult as other administrations, both across students and for each individual student. The National Center on Response to Intervention (NCRTI) has given the STAR assessments its highest rating for "equal alternative forms."[2]

It is important to note that, as a computer-adaptive assessment, STAR Math differs from conventional tests in its approach to multiple forms. STAR's adaptive-branching mechanism individually tailors each student's test to items the student can answer correctly a specified percent of the time. To accomplish

[1] For information on minimum and recommended technical requirements of this platform, please see the document titled *Renaissance Place Real Time Technical Recommendations* in appendix 2.

[2] These ratings can be viewed on the NCRTI's website at http://www.rti4success.org/tools_charts/popups_progress/STAR_Math_alternateForms.php.

this, STAR Math continuously updates an estimate of the student's developed ability and selects test items one at a time from a pool of thousands of items. STAR Math applies item response theory to accomplish its percent difficulty target. The key is having accurate measures of each item's IRT difficulty parameter, as well as an accurate measure of each student's current ability.

The *STAR Math Enterprise Technical Manual* provided in appendix 5 documents the reliability of the STAR Math assessment overall and by grade. Evidence for STAR Math's alternate form reliability can be found in the technical manual (pages 43–44).

Test Blueprint

A test blueprint for STAR Math Enterprise is provided in appendix 5.

Requirement: 1.2

1.2
Provide evidence of alignment of test items to the Virginia Standards of Learning (SOL) for existing assessments. For assessments developed in response to the RFP, provide a plan for assuring the alignment of test items to the SOL.

The alignment reports included in appendix 5 provide evidence that the skills assessed by STAR Reading Enterprise align to the Virginia Standards of Learning and to Virginia's 2011 College and Career Ready Performance Expectations for Math, grades 9–12.

In addition to the standards content alignment, Renaissance Learning uses a rigorous five-step approach to align Virginia standards to the STAR scale in order to determine estimated difficulty levels of the standards. The difficulty level calculations are used on the STAR State Standards Reports. (See the document titled *Key Report Samples: STAR Enterprise*, included in appendix 2 for a sample State Standards Report.) This approach is described below:

- **Identify standards.** Each standard is unpacked and the skills, concepts, and expectations inherent in the standard are identified. After review of the standards, Renaissance Learning identifies which standards will be included on the STAR State Standards Reports. A subset of standards is included to give educators an actionable, not overwhelming, amount of information.

- **Identify the STAR skills and the items for those skills that assess the standard's expectations for the skills and concepts.** The STAR Math Enterprise item bank has more than 5,000 items, all of which were created using a rigorous, research-based item-development process. As part of this process, the Virginia Standards of Learning are reviewed and items are written to address the skills identified within these standards. Each STAR skill is analyzed and compared to the standards. Through this process, the skills in the items are aligned to the skills in the standards. Where they match, an alignment is made.

- **Review the calibrated difficulty level of the STAR items associated with the skills and concepts embedded in the standard.** Each STAR test item is calibrated to determine its exact point on the STAR difficulty scale. This is done by administering test items to large samples of students, collecting student-response and other data, and performing a statistical analysis of the

Requirement: 1.2

response data to determine the scale values.

- **Assign a difficulty level to the standard based on the review of this empirical data.** This difficulty level is equated to a scaled score on the STAR scale, which ranges from 0–1400. An Estimated Mastery Range identifies a band of scores just below or above the difficulty level of the standard. Students who fall within the Estimated Mastery Range should be monitored to confirm their understanding of that standard.

- **Review the assigned STAR difficulty level.** A content-area expert analyzes the assigned difficulty level, which is expressed on the same scale as STAR test scores, in relation to the Virginia Standards of Learning to ensure that the placement is accurate and appropriate. Renaissance Learning's standards experts combine empirical data with educator expertise to review the accuracy and reasonability of the difficulty level assigned to each standard. This review includes analyzing each standard for its relative skill difficulty rating within the Core Progress learning progression and then comparing the standard to other standards within the same grade level and to similar standards in higher and lower grade levels.

Section 2: Technical Characteristics

Requirement: 2.1

2.1
Provide evidence of content, construct, concurrent, and predictive validity as appropriate. Include validity evidence that supports the use of scores from the proposed assessment in teacher evaluation, addressing specifically the validity of using assessment results to support inferences about effectiveness of teacher in producing growth in student performance (if available).

The *STAR Math Enterprise Technical Manual* (pages 57–88) provided in appendix 5 contains detailed evidence of content, construct, concurrent, and predictive validity for the assessment. Below we have provided summaries of this evidence. In addition, we describe how assessment results may be used to support inferences about teacher effectiveness in producing growth in student performance.

Content Validity

The content for STAR Math Enterprise is based on analyses of state and college- and career-ready standards (including the Virginia Standards of Learning), curriculum materials, test frameworks, and content-area research, including best practices for mathematics instruction.

In the *STAR Math Enterprise Technical Manual* (pages 14–25), content-related validity evidence is contained in the extensive treatments of the content of the item banks; the rationale and research basis for selecting that content; and the methodologies employed to develop, evaluate, and select test items. Content-related validity evidence is also provided to address the methods used to align skills and items to a wide variety of standards, including but not limited to state curriculum standards and college- and

Requirement: 2.1

career-ready standards.

Criterion-Related and Construct Validity

In the *STAR Math Enterprise Technical Manual*, criterion-related and construct validity evidence is contained in numerous tables of correlations between STAR assessments and other measures of the same skills and abilities, including teacher ratings and many widely used tests of reading, English language arts, and math (see pages 57–83). These correlations include both nationally published standardized achievement tests and state accountability tests.

Evidence for the criterion-related and construct validity of the STAR assessments is based on several different types of data, including correlations with teacher ratings of student skills, concurrent and predictive relationships with other standardized tests, predictive relationships to summative tests such as state accountability tests, and construct equivalence to established measures of reading comprehension. The technical manual reports reliability coefficients by grade, including measures of internal consistency reliability as well as test-retest reliability over brief intervals, and from fall to spring. (See pages 42–57 of the *STAR Math Enterprise Technical Manual* provided in appendix 5.)

Concurrent and Predictive Validity

Pages 57–75 of the *STAR Math Enterprise Technical Manual* in appendix 5 contain concurrent and predictive validity evidence for STAR Math. Concurrent validity data includes correlations between STAR Math norming study test scores and other tests administered within a two-month time period. The data concerning predictive validity provides an estimate of the extent to which scores on the STAR Math test predicted scores on criterion measures given at a later point in time, operationally defined as more than two months between the STAR test (predictor) and the criterion test. It provides an estimate of the linear relationship between STAR scores and scores on measures covering a similar academic domain. Predictive correlations are attenuated by time due to the fact that students are gaining skills in the interim between testing occasions, and also by the differences between the tests' content specifications.

Meta-Analysis of the STAR Math Validity Data

Meta-analysis is a statistical procedure for combining results from different sources or studies. When applied to a set of correlation coefficients that estimate test validity, meta-analysis combines the observed correlations and sample sizes to yield estimates of overall validity, as well as standard errors and confidence intervals, both overall and within grades.

To conduct a meta-analysis of the STAR Math validity data, the 568 correlations reported in the current technical manual were combined and analyzed using a fixed-effects model for meta-analysis. The results are displayed in Table 31 (page 76) of the *STAR Math Enterprise Technical Manual* provided in appendix 5. This table lists results for the correlations within each grade, as well as results from combining data from all twelve grades. For each set of results, the table lists an estimate of the true validity, a standard error, and the lower and upper limits of a 95 percent confidence interval for the validity coefficient. Using the 568 correlation coefficients, the overall estimate of the validity of STAR Math is 0.69, with a standard error of 0.0001. The true validity is estimated to lie within the range of 0.69 to 0.70, with a 95 percent confidence level. Because the 568 correlations were obtained with widely different tests, and among students from twelve different grades, these results provide support for the validity of STAR Math as a

Requirement: 2.1

measure of math skills.

Using Assessment Results to Support Inferences about Teacher Effectiveness

Renaissance Learning continues to gather data on the validity of STAR to support inferences about educator and system effectiveness. Below are the types of analyses we continue to work on in this area:

- Evidence of year-to-year stability in STAR's measure of teacher impact on student learning. Specifically, do teachers with high performance one year with high performance the next? Renaissance Learning continues to gather this evidence.
- Evidence on how reliable STAR measures are in distinguishing among teachers, based on the number of test administrations and the number of students. Renaissance Learning continues to examine standard error around STAR's measure of teacher impact on student learning.
- Evidence that STAR correlates with other accepted measures of teacher effectiveness, including observational ratings of practice and value-added/growth on state tests.
    - Renaissance Learning has correlations between STAR and several state tests. We suggest that if the state assessments are valid measures for making inferences about teacher effectiveness, STAR can also be useful.
    - Renaissance Learning is beginning to correlate the student growth percentile within STAR to other state's growth measures.
    - Renaissance Learning can demonstrate that attributes of teacher practice impact STAR results. Within Renaissance Learning's vast databases, we have tens of millions of records on reading comprehension (average percent correct); percent correct of math problems, math fluency, and number of English words learned (for ELL students). These variables reflect the research-based best practices taught in professional development to improve teacher practice. Renaissance Learning has observed consistently over the last 20 years that high performance on these variables is strongly correlated to student performance on the STAR assessment.

In appendix 2, Renaissance Learning provides a policy brief titled *Using Short-Cycle Interim Assessment to Improve Educator Evaluation, Educator Effectiveness, and Student Achievement*. This policy brief explains how the use of multiple measures of student growth can potentially make educator evaluations more objective and actionable. Renaissance Learning commissioned this brief to explore how multiple measures can be used to help educators improve learning in their classrooms.

Requirement: 2.2

2.2
Provide evidence of reliability, both for the total test and for any subtests for which scores are reported. Include estimates of error in measurement.

Requirement: 2.2

In addition, the STAR assessments have received the highest ratings for reliability and validity from the National Center on Response to Intervention (NCRTI). STAR assessments met NCRTI's rigorous criteria for screening and progress monitoring. You'll find the ratings on the NCRTI website at http://www.rti4success.org/toolschartsLanding.

STAR Math Enterprise provides two ways to evaluate the reliability of its scores: reliability coefficients, which indicate the overall precision of a set of test scores, and conditional standard errors of measurement (CSEM), which provide an index of the degree of error in an individual test score. A reliability coefficient is a summary statistic that reflects the average amount of measurement precision in a specific examinee group or in a population as a whole. In STAR Math, the CSEM is an estimate of the unreliability of each individual test score. While a reliability coefficient is a single value that applies to the overall test, the magnitude of the CSEM may vary substantially from one student's test score to another's.

For more information on the reliability of the STAR Math test, please see the *STAR Math Enterprise Technical Manual* in appendix 5. The chapter titled Reliability and Measurement Precision (see pages 41–56) presents three different types of reliability coefficients: generic reliability, split-half reliability, and alternate forms reliability. This is followed by statistics on the conditional standard error of measurement of STAR Reading test scores.

Estimates of Measurement Error

The table on the following page contains two different sets of estimates of STAR Math Enterprise measurement error: conditional standard error of measurement (CSEM) and global standard error of measurement (SEM). The estimates of CSEM in the table are the average CSEM values observed for each grade.

Global standard error of measurement is based on the traditional SEM estimation method, using internal consistency reliability and the variance of the test scores to estimate the SEM:

SEM = SQRT(1 – ρ) σx, where

- SQRT() is the square root operator
- ρ is the estimated internal consistency reliability
- σx is the standard deviation of the observed scores (in this case, scaled scores)

Global estimates of SEM can be expected to be more conservative (larger) than CSEM estimates, because the former are calculated from observed data, while the individual CSEM values are theory-based. To the extent that students' item responses do not perfectly fit the IRT model used (here, the Rasch model), CSEM should underestimate measurement error. Consistent with that, the global values of SEM shown in the table are equal to or greater than the counterpart CSEM values at every grade. The similarity of the values provides confidence that these estimates of STAR Math Enterprise measurement error are reasonably accurate.

Requirement: 2.2

Estimates of STAR Math Enterprise Measurement Precision by Grade: Conditional and Global Standard Error of Measurement

Grade	Sample Size	Conditional Standard Error of Measurement Average CSEM	Standard Deviation	Global Standard Error of Measurement (SEM)
K	53	31	7.1	37
1	1,425	30	2.4	32
2	1,560	30	2.6	31
3	1,791	30	3.1	31
4	2,223	30	2.2	31
5	2,432	30	3.1	31
6	1,533	30	2.6	31
7	1,213	30	2.1	32
8	876	30	3.1	32
9	439	30	2.4	30
10	148	31	3.2	31
11	112	30	1.6	30
12	211	30	3.2	33
Average		30		32

Requirement: 2.3

2.3
Provide evidence that the assessment is appropriate for use with student subgroups, including English language learners and student with disabilities. Include documentation that the assessment does not exhibit bias toward any major subgroups (e.g., through an analysis of differential item functioning). In addition, provide a sensitivity review to demonstrate the assessment tasks and items are designed to be accessible and fair for all students.

Item development is of critical concern to Renaissance Learning. All content for Renaissance Learning products is developed by professional designers, writers, and editors with education backgrounds and content-area expertise. Writers and editors receive ongoing item writing and bias and fairness trainings to ensure that test items are appropriate for use with all student subgroups. All content is subject to a

Requirement: 2.3

stringent development process that includes adherence to content specifications and internal and external expert review and auditing. Items for the STAR Reading Enterprise assessment go through a rigorous calibration process.

Bias and Fairness Philosophy

All STAR Math Enterprise items are reviewed for bias and fairness by content experts before being included in the item bank. Our content specifications for each content area are research based, and they include bias and fairness criteria that address stereotypes, misleading characterizations of people or events, or offensive language that could be construed as demeaning, patronizing, or otherwise insensitive. Content development tools track and report on attributes such as gender, age, ethnicity, subject and topic areas, occupation, disability, and geographic region. The individual attributes as well as the intersection of multiple attributes are tracked throughout the development process to ensure that final content is demographically balanced, free of bias, and accessible and fair for all students. Items that are deemed to be biased or unfair are either revised and resubmitted for review or discarded. For more information on Renaissance Learning's bias and fairness philosophy, please see pages 24– 25 of the *STAR Math Enterprise Technical Manual* provided in appendix 5.

Differential Item Functioning Analysis

Please refer to appendix 5 to view the STAR Math item bias study, titled *Evaluating the Influence of Differential Item Functioning for Race and Gender on STAR Math Items* by Damien C. Cormier.[3]

This study concludes that the STAR Math items analyzed are not biased as regards gender (Male and Female) or as regards the racial groups studied (Whites, Hispanics, and Blacks). Additional item bias studies were completed more recently for the Asian and Native American racial groups. Sample size limitations restricted the analyses to fewer items than had been studied for the White, Hispanic, Black, and gender groups. However, the additional item bias study also concluded that STAR Math is not biased as regards Asian and Native American racial groups.

The Cormier report cited above evaluated more than 500 STAR Math Enterprise items. New STAR Math test items are periodically field tested and calibrated in batches of hundreds of test items. Going forward, DIF analyses will be performed on all STAR Math Enterprise items.

Protocol for Review, Modification, and Replacement of Items

Since our item bias study determined that none of the STAR Math items evaluated are biased, we did not modify or replace any of the studied items for bias and fairness reasons. However, Renaissance Learning follows a strict protocol for developing and retaining items in the STAR item banks, as follows:[4]

- Analyze standards to be assessed in the categories of skill, action, vocabulary, and context; and refer to national or state resources for appropriate standard and grade-level expectation

[3] Cormier, D.C. (2011). Evaluating the Influence of Differential Item Functioning for Race and Gender on STAR Reading Items. Technical report. Renaissance Learning, August 2011.

[4] For more information on how Renaissance Learning works to ensure quality item development, refer to the document titled *The Foundation of the STAR Assessments* (pages 7–8) provided in appendix 2

Requirement: 2.3

interpretation.

- Write item specifications and provide specifications training to item writers and editors.
- Establish item metadata to guide development, including standards-related and item-related data.
- Use a multistep recursive writing and editing process that ensures adherence to specifications and alignment to standards and item metadata.
- Post items for calibration and acquire student-response data through STAR's dynamic calibration process.
- Examine psychometricians' analyses of item testing results.
- Add successful items to the operational assessment item bank.

Assessment items must also pass strict quality reviews which check for discipline-specific criteria, accuracy, language appropriateness and readability level, bias and fairness, and technical quality control.

Following these analyses, all information pertaining to each test item—including traditional and item response theory (IRT) analysis data, test level, form, and item identifier—is stored in an item-statistics database. Then a panel of content reviewers examines each item within content strands to determine whether the item meets all criteria for use in an operational assessment. After all content reviewers have designated certain items for elimination, the recommendations are combined and a second review is conducted to resolve any issues.

Requirement: 2.4

2.4
Provide evidence that the assessment includes items of varying difficulty to ensure accurate measurement of student achievement across the ability continuum, including the tails of the score distribution.

STAR Math Enterprise is a computer-adaptive assessment built on item response theory (IRT). Item response theory allows the assessment to take advantage of the calibrated difficulty of every test item as it selects items dynamically during each test and matches item difficulty to each student's performance one item at a time. The assessment then calculates each student's score on a single vertical score scale that spans Grades K through 12. The STAR assessments use the Rasch—or one-parameter logistic—item response model.

In general, the assessment continually tailors the difficulty of the questions to the student's responses. For example, if the student's response is correct, the difficulty level is increased. If the student misses a question, the difficulty level is reduced—in real time. As the test progresses, a student may be "routed" to items at the lowest reading level or to items at higher reading levels within the overall bank of items,

depending on the student's performance during the testing session. Thus, a low-performing student's reading skills may branch to easier items in order to better estimate his or her reading achievement level. High-performing students may branch to more challenging reading items in order to better determine the breadth of their reading skills and their reading achievement level. This saves testing time and spares students the frustration of responding to items that are too difficult and the boredom of responding to items that are too easy. Another key benefit of a well-calibrated, scaled computer-adaptive test is that it eliminates "testing to the middle." There are no grade-level test forms; each STAR test "meets them where they are," and each individual test concentrates its content at the manifest performance level of the student. This method yields extremely accurate information about what that student has and has not mastered.

Each STAR Math Enterprise assessment administers 34 items to the student. At each point during the test, the adaptive testing software seeks an item with about a two-thirds probability of a correct answer by a student at the student's current estimated ability level. This can be achieved by finding an item with Rasch difficulty approximately equal to the Rasch ability estimate minus 0.73 Rasch units. To do so requires a large set of test items with a wide range of difficulty. For this reason, the STAR Math Enterprise assessment has a large item bank, containing more than 5,000 unique items. Each item has been designed to measure a skill at a specific grade level. All items have been calibrated so that the difficulty of each item is expressed on a Rasch difficulty scale that spans the range of math proficiency from kindergarten through grade 12.

The table on the following page summarizes the Rasch difficulty statistics by item grade level (0 to 12) for 4,442 STAR Math Enterprise test items. The statistics show the number of items at each grade level (N), the mean Rasch difficulty parameter of the items as well as the standard deviation, the median value, and the range of item difficulty from the 1st to the 99th percentile. For each grade level, the 1st percentile (column 5) and the 99th percentile (last column) represent the Rasch item difficulty at the lowest and at the highest end of the distribution, respectively.

The Range of Rasch Difficulty by Item Grade Level for 4,442 STAR Math Enterprise Items						
Item Grade Level	N	Mean	Std Dev	1st Pctl	Median	99th Pctl
0	34	-5.182	0.944	-6.810	-5.435	-2.911
1	573	-5.124	9.716	-6.000	-4.011	-2.088
2	357	-2.888	1.012	-5.402	-2.870	-0.647
3	688	-2.655	9.084	-4.920	-1.810	0.444
4	538	-0.600	0.873	-2.896	-0.559	1.680
5	611	0.738	1.005	-1.876	0.800	3.360
6	619	1.324	0.998	-1.401	1.280	4.000
7	396	-0.041	12.342	-99.000	1.475	4.810
8	330	1.898	0.722	0.297	1.920	3.550
9	208	2.087	0.757	0.255	2.218	3.620
10	43	2.821	0.953	1.302	2.580	5.040
11	20	3.098	0.669	1.490	3.160	4.320
12	5	3.860	0.408	3.370	3.700	4.290

Section 3: Use of Assessment as a Measure of Growth

Requirement: 3.1
3.1 Provide evidence that the scores resulting from the assessment have been used as measures of growth by other local or state education agencies.
Currently, approximately 30,000 schools use the STAR assessments on the Renaissance Place Real Time platform, including many schools in Virginia. During the 2011–2012 school year alone, more than 38 million STAR assessments were taken by students on this platform. The STAR Enterprise assessments have been approved for use as growth measures specifically for educator evaluation in New York, Ohio, and Tennessee, giving us valuable experience helping educators work through some possibilities for using STAR for this purpose. In addition, STAR Enterprise has been approved for use in the state of Wisconsin as an effective Response to Intervention (RTI) measure, and STAR Math Enterprise has been approved in Kansas for use within the state's multi-tier system of support

Requirement: 3.1
(MTSS). With STAR as part of these frameworks, educators establish goals for growth and then monitor progress toward these goals. And with the valuable reporting available with STAR, educators have enough time to provide the deliberate, personalized practice and feedback students need for growth.

Requirement: 3.2
3.2 Describe the methodology used to measure growth. For example, does the assessment employ a vertical scale, use a computer-adaptive model to measure growth over time, or employ some other methodology. Does the methodology allow for the longitudinal measure of growth across academic years? What about the measurement of required growth on the proposed assessment to reach proficient on the statewide assessments (the Standards of Learning tests) in a specified amount of time? Include standard setting studies or other analyses conducted to establish measures of growth.

When it comes to measuring student growth over time and conveying the results to educators in ways that are actionable and instructionally relevant, the STAR assessments lead the way. STAR Early Literacy Enterprise offers three primary ways of measuring student growth: student growth percentiles, weekly growth rates, and pre- to posttest change in scores.

- The **student growth percentile** provides a normative growth comparison on a scale of 1–99. It is particularly useful in educator evaluation formulas.
- **Weekly growth rates** offer an easy-to-calculate rate of improvement toward student academic goals and are especially helpful in Response to Intervention (RTI) settings.
- **Pre- to posttest growth**, usually determined by changes in a student's STAR scaled score, shows the raw gains a student has made over a specific period of time.

As described below, each of these methods delivers growth information to teachers in a digestible and meaningful way they can act on as soon as students complete a STAR test.[5]

Student Growth Percentile

Student growth percentiles (SGPs) are a norm-referenced growth score that compares a student's growth to that of his or her academic peers nationwide. Using quantile regression techniques, this score quantifies individual student growth to show how a student changed from one STAR testing window to the next, relative to other students with similar starting STAR scores. This score is displayed on the STAR Growth Report, which can be run for a group, class, grade, school, or district. A sample Growth Report for STAR Math Enterprise is provided in *Key* the *Report Samples: STAR Enterprise* in appendix 2.

SGPs range from 1–99 and interpretation is similar to that of percentile rank scores; lower numbers

[5] Additional information about STAR Math Enterprise growth norms can be found in the *STAR Math Enterprise Technical Manual* in appendix 5.

Requirement: 3.2

indicate lower relative growth and higher numbers show higher relative growth. For example, an SGP of 70 means that the student's growth from one test to another exceeds the growth of 70 percent of students in the same grade with a similar beginning (pretest) STAR score. All students, no matter their starting STAR score, have an equal chance to demonstrate growth at any of the 99 percentiles.

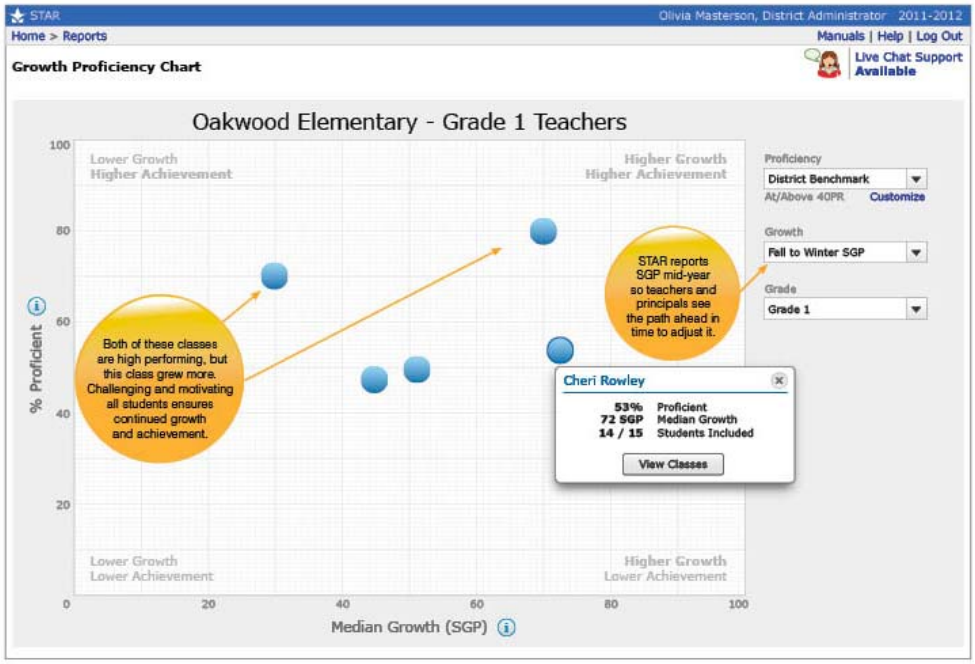
SGP is often used to indicate whether a student's growth is more or less than can be expected. For example, without an SGP, a teacher would not know if a scaled score increase of 100 represents good, not-so-good, or average growth. This is because students of differing achievement levels in different grades grow at different rates relative to the STAR scale. For example, a high-achieving first-grader grows at a different rate than a low-achieving first-grader. Similarly, a high-achieving second-grader grows at a different rate than a high-achieving third-grader.

SGP can be aggregated to describe typical growth for groups of students—for example, a class, grade, or school as a whole—by calculating the group's median, or middle, growth percentile. No matter how SGPs are aggregated, whether at the class, grade, or school level, the statistic and its interpretation remain the same. For example, if the students in one class have a median SGP of 62, that particular group of students, on average, achieved higher growth than their academic peers.

Typical growth is characterized by individual states. For example, in Virginia, typical growth for a student is defined as an SGP that falls between 35 and 65. A student with low growth would have an SGP between 1 and 34, and a student with high growth would have an SGP between 66 and 100. An SGP of 50 represents one year's growth. When looking at a school or group, a median SGP above 50 indicates a tendency toward higher growth, and a median SGP below 50 indicates a tendency toward lower growth. As additional states define their categories for typical growth, we can begin to tailor our advice to each system and do the work of correlating qualitative data with STAR results.

In addition to reporting SGP in STAR, Renaissance Learning has partnered with the SchoolView® Foundation to provide a visual experience of this score; within the STAR software, SGP is represented visually by the Growth Proficiency Chart, shown on the following page. This visual emulates Colorado's SchoolView visualization tool for student growth. Renaissance Learning continues to work in partnership with SchoolView, a non-profit organization that enables states and their partner stakeholders to leverage education data to enhance accountability measures and student academic outcomes.

Requirement: 3.2



SGPs were initially developed by Dr. Damian Betebenner of the Center for Assessment for use with state summative tests. The STAR assessments are believed to be the first interim assessments that report this score. To adapt the SGP methodology for use with the STAR assessments, Renaissance Learning worked with Dr. Betebenner and also with Dr. Daniel Bolt, a professor of quantitative methods at the University of Wisconsin–Madison. Through an extensive research and development process, these renowned measurement experts helped establish a valid methodology for creating the SGP scores and provided guidance on how SGP should be presented to educators on STAR Math reports and in professional development.

Weekly Growth Rates

STAR Math Enterprise growth norms are based on a student's grade and level of achievement. While similar tests only provide norm-referenced information with respect to a student's performance at a particular point in time, STAR's growth norms go a step beyond this approach by providing a reference to student growth over time within the academic year. This approach allows a student's observed growth over a period of time to be compared to growth made by students of a similar grade and achievement level.

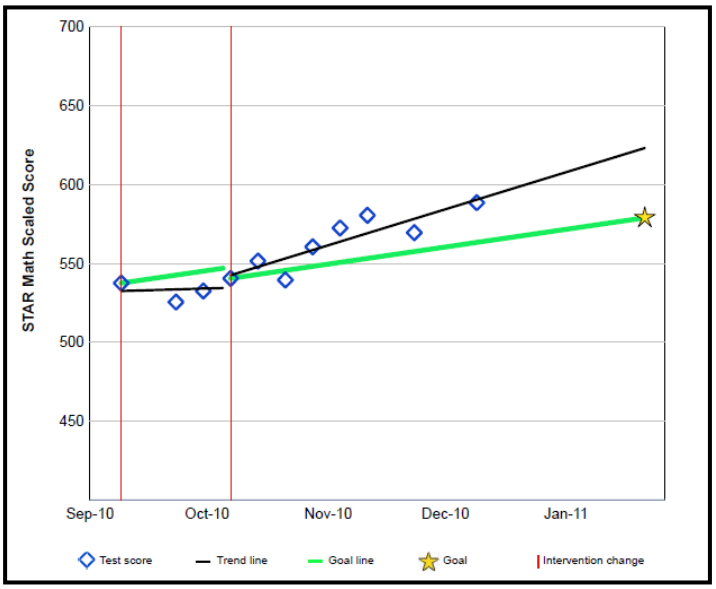
Growth norms are especially helpful for setting short-term growth targets for incremental progress monitoring in Response to Intervention. Of particular note is the Student Progress Monitoring Report, which includes a built-in Goal-Setting Wizard.[6] This tool helps teachers select a personalized growth target for each student, providing two recommendations (moderate and ambitious growth) and also allowing users to select a customized target. The moderate growth target is the 50th percentile of the

[6] To see a screen shot of the Goal-Setting Wizard and a sample Student Progress Monitoring Report, see the *Key Report Samples: STAR Enterprise* document in appendix 2.

Requirement: 3.2

student's decile. The ambitious target is the 75th percentile of the student's decile.

Once goals are set, the Goal-Setting Wizard automatically graphs a student's progress toward goals on the Student Progress Monitoring Report, as shown at right. This information can help teachers decide how to adjust instruction for each student. Teachers who decide to start a new intervention can simply designate the start date of another intervention using the Goal-Setting Wizard. When the Student Progress Monitoring Report is run again, the new intervention is indicated by a new red intervention line and the start of a new black trend line. This gives teachers a quick lens through which to observe the effects of a particular intervention over time, as well as monitor progress toward a specific achievement goal.



Using Scores to Indicate Pretest to Posttest Growth

The third type of growth reported by STAR is pretest to posttest growth, which is sometimes referred to as a student's "raw gain score." A student's raw gains are best illustrated by pre- to posttest changes in the student's scaled score. This score represents a student's reading ability on a continuous vertical scale that ranges from 0–1400, spans all grades 1–12, and can be used to indicate longitudinal growth. A student's scaled score maps a student to a specific location on the underlying ability continuum. This score is especially useful for helping educators gain an understanding of the general achievement level of a student or to evaluate the effectiveness of an instructional intervention with the group as a whole. The scaled score is shown on a variety of STAR Math Enterprise reports, including the Growth Report. (See the document titled *Key Report Samples: STAR Enterprise* in appendix 2 to view a sample of the Growth Report.) It is also the basis for all of the other scores reported by STAR Math Enterprise (e.g., grade equivalent, percentile rank, normal curve equivalent, etc.).

Predicting Performance on Statewide Assessments

Renaissance Learning has linked the STAR tests to state summative tests. Customized metrics depict a variety of meaningful aggregated data that demonstrates benchmark and proficiency projections for each of these states. In addition, Performance Reports predict student performance on state tests—at the student, class, school, and district levels—early in the year, while there is still time to change the course of each student's achievement path. For Virginia, this family of reports will show student proficiency and a growth trajectory in relation to the Virginia SOL's four performance levels (Pass/Advanced, Pass/Proficient, Fail/Basic, Fail/Below Basic), giving educators an early warning sign as to which students may be at risk for not meeting state requirements. Predictions will account for the amount of growth that typically occurs between the date of the last STAR test taken and the date of the SOL.

Requirement: 3.2
Please note that linking STAR and state test scores requires that we match STAR test scores to Virginia's Standards of Learning (SOL) test scores. Renaissance Learning has partnered with 43 states to link the STAR scale to the scale of their state summative test; in order for educators in Virginia to take advantage of these critical reports, we invite Virginia to partner with us to complete a linking study for the STAR tests and the SOL exams. The linking process generally takes about six weeks to complete. Renaissance Learning will work with your state to determine an appropriate schedule for the project as soon as sufficient data are available. SOL performance predictions and reports will be available in the software as soon as the linking study is complete. In addition to performance predictions for state assessments, Renaissance Learning is actively conducting a research study to understand the predictive relationship between the STAR assessments and the test used in Readistep, PSAT, and SAT tests. STAR growth models can suggest a personalized growth trajectory given the student's performance to date and the end-of-grade STAR score equated to the college readiness benchmark. This research will make it even more possible for Virginia educators to determine which students are college- and career-ready—and enable them to fine-tune instruction while there is still time to improve performance before the regular test cycle.

Requirement: 3.3
3.3 Describe the methodologies used to control item exposure so that the accuracy of students' scores is not impacted by multiple exposures to the same items.
The STAR Math Enterprise item bank includes more than 5,000 carefully calibrated items. STAR software keeps track of the specific items presented to each student from test session to test session, and does not present the same item more than once in any 90-day period. By doing so, the software keeps item reuse to a minimum. Because items are dynamically selected from the STAR item banks each time the student responds to an item, STAR only selects items that the student has not had in a recent previous test. In addition to restricting the use of items previously administered to each student, STAR also minimizes excessive exposure of individual test items in two ways: 1) each item administered to a student at any point in the test is randomly chosen from the full set of all items that meet the criteria for content and difficulty at that point; and 2) with few exceptions, no more than one item measuring the same skill is administered in the course of a student's test. These two precautions serve to balance the exposure of all items. In this way, each test is individually assembled based on the student's past and present performance, ensuring that there is little risk of item overexposure. Furthermore, if a student is progressing in reading development throughout the year and from year to year, item overexposure should not be an issue at all. More information on the content of the STAR Math item bank is available in the *STAR Math Enterprise Technical Manual* provided in appendix 5 (see pages 14–25).

Requirement: 3.4

3.4
Describe the procedures used to validate the measures of growth.

Below are descriptions of the procedures used to validate the three primary measures of growth reported by STAR: student growth percentiles, weekly growth rates, and pre- to posttest changes in scaled scores.

Student Growth Percentiles

Because STAR Math Enterprise is so widely used, Renaissance Learning has data for millions of testing events. With these scores, we are able to calculate growth norms that approximate typical growth for students of different achievement levels in different grades from one time period to another.

SGP was initially developed by Dr. Damian Betebenner for use in state summative assessments and is currently used by 12 state accountability assessments, including Virginia's Standards of Learning tests. Renaissance Learning worked closely with Dr. Betebenner as well as Dr. Daniel Bolt from the University of Wisconsin–Madison to adapt the state SGP model for year-to-year growth to interim testing within the school using a dataset that approximately two million students. As a result of this work, the STAR assessments are believed to be the first interim assessments to incorporate SGP.

The SGP used in state accountability tests measures growth from one school year to the next. In the STAR interim assessments, SGP can track growth for multiple testing windows—fall to winter, winter to spring, and fall to spring.[7] Using the fall-to-winter SGP, teachers have the opportunity to reflect on practice and instruction and take appropriate action in the classroom to facilitate student growth. Similarly, teachers can see the growth that takes place between winter and spring STAR testing, and/or fall and spring STAR testing.

To calculate student growth percentiles, Renaissance Learning collected hosted student data from the five most recent school years (2006–07, 2007–08, 2008–09, 2009–10, and 2010–11). Sample sizes were approximately two million students.[8] Quantile regression was applied to characterize the bivariate distribution of students' initial scores and ending scores. Students were grouped by grade and subject, and then quantile regression was used to associate with every possible initial score and ending score combination a percentile corresponding to the conditional distribution of end score given the initial score. The result of these analyses was the creation of a look-up table in which students' beginning and ending STAR scores are used as input to define a student growth percentile for each grade, subject, and time period (e.g., fall to winter, winter to spring, fall to spring). The use of quantile regression techniques makes construction of such tables possible even though not all possible initial and ending score combinations were observed in the student data. Loosely speaking, the quantile regression approach can be viewed as a type of smoothing in which information from neighboring score values (initial and ending) can be used to inform percentiles for hypothetical score combinations not yet observed. As such, application of the methodology allows us to look up any score combination to obtain the percentile cut points for ending

[7] Renaissance Learning collects data for our growth norms during three different time periods: fall, winter, and spring. For SGP to be reported, students must be tested within at least two of the following date ranges: August 1–September 30; December 1–January 31; and April 1–May 31.

[8] Table 55 in the STAR Math Enterprise Technical Manual (page 120) in appendix 5 shows details of the demographics of these students.

Requirement: 3.4

score conditional achievement distribution associated with the given initial score. These cut points are the percentiles of the conditional distribution associated with the student's prior achievement. Specifically, using the quantile regression results of STAR Math Enterprise spring scores on fall scores, estimation of the 1st through 99th percentile growth from fall to spring can be calculated. Using each of these cut points, we are able to calculate a student growth percentile for every subject, grade, and score combination.

Weekly Growth Rates

Growth norms for the weekly growth rates in the STAR Math assessment were developed for each grade by following students across the entire academic year, ranging from August to June, with some minor variation in schedules across schools. Students were tested both at the beginning and end of the school year (fall and spring semesters), allowing the student growth estimates to be computed across the academic year. To normalize differences in time between the initial test and the final test at the end of the school year, the measure of growth (change in score from fall to spring testing) was divided by the number of weeks between the assessment occasions to obtain an estimate of typical growth per week for all students.

Since students develop academic skills at different rates as they mature and move across the grades, they also develop and grow at different rates within each grade depending on where they score in the overall distribution of performance. Students who score in the top decile for a grade do not, and should not be expected to, grow at the same rate across the academic year as students in the middle or lower deciles, and vice versa. Therefore, it would be problematic to use growth rate expectations obtained from a group of students who were not of comparable achievement levels. Growth rates of students should be compared to students of similar academic achievement levels; otherwise, there is the potential to expect too much or too little growth from certain students.

To account for differences in student growth, both across grades and within grades during an academic year, growth norms were developed by using information about grade and level of performance to construct homogeneous student groupings for comparison. The within-grade groupings were done by partitioning students into decile groups based on their initial percentile rank scores within a school year; STAR Math percentile ranks are based on the nationally representative samples used to generate the norms, as described earlier. For example, students in different grades will have different expectations of growth, and students within the same grade can have different levels of growth depending on their performance and achievement level. Students within the same grade in the lowest decile group will have a different growth expectation from students in the other nine decile groups within that grade.

STAR growth norms were constructed by following students within each decile of each grade across the entire academic year. This provided a means to compute a distribution of growth scores for every decile group for all grades, i.e., 10 decile groups for each grade will each have their own growth norms distribution. The growth norms are thus conditional on both grade and decile level of student initial performance during the academic year.

All data was retrieved from the hosted customer database (Renaissance Place Real Time), and the growth norms are updated periodically. This allows for the norms to continually reflect the effects of changes in educational practices that can have differential effects on student learning, and also keeps the growth norms up-to-date to ensure students' observed growth is being referenced against a current

Requirement: 3.4

student group.

Growth norms provide a norm-referenced method of computing expected growth distributions for students. This allows student growth over time to be referenced to a known normative distribution of growth. Therefore, a relative measure of student growth can be obtained by placing a student's growth over time in the context of the distribution of growth observed among students of a similar grade and initial performance level. To provide an estimate of the normal growth for students of any grade and decile level, the median growth rate from the growth rate distribution is used. However, in principle, any observed growth for a student can be referenced against the appropriate growth distribution to evaluate where that student's growth ranks in comparison to other similar students.

Pretest to Posttest Changes in Scaled Scores

Another way educators can track growth is to look at changes in test scores, specifically the scaled score. The STAR Math Enterprise score scale spans from 0–1400. This vertical scale was derived as part of the test development process. In adaptive testing, students can receive different sets of test items and still receive a comparable scaled score that represents their unique underlying ability level. Because scaled scores essentially map a student to a specific location on the underlying ability continuum, they can be useful in measuring absolute growth.

Section 4: Test Administration Procedures

Requirement: 4.1

4.1
Describe the administration procedures necessary to produce growth scores. For example, is the assessment designed to be administered multiple times during the year or administered once in the fall and once in the spring?

Below we describe the administration procedures necessary to produce student growth percentiles, weekly growth rates, and scaled scores.

Student Growth Percentiles

Renaissance Learning collects data for our growth norms during three different time periods: fall, winter, and spring. To receive a student growth percentile (SGP), students must be tested during these periods. We provide student growth percentiles for achievement that takes place between fall and winter, winter and spring, and fall and spring. For SGPs to be reported, students must be tested within at least two of the following date ranges:

- **Fall**: August 1–September 30
- **Winter**: December 1–January 31

Requirement: 4.1

- **Spring**: April 1–May 31

We recommend that educators do universal screening with the STAR assessments within the SGP testing windows. For example, fall screening could be scheduled for September 1–September 15. Winter screening could be scheduled for January 1–January 15, and spring screening could be scheduled for May 1–May 15. That way, data from a single test would be used for both screening and growth reporting, and students would take fewer tests. These default screening dates can be edited and more screening periods can be added. The first screening period should be as close to the beginning of the school year as possible so that educators can address instructional needs quickly.

Because STAR measures each student's achievement relative to that of other students, we recommend administering the test to all students within a fairly short time period. While the software allows users to define a 30-day screening period, two weeks or less is optimal. Testing all students within a narrow timeframe creates more useful data. For example, let's suppose you are evaluating the growth of two students between fall and winter. You tested Student A on September 1 and January 30, but you tested Student B on September 30 and January 1. Both students had the same fall test score and both gained 100 scaled scores by the time of the winter test. Because they both were tested within the fall and winter SGP testing windows, the software will consider them to have the same student growth percentile. But in reality, their rate of growth will not have been the same. Student A will have gained 100 scaled scores in five months (September 1–January 30), while Student B will have gained 100 scaled scores in three months (September 30–January 1). To avoid this situation, we recommend that all students are tested within the narrowest timeframe possible so that they have approximately the same amount of time to grow and so educators can more accurately compare one student's growth to another's.

Weekly Growth Rates

Using the STAR Goal-Setting Wizard, teachers may monitor a student's growth as often as weekly. First, the teacher selects a start date for the intervention goal and administers a STAR assessment close to the start date to establish baseline data. Then the software calculates the rate of growth that the student must achieve per week to reach his or her goal by the end of the school year. This growth rate will vary depending upon how ambitious the student's goal is. For example, a "moderate" goal displays the growth rate achieved by 50 percent of students with a similar percentile rank as the student for whom you are setting goals. An "ambitious" goal displays the growth rate achieved by 25 percent of students with a similar percentile rank. Students then take the STAR assessments as often as weekly so that teachers can monitor students' progress. The Student Progress Monitoring Report displays a student's scaled score and growth rate from one test to the next. The growth rate appears after the student has completed four tests, the number of tests statistically required to report a growth trend with confidence.

Scaled Scores

After any assessment event, educators can look to a student's scaled score to see whether a student has improved. Comparing a student's scaled score from one assessment event to another will help educators see a student's pre- to posttest changes and determine the growth that student has made.

Requirement: 4.2

4.2
Describe any processes used for pre-identifying and/or registering students for testing. Include what data, including the State Testing Identifier, are collected for each student, how data are collected or transmitted, and how data are maintained and securely managed.

Registering Students for Testing

School division administrators, school division staff, and school administrators can easily register students for the STAR assessments on the Renaissance Place Real Time platform. Student information may be entered manually or imported from your student information system (SIS). In either case, student information need only be imported once and is updated in one central location, making it easy to add and maintain the information in real time. The registration system also contains a field that permits Virginia users to include the State Testing Identifier.

Data Collection and Integration

The STAR assessments are run on the Renaissance Place Real Time platform, which is a client/server platform located on a central web-based server. Through Renaissance Place, customers can store, access, and manage the data associated with the Renaissance Learning products they purchased. In addition, customers may add, edit, or delete information about their district, schools, school years (including marking periods and days off), personnel (including teachers), students (including customizable demographic categories), courses, and classes across all of their Renaissance Learning products simultaneously.

Renaissance Learning is SIF compliant. We also offer an interface service called Renaissance Data Integrator (RDI), a custom-built program that automatically refreshes your Renaissance Place applications daily with new data from your student information system (SIS). RDI is capable of processing both user and schedule items from your SIS. These items are presented in the table below.

User and schedule items processed in Renaissance Data Integrator	
User Items	**Schedule Items**
• School Information • Student Information • Staff Information • User names/Passwords • Ethnicities • Student Characteristics	• Student School Enrollment • Student Class Enrollment • Terms • Courses • Classes • Default Homerooms

The SIS data is transferred on a secure FTP connection over an SSL/HTTPS port 443 connection. The SIS extracts the user and schedule items you entered, converts them to readable text files, and transfers them to the Renaissance Place server. RDI then imports these files, identifies any transfers,

additions, changes, or deletions in your SIS data, and processes these changes in your Renaissance Place software applications. Therefore, the student information in your Renaissance programs is always as current as your last RDI run, eliminating the need for manual entry and data updates. Moreover, because our support team tracks and manages your SIS-RDI interface on a daily basis, your administrative staff and teachers are free to focus on other pressing issues. Please note that if data is to be extracted and used in larger organization systems, all participating districts will need to include State Testing Identifiers/Unique Student Identifiers as part of their daily SIS extract.

Customer Data Security

Renaissance Learning practices industry-standard procedures to maintain optimal security once customers are ready to create user accounts in Renaissance Place. These practices include defining user permissions based upon individuals' roles within the school district, authenticating users to guarantee that only authorized individuals gain access to the system, and encouraging users to follow recommended password security protocols.

User Permissions and Authentication. Renaissance Learning's product software contains seven defined user groups, including district administrator, district staff, school administrator, school staff, teachers, parents, and students. The system administrator assigns each person using the software to one of these user groups and to a specific position within that group. Each person is then assigned capabilities, which gives users the right to perform specific, pre-assigned tasks in the software based on their user group and the tasks that the user group usually performs. These settings also allow different levels of access to the software setup and data. For example, district administrators have access to reports for all levels—individual students, classes, teachers, schools, and the district—while teachers will only be able to view reports for their students and classes. Some staff, however, may perform more than one role in the school or district, or they may be assigned to more than one school. The system keeps track of these multiple roles and school assignments and allows these users to switch between roles or schools while they are using the software.

Authentication of users is self-contained within the Renaissance Place system. To authenticate users, Renaissance Place utilizes the forms authentication model included in Microsoft's ASP.net web application framework. Within this model, a user logs in to Renaissance Place and enters his or her password, which is validated against the encoded password associated with that user and stored in the system database.

Password Security. To access programs in Renaissance Place, all users—including administrators, teachers, parents, and students—must enter a user name and password. These user names and passwords can be any combination of upper case, lower case, numeric, or symbol characters. When a system administrator creates user accounts in the system, these users are automatically assigned a temporary password. The first time users log in with their temporary passwords, the system requires them to change their passwords. If users forget their passwords, they should contact Renaissance Learning via e-mail. We require customers to e-mail us with a password request; Renaissance Place user passwords are never provided over the telephone. Customers must send their requests from a valid district or school domain e-mail account. This also applies to any requests for backdoor passwords or utilities that may be necessary to reset the district administrator's user account. Additionally, if customers try unsuccessfully to log in three times, they will be locked out of the system. The locked-out login will automatically clear every evening at midnight.

Users should change passwords periodically as a safeguard against illegal access of their account. While Renaissance Place requires users to update their passwords annually, they may manually update passwords more frequently. System administrators can also force the system to prompt password changes.

Requirement: 4.3

4.3
Describe all materials needed for test administration and how school divisions will order and obtain sufficient quantities. Include details of test booklets and answer documents for paper/pencil testing (if applicable), test administration manuals, etc. If applicable, identify any test administration materials school divisions would be responsible for supplying locally (manipulatives, copies of test materials, etc).

To administer the STAR Enterprise assessments, the following materials are required: STAR Enterprise software, an Internet connection, and classroom or lab computers. The software comes with a variety of print and online user resources, including a software manual that contains general information and step-by-step instructions for working with the program, and a technical manual that explains the development of the assessment as well as information about the test's reliability and validity.[9] Though not required for test administration, these manuals are designed to assist your educators with the implementation of the STAR assessments. They are available for free to educators online with your Renaissance Place Real Time subscription.

School divisions may order student subscriptions for the STAR Enterprise assessments at any time by contacting Renaissance Learning at (800) 338-4204. A sales representative will work with each school division to determine the amount of student subscriptions necessary for the implementation, based on the number of students who will be using the assessments at each school.

Requirement: 4.4

4.4
Provide examples of the test administration manuals to be used with the assessment(s).

Please refer to the *STAR Math Software Manual* provided in appendix 5 for information on best practices for test administration. While not required to administer the STAR test, the software manual can assist educators with all functionalities of the software.

[9] Copies of the software manual and technical manual for STAR Reading Enterprise are provided in appendix 5.

Requirement: 4.5

4.5
Describe all technology requirements related to school personnel managing the administration of tests and to students completing tests if assessments include technology-based delivery. Include the minimum and recommended hardware and software requirements and network requirements for test administration by school personnel and test delivery to students. Include how assessments are hosted (e.g., locally, vendor, 3rd party). Provide examples of user interfaces for test administration by school personnel and test delivery to students. Include descriptions or examples of test navigation and any test tools (e.g., calculator, ruler, highlighter) available to students for testing.

Technology Requirements

The STAR assessments are web-based, computer-adaptive tests. They operate on the Renaissance Place Real Time platform, a comprehensive data management platform that can be accessed on both PC and Macintosh operating systems. In addition, Renaissance Learning is in the process of developing the STAR assessments to be supported on Apple's iOS. Detailed information on minimum and recommended hardware, software, and network requirements for Renaissance Place are provided in appendix 2.

Web-Hosting Services

Renaissance Learning provides web-hosting services to schools and districts through our Enterprise-Class Data Center. This allows us to deliver all new content, features, and other updates instantly and automatically, freeing up district technical staff to focus on other important issues.

Assessment Interface Examples

Appendix 5 contains a sample sequence for a student taking a STAR Reading Enterprise assessment. The screenshots depict how students navigate through the assessment from login to test completion. Appendix 5 also contains a school personnel sequence that depicts how teachers and administrators will use the Renaissance Place Real Time platform to run STAR Enterprise reports and manage data from the assessments. You will also see a screen shot of some customizable reporting options. Educators can control what information they access and view in the system, and they can organize this information based on local preferences.

Requirement: 4.6

4.6
Describe accommodations available to students with disabilities and limited English proficient students. Include procedures related to the provision of accommodations to eligible students.

Renaissance Learning takes the subjects of testing accommodations and alternate assessments seriously, and we are confident that we provide outstanding software programs that come a long way toward obviating the need for alternate assessments, due to their adaptive nature. Detailed information on

Requirement: 4.6

these topics is provided below.

Renaissance Learning is committed to making the STAR Enterprise assessments accessible to all types of students, including English language learners and students with disabilities. With that objective in mind, we continue to explore a variety of options for accommodating most students. Currently, we are engaged in several opportunities and endeavors to add new software features and hardware accessibility with the end goal of making the STAR assessments accessible to all. Further, we recently created a new assessment that measures Spanish reading comprehension for Spanish speakers, called **STAR Reading Spanish™**. The STAR Reading Spanish items were developed by native speakers (the test is not a translation of STAR Reading Enterprise). Finally, we are in the process of incorporating recent research on best practices for students with special needs into our future development plans. Many of these new features and offerings are slated to become active within the 2012–13 school year.

Setting Accommodations:

- Administer the test in a location with minimal distractions.
- Administer the test in a separate location.
- Administer the test in a small group.
- Administer the test individually.
- Administer the test in a room with special acoustics.
- Administer the test in a room with special lighting.
- Administer the test in a room with special furniture.

Presentation Accommodations:

- Magnify the size of the type.
- Read the testing directions orally to students.
- Sign the testing directions to students.
- Read the testing directions in students' native languages.
- Read the entire test orally to students.

Response Accommodations:

- Alter method of input from mouse to keyboard or vice versa, or to an assistive device.
- Allow students to use scratch paper while testing.

Requirement: 4.6

- Assign an assistant to input student responses

Timing Accommodations:

- Allow extended time. STAR Math Enterprise allows teachers and test monitors to extend time limits for students who have special needs. Extended time may be a valuable accommodation for English language learners as well as for some students with disabilities. Test users who elect the extended time limit for their students should be aware that STAR Math™ norms, as well as other technical data such as reliability and validity, are based on test administration using the standard time limits. The standard time limits are 90 seconds for practice questions and 180 seconds for actual test questions; extended time limits allow 180 seconds for practice questions and 360 seconds for actual test questions.

Miscellaneous Accommodations:

- Provide a drink or snack during testing.

STAR and the Topic of Alternate Assessments

Each STAR Enterprise assessment comes in one computer-adaptive version. Students who are performing above or below grade level do not need to take a different test, because the STAR tests will adapt to their ability levels, no matter where they fall within the K–12 learning progression for math or the K–12 learning progression for reading. This makes the tests highly suitable for many different types of students, including English language learners and students with disabilities. Following is more information about using STAR Enterprise with English language learners and students with disabilities.

English Language Learners: STAR Math Enterprise can be used successfully with English language learners to enhance the data picture for these students. For example, English language learners may benefit from testing directions being read in their native language and/or extended time limits. Test administrators can, if they choose, read aloud the entire STAR Math Enterprise test to students in their native language. However, students who have received math instruction in English typically will not have the equivalent mathematical vocabulary in their native language. Moreover, taking STAR Math Enterprise in English helps prepare students for high-stakes, end-of-year math assessments in English. It also allows the software to accurately predict how students will perform on those tests.

While STAR Math Enterprise is an English-only assessment, parent reports are available in English and Spanish.

Students with Disabilities: The STAR Enterprise assessments are suitable for use with students with disabilities. The difficulty of STAR items is “adapted” based on each student’s response pattern. If a student misses a question, the next question will be easier. If a student answers a question correctly, the next question will be more difficult. Because every student sees a different series of questions— and the same is true of one student over a period of growth (such as between fall and spring) as the software delivers more difficult questions—each individual STAR test is unique to each student’s needs at his or her level of ability. Thus, in many cases, the STAR Enterprise assessments eliminate the need for an

Requirement: 4.6

alternate assessment.

This method of testing reduces student frustration, since test questions tend to be very close to each student's actual ability level. Another benefit of using STAR with students with disabilities is that the tests gather valid and reliable information in a relatively short period of time. Thus, they are convenient for testing students who may require more time than usual to complete a test.

Please refer to the above section called *Allowable Accommodations* for information on which accommodations are permissible for students with disabilities taking the STAR Enterprise tests. Students with disabilities may particularly benefit from extended time limits or alternate settings.

In addition, many accommodations are already built into the STAR software or into the computer's operating system and are available to all students who take the tests. Here is further information on using STAR with students with disabilities:

Students with Visual Impairments: The introductory screens of the STAR assessments respond to the "High Contrast" accessibility feature within Windows and the "Switch to Black and White" accessibility feature in MAC OS 10. In addition, the assessment screens within STAR already provide visual contrast through a light background and black writing. STAR assessments are compatible with Mac OS 10's "Zoom In" accessibility feature, which allows users to magnify nearly all STAR screens.

Students with Hearing Impairments: The STAR assessments may be used with students who are deaf or hard of hearing using standard administration procedures. At the discretion of the teacher or specialist, STAR Math may be signed to a student if there is a question about the student's ability to understand the text. If this adaptation is used, the results of the assessment should be evaluated cautiously. It is reasonable to assume that the same adaptations or accommodations permitted for standardized or state assessments may be used with STAR Math.

Students with Limited Motor Skills: STAR Enterprise supports both mouse and keyboard input. Also, STAR offers accommodations for students with limited motor skills through the accessibility options built into a computer's operating system. Finally, the STAR assessments are compatible with most adaptive devices typically used by students with limited motor skills. These include add-on devices such as cursor or keyboard controllers.

External Review

The federally funded National Center on Response to Intervention (NCRTI) gave the STAR assessments some of its very highest ratings for screening and progress monitoring, thereby indicating that these assessments are among the best software tools available for RTI purposes, including the assessment of students in special education and intervention programs. To see the reviews of STAR, please visit the NCRTI website at http://www.rti4success.org/screeningTools and http://www.rti4success.org/progressMonitoringTools.

Requirement: 4.7
4.7 Describe procedures for completed student tests to be submitted for scoring and reporting purposes.
As soon as students complete a STAR assessment, the software automatically scores the test, and teachers have immediate access to numerous easy-to-read reports. Reports are accessible through the Renaissance Place Real Time platform, and they can be printed, saved, or viewed online.

Section 5 : Scoring and Reporting

Requirement: 5.1
5.1 Describe scoring procedures for all item types and test forms administered, including implemented quality control measures.
Once a student completes a STAR Enterprise assessment, educators have access to a variety of norm-referenced and criterion-referenced scores that demonstrate a student's performance in the five domains of the Core Progress learning progression for math.[10] The primary score—the scaled score—is generated from a student's responses to the items administered to that student. This score is available immediately after a student completes an assessment. As mentioned, all items have been calibrated using the Rasch IRT model. A Rasch ability estimate is calculated for the pattern of the student's right and wrong answers, using maximum likelihood ability estimation. STAR Math Rasch ability estimates are transformed to the scaled score, which spans a 0–1400 unit scale with unit intervals. This process employs an equipercentile equating transformation of the Rasch score to equivalent scaled scores on the original version of STAR Math, which was not based on item response theory. The scaled score is then used to access numerous norm-referenced scores. Norm-referenced scores provided by STAR Math Enterprise are as follows: • **Student Growth Percentile (SGP)** compares a student's growth to that of his or her academic peers nationwide. An SGP provides a measure of student growth from one SGP testing window (fall, winter, or spring) to the next, relative to other students with similar initial STAR scores. A student's SGP contributes significantly to an educator's understanding of how well a student is doing in school because it indicates what kind of progress the student is making. For example, a student may be performing at a low level, yet be growing at a significant rate. Conversely, a high-performing student could be stagnating. • **Percentile Rank (PR)** compares the student's performance with other students of the same grade and month of the school year. Percentile rank indicates how a student performed compared to

[10] To learn more about Core Progress for Math, see the document titled *Core Progress for Math: An Empirically Validated Learning Progression* provided in appendix 2.

Requirement: 5.1

other students who took STAR Math tests as a part of the national norming program.

- **Normal Curve Equivalent (NCE)** transforms the percentile rank value to an equal-interval scale. NCEs are useful for purposes of statistically manipulating norm-referenced test results, such as when interpolating test scores, calculating averages, and computing correlation coefficients between different tests. For example, in STAR Math score reports, average percentile ranks are obtained by first converting the PR values to NCE values, averaging the NCE values, and then converting the average NCE back to a PR.

- **Grade Equivalent (GE)** expresses the grade and month of the school year for which the student's present performance was the median score in the distribution of students, thus providing a grade and month of the academic year for which the student's present performance was the median.

Quality Control Measures

The STAR Enterprise assessments are computer-administered and computer-scored tests, eliminating the human factors of judgment and test administration that can influence test scores. Students are assessed fairly and in an accurate, standardized way. Reports deliver valuable information about each student, and that information is not subject to differences in how the test was administered or the subjectivity that can come into play when human judgment is a factor of the scoring process. This is especially important in today's classrooms, since many district and state administrators are now using assessment results to make high-stakes decisions about academic readiness and educator effectiveness.

Renaissance Learning makes every effort to help educators implement the STAR assessments with fidelity. Free educator's guides and other resources are available online, and a variety of professional development opportunities are also available to address essential testing conditions and protocols to maintain the fidelity of the implementation.

Requirement: 5.2

5.2
Describe the type of reporting provided (e.g, static and/or dynamic, electronic and/or paper-based, item-level, strand-level, and/or test-level scoring). Include approximate timelines for score reports to be available to divisions, how score reports will be accessed and/or obtained, and samples of student, class, school, and division score reports and sample record layouts for electronic data files.

The STAR Enterprise assessments provide a variety of reporting options, as explained below. All reports are available to school divisions immediately after students complete the assessment. The reports are accessible online via the Renaissance Place Real Time platform.

Static and Dynamic Reporting: Several STAR reports offer onscreen dynamic drill-down capabilities, which allow users to focus on one particular grade level, one particular teacher's class, or one particular student. Most STAR reports offer flexibility through user-selected options. Once the user is satisfied with

Requirement: 5.2

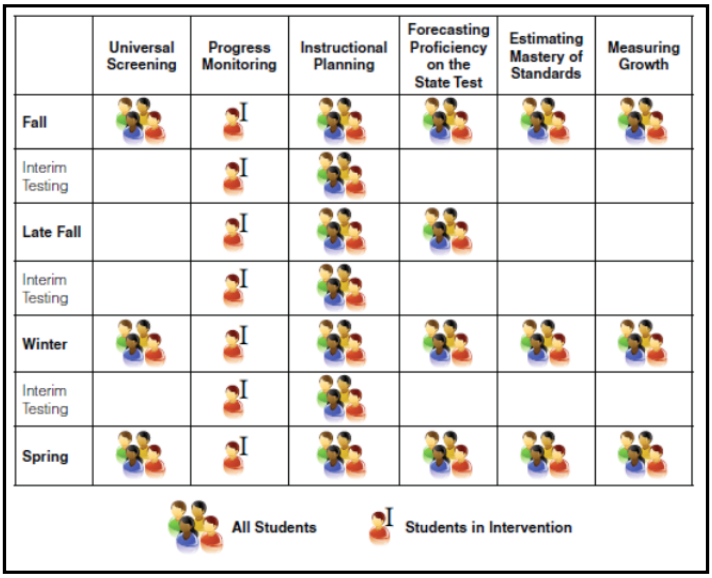
the report options, a report can be printed or saved as a static PDF file.

Electronic and Paper-Based Reporting: All STAR reports can be printed, saved, or viewed online.

Strand-Level Scoring: Both the STAR Student Diagnostic Report and Instructional Planning Report show student performance by strand, or domain. The Student Diagnostic Report provides domain scores for each of the four math domains. These scores estimate a student's percent of mastery of skills in each of the domains for that grade level. Subsequent pages of this report identify a student's skill area scores, which estimate a student's percent of mastery within each skill area assessed by STAR Math Enterprise. The report also highlights skills that a student is likely ready to learn and practice, based on his or her domain scores.[11] Similarly, the Instructional Planning Report uses the student's scaled score (0–1400) to identify skills for instruction that individual students and groups need to learn in each domain in order to advance in the reading learning progression.

Test-Level Scoring: Each STAR assessment provides a test-level scaled score that is linked to norm- and criterion-referenced scores, such as the percentile rank, grade equivalent, normal curve equivalent, student growth percentile, and more. To learn more about these scores, please see pages 100–111 of the *STAR Math Enterprise Technical Manual* provided in appendix 5.

Additional Reporting Functionalities: As a result of our diligent efforts to identify a learning progression for reading and map the STAR Enterprise scaled scores to this learning progression, the STAR Enterprise assessments provide comprehensive reporting to help teachers and administrators improve teaching and learning. All reports are available immediately after a student completes a STAR assessment and are accessible via the online Renaissance Place Real Time platform. Samples of the reports described below are provided in the document titled *Key Report Samples: STAR* Enterprise in appendix 2.



- **Universal screening.** *Which students are reaching benchmark goals? Which need intervention?* STAR assessments generate screening reports that show which students are succeeding with core instruction and which may need intervention. These reports are customizable to compare student performance to school or district, and state test benchmarks.

- **Instructional planning.** *What are my students ready to learn next?* STAR assessments identify skills for instruction individual students and groups need to learn in order to advance in the learning progression. Each student's Instructional Planning Report graphs the student's current

[11] A sample STAR Math Enterprise Diagnostic Report can be found in appendix 5.

Requirement: 5.2

and projected scaled score against school, district, or RTI benchmarks. In the future, Virginia's SOL benchmarks will also be available. This information provides teachers a fast and efficient way to understand where students are in the learning progression and the types of skills they are ready to learn next.

- **Learning progression.** *What prerequisite skills does my student need in order to fill knowledge gaps?* STAR Enterprise simplifies lesson planning by providing an interactive web portal to the Core Progress learning progression. Teachers can search within Core Progress for core objectives and prerequisites skills. Through this portal, teachers also access instructional resources, sample items, and worked examples which can be printed or displayed on an interactive whiteboard.

- **Progress monitoring.** *Is my student responding to the intervention?* Teachers first set an individualized goal and intervention for students using the STAR Goal-Setting Wizard. The Goal-Setting Wizard suggests personalized goals based on a decile-based growth model. This tool automatically graphs a student's progress toward goals on the Student Progress Monitoring Report. Students can be retested multiple times with STAR—even weekly—without the risk of item overexposure for progress monitoring.

- **Mastery of standards.** *Are my students mastering Virginia Standards of Learning?* STAR assessments help educators gauge progress toward mastery of state standards at the student, class, school, and district level. Renaissance Learning's nationally recognized standards experts have placed the state standards from all 50 states and the Common Core on the STAR scale using a systematic, qualitative method. As a result, the STAR standards reports generate reliable estimations of student progress toward standards mastery.

- **Prepare for state testing**. *Are my students on track to reach proficiency on the Virginia Standards of Learning (SOL) Test?* STAR Performance Reports predict performance on high-stakes tests. STAR offers many reports that include forecasts of state test performance. Using a psychometrically sound growth model, STAR performance predictions are based on extensive research equating the STAR scale to proficiency cuts for each performance category on the state test. Predictions account for the amount of growth that typically occurs between the date of the last STAR test taken and the date of the state test.[12]

- **Measure growth.** *Are each of my students reaching their maximum potential?* The STAR Growth Report and the interactive Growth Proficiency Chart help teachers, principals, and district staff focus not just on achievement, but on potential. Emulating the Colorado Growth Model, STAR's Growth Proficiency Chart combines achievement and growth in one view. As a result, educators can start to see their data in new ways, such as the following:

    - How can we make sure all students, including high-achieving students, are challenged and growing every year?

[12] State Performance Reports will be available for STAR Math Enterprise once a linking study is complete. The linking process generally takes about six weeks to complete; Renaissance Learning will work with your state to determine an appropriate schedule for the project as soon as sufficient data are available.

Requirement: 5.2

- How can we make sure each student is moving forward, no matter what their starting point?
- How can we make sure that students with low growth and low proficiency are getting the services they need? Are these students responding to intervention? Is the intervention being implemented with fidelity? Are these students attending school?

Requirement: 5.3

5.3
Describe all data tools available to school division staff for the analysis of data and the creation of customized reports.

As described below, Renaissance Learning offers a variety of real-time data analysis and instant-reporting options to keep school division administrators, teachers, students, and parents aware of students' reading achievement and progress.

Web-Hosted, Centralized Data

The Renaissance Place Real Time platform consolidates all Renaissance Learning software and provides a central location for assessment results and daily test scores across the entire learning organization. This system is accessible via a variety of hardware platforms, including Macintosh and Windows operating systems. In addition, Renaissance Learning is in the process of developing the STAR assessments to be supported on Apple's iOS. This functionality is scheduled to be available in spring 2013. Accessing the platform requires a user name and password. To take a STAR assessment, students simply log in to any Internet-connected computer at the school and take the test. Afterward, the teacher or administrator logs in to any Internet-connected computer and runs easy-to-read reports.

This centralized data management system makes it easy to add and maintain student information in real time because the information is imported only once and updated in one central location. There is no need to manually score the tests or transfer scores from one system to another, and all of the software and content updates are handled by Renaissance Learning. It's also easy to add and maintain other Renaissance programs. And when schools host their Renaissance Learning software in our Enterprise-Class Data Center, they receive all new content, features, and software updates instantly and automatically, and at no cost to your districts. This frees up state and district technical staff to focus on other important issues.

User-Friendly Reports for All Stakeholders

The online Renaissance Place system manages student data and generates customized reports for all Renaissance programs, including the STAR assessments. Educators can control what information they access and view, and they can organize this information based on local preferences.

Administrators can customize many of the reports in Renaissance Place to view information about participation and performance across the district and by various demographic subgroups (for example,

Requirement: 5.3
students receiving free and reduced lunch, English language learners, etc.). In addition, each STAR assessment immediately produces numerous easy-to-read reports, and educators can choose to run them at the student, classroom, school, or district level. Users can specify a group of students whose reports are to be printed in one batch. They can also choose to drill down to focus on one particular grade level, one particular teacher's class, or one particular student. Reports can be printed or viewed online. Descriptions of key STAR reports are provided in response to item 5.2, and report samples are provided in appendix 2.