The Science Behind i-Ready's Adaptive Diagnostic

Curriculum Associates

i-Ready

Table of Contents

Page

Appendix

An Ideal Assessment

Adaptive assessments are not new. However, the rise of technology and the growth of computer usage in schools have made large-scale, computer-adaptive testing more feasible and increasingly common. Adaptive assessments, like i-Ready Diagnostic, leverage advanced technology to provide a deep, customized evaluation of every student and to track student growth consistently and continuously over a student's entire K–12 career. This is especially beneficial for identifying gaps from prior years as districts transition to the Common Core. i-Ready also provides valid and reliable growth metrics across a district and school environment to optimize administrative decision making for long-term performance improvements.

Adaptive Assessments Maximize Information on Student Performance

Adaptive assessments are frequently chosen for their high precision and efficiency, allowing educators to pinpoint student needs more accurately and in less time than with traditional fixed-form assessments. By dynamically selecting test items based on student response patterns, adaptive assessments are able to derive large amounts of information from a limited number of test items and can adapt to students with low and high ability to get a better assessment of student performance.

Many educators familiar with fixed-form assessments may have some questions about the information gained from an adaptive assessment: With a limited number of test items, how can I be sure of the skills my students have and have not mastered? How do I know that my student has mastered a skill, if he has not been tested on it? This is where i-Ready's sophisticated adaptive logic and a bank of thousands of test items come into play—pinpointing students' needs in reading and math down to the domain and sub-skill levels.

Great effort was taken in building out the i-Ready item bank and adaptive logic to ensure that, for example, when a 5th grade student is still lacking mastery of Grade 4 standards, the system provides the teacher with what would help the student the most—recommendations for the below-level skills the student still lacks. On the other hand, when the student's initial performance demonstrates the mastery of higher level skills, no time is wasted on needlessly assessing lower-level prerequisite skills.

For example, if a student is able to correctly solve a two-digit multiplication problem that requires re-grouping, then there is no need to assess that student on single-digit addition, a skill that is necessary to solve the initial multiplication problem. Yet, with a fixed-form test, multiple test items would be required to gain this same information! Because i-Ready Diagnostic already knows the student has a very high probability of answering questions aligned to these standards correctly, it tries to gain more information about the student's ability level by providing questions that will offer more information about the student.

Understanding the Difference between Fixed-Form and Adaptive Assessments

To explain the difference simply, let's consider a test item:

Mary goes to the coffee shop. She can purchase a pound of coffee for $9 or 12 ounces for $7. Which is the better bargain?

The above example tests students on three different sets of skills:

1) Do they possess the algebraic thinking skills to set up the problem to compare fractions?

2) Do they know their measurement conversions?

3) Do they possess the computational skills to manipulate and solve the problem?

On a fixed-form assessment, this problem may simply be considered an example of comparing fractions for a student who is "average;" it may in fact be too easy or difficult for a number of students.

On an adaptive assessment, items are tagged so that trends can be seen and more information can be efficiently gathered. Once a student fails an item, additional items assessing the relevant sub-skills are drawn to get to the root cause of getting the first question wrong.

This is powerful to educators as it drives more precise targeting of instruction.

Adaptive Assessments Promote Accurate Measurement of Growth Across a Student's Career

i-Ready makes measuring student growth easy, because of its use of a vertical scale for scoring. Think of it like a growth chart seen at a pediatrician's office—every child can be measured on one chart. Similarly, i-Ready uses a vertical scale to measure which skills a student has gained from one point in time to the next, on a "chart" of skills that spans kindergarten through 12th grade. Educators can thereby measure student growth on a consistent scale throughout a student's entire career. Because i-Ready Diagnostic was built on the Common Core, this "chart" consists of Common Core skills expected of students at each grade level.

For example, consider a student who takes a fixed-form summative assessment at the end of each year in grades 3, 4, and 5. Each year he answers 60% of the items correctly on the test. Because the fixed forms for each grade are different, the percent correct does not tell the teacher how much growth the student has made. Alternatively, if this student took an i-Ready Diagnostic assessment at the end of each year, his placement may go from Level 1 the first year, to Level 3, the next year and Mid 5 the following year, measuring how much growth the student has made from year to year.

Key Distinctions of Fixed-Form and Adaptive Assessments

Fixed-Form Assessment

Assesses proficiency on grade-level skills, but does not allow educators to measure student proficiency on the same scale from year to year

- Fixed forms, fixed item selection
- Presents items based on prior design
- Can be paper- or computer-based
- Narrower scope (single grade level)
- Score usually presented as percent correct—e.g. 90%
- Test has difficulty providing detailed information about very high performing or very low performing students

Adaptive Assessment

Assesses proficiency on both on-grade and off-grade level skills without the need for additional test items and testing time; a vertical scale provides a consistent metric for measuring student progress across multiple grade levels

- Adaptive forms, dynamic item selection
- Presents items based on ongoing calculations of student ability
- Computer-based
- Broader scope possible (multiple grade levels)
- Score presented on the spectrum of ability across grades—e.g. 750 (on an 800-point vertical scale)
- Questions within the test adjust to the student's ability

Adaptive Assessments Help Administrators Make Long-Term Decisions and Measure Impact

For administrators, an adaptive assessment has proven to be the most precise measure of student growth (Growth, Precision, and CAT: An Examination of Gain Score Conditional SEM by Tony D. Thompson, Research Report, December 2008). This real-time visibility enables immediate, effective course corrections.

Administrators using i-Ready are given insight into:

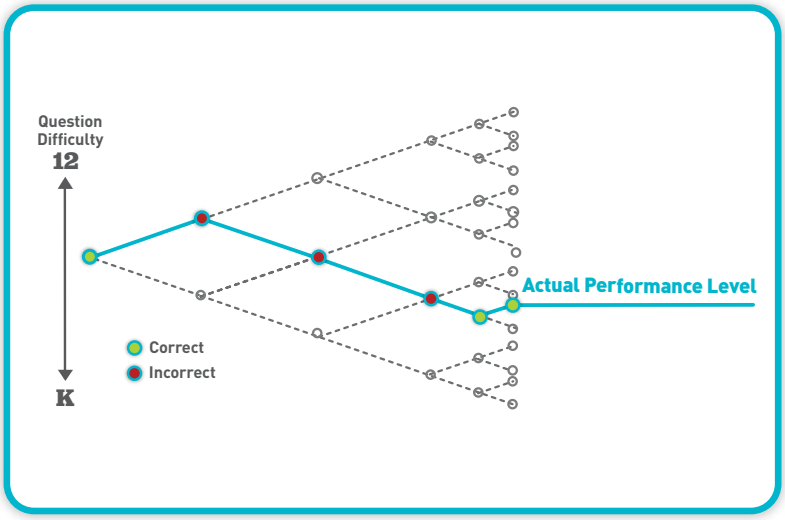
- Percent of students performing below, on, and above grade level
- Percent of students on track to meet annual growth expectations
- Details by school, grade, class, and student

How i-Ready Diagnostic Works

Adaptive Structure:

i-Ready Diagnostic adapts, or adjusts, until it finds exactly the level at which students need to receive instruction.

- When students answer questions correctly, i-Ready gives them more challenging questions
- When students answer questions incorrectly, i-Ready gives them less challenging questions
- This process continues. In the end, i-Ready pinpoints which skills each student has mastered and which skills need improvement



How i-Ready Diagnostic Works (continued)

Upon completion of the adaptive Diagnostic, multiple types of scores are reported by i-Ready to enable a well-rounded view of each student's proficiency levels:

- **Scale Scores** – a common language across grades and schools. Scale scores put everything on a single continuum so that educators can compare across grade levels. They provide a metric, which indicates that a student has mastered skills up to a certain point and still needs to work on skills that come after that point
- **Placement Levels** – the practical day-to-day language that helps teachers determine what grade level of skills to focus on with a particular student. Placement levels indicate where students should be receiving instruction
- **Norm Scores** – identify how students are performing relative to their peers nationwide. Based on a nationally representative sample of students taking the i-Ready Diagnostic, they specify a student's ranking compared to students in the same grade. For example, if a student's percentile rank is 90%, this means the student scored better than or equal to 90% of her national peers from the same grade level
- **Lexile® Measures** – developed by MetaMetrics®, Lexile measures are widely used as measures of text complexity and reading ability, allowing a direct link between the level of reading materials and the student's ability to read those materials
- **Quantile® Measures** – developed by MetaMetrics, the Quantile Framework for Mathematics is a unique resource for accurately estimating a student's ability to think mathematically and matching him/her with appropriate mathematical content

Educators are also given explicit qualitative information on each student's abilities:

- The specific skills students have mastered and those that need to be prioritized for instruction
- **Standard-by-standard analysis** that details student performance against Common Core standards and sub-skills

i-Ready Diagnostic: Quick Facts

Assessment Length:

- Students receive 54–72 items per subject
- Students typically take 30–60 minutes per subject to complete the Diagnostic. Average duration varies by subject and grade level, with grades K–3 tending towards the shorter end of the range. Additionally, variability exists in every grade given different student performance levels.

Content Areas:

i-Ready assesses across the following content areas, also known as domains:

Reading

- Phonological Awareness
- Phonics & Word Recognition
- Vocabulary
- Reading Comprehension: Literature
- Reading Comprehension: Informational Text

Mathematics

- Counting and Cardinality
- Number & Operations in Base Ten
- Number & Operations – Fractions
- The Number System
- Number and Quantity
- Operations & Algebraic Thinking
- Ratios and Proportional Relationships
- Expressions and Equations
- Functions
- Algebra
- Measurement and Data
- Statistics and Probability
- Geometry

Underlying Theory

Computer adaptive testing and the Rasch Item Response Theory model form a strong foundation for ensuring valid inferences are reported by i-Ready Diagnostic.

In 1960, Georg Rasch developed the Rasch Item Response Theory Model. In this model, the logit value or difficulty level of the items are independent of the ability level of the student. These logit values can also be used to describe the ability level of the student. Using the Rasch Equation, it is possible to calculate the probability of success that a student of a certain ability would have with an item of a certain difficulty. In fact, if the difficulty level of the item and the ability level of the student are the same, then the student will have an even chance of answering the item correctly or incorrectly. This phenomenon is shown graphically in Appendix II using a Wright Map to show the progression of item difficulty through the grades.

i-Ready Diagnostic uses both adaptive testing and item response theory to determine the ability level of the student. From extensive field-testing of items with over 2,000,000 students, there exists a very strong and reliable foundation for determining the difficulty level of each item as well as each indicator group. An indicator group is a set of items aligned to a specific skill. From the ability level of the student and the difficulty level of these indicators, i-Ready can make probabilistic inferences about what students know and are likely able to do. Using this information, the assessment can accommodate students of far-ranging ability levels. Moreover, the results from the i-Ready Diagnostic can pinpoint students' strengths and provide teachers with actionable information on what students should work on next.

Designed for Common Core Success

Successful transition to the CCSS requires visibility into student performance on the more rigorous assessments that are to come. Using measures that are highly correlated to Common Core-based assessments is a critical step, and i-Ready offers that solution.

Common Core support embedded into the entire program

- Covers more than 90% of assessable standards in Grades K–8 as well as most standards in High School Math and Reading
- Assesses both procedural and conceptual fluency
- Presents a range of challenging informational and literary texts, including authentic texts and multimedia items
- Prepares for College and Career Readiness expectations, including the Smarter Balanced Assessment Consortium (SBAC) and Partnership for Assessment of Readiness for College and Careers (PARCC) expectations

Proven to be Valid and Reliable

- Developed and reviewed by well-known experts in Educational Measurement, Computer Adaptive Testing, Mathematics, English Language Arts and the Common Core
- Adheres to the Standards of Psychological and Educational Testing (AERA, 1999) and was independently audited for adherence to the Standards by researchers from the University of Massachusetts at Amherst
- Extensive stand-alone and embedded field testing with over 2 million students
- Approved by high profile review committees for New York, Ohio, Virginia, Chicago, Dallas, and many more
- Strong test metrics: Low SEMs; good item discrimination among students of different abilities
- Linked to National Measures recognized by Common Core: Lexiles measures, Quantile measures (refer to Appendix III for more details on these linking studies)
- Strongly correlated to Common Core assessments based on third-party research from the Educational Research Institute of America (ERIA)

Text complexity backed by research

During the development of all passages within the Diagnostic, the recommendations from the Common Core State Standards that readability be evaluated both quantitatively and qualitatively were followed. Lexile and Flesch-Kincaid (F/K) are the quantitative tools used, which provide scores based primarily on the length of syllables, words, and sentences in a text. The Lexile Range scores from MetaMetrics as well as the Flesch-Kincaid tool in Word were used to focus in on proper readability levels. In addition, using a qualitative approach, content experts reviewed all i-Ready reading passages for the qualitative measures that contribute to text complexity. All items went through extensive field testing to confirm appropriate grade placement, and passages were reviewed by subject matter experts for their appropriateness for Reader and Task complexity.

i-Ready® Accurately Predicted Individual Proficiencies on a Common Core-Based Assessment

Highly Correlated

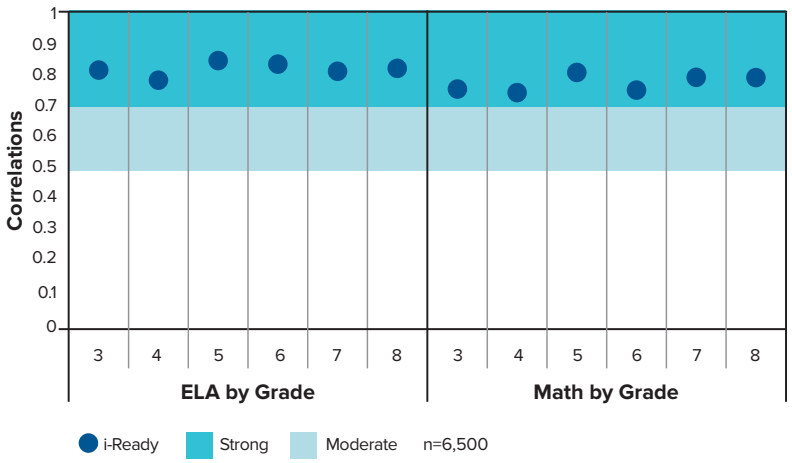
Strong correlations mean the right preparation

In a recent independent study conducted by the Educational Research Institute of America, i-Ready was found to have strong correlations to the 2013 NY State Assessment, one of the first truly Common Core-based summative assessments (correlations ranged from .77-.85 across grades and subjects).

Why it matters

Because of these strong correlations, you can be confident that your students are gaining crucial exposure to the key skills and concepts that they need for success on new, more rigorous assessments.

i-Ready and Common Core-based state assessment correlations



Predictive

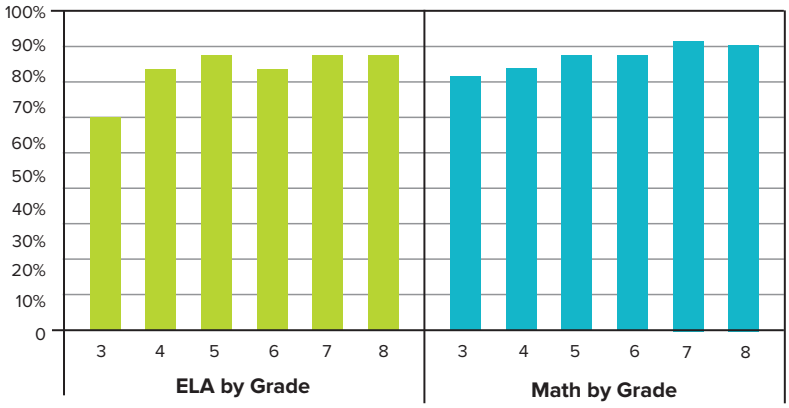
Critical insight to inform decisions

Correlations are just the beginning of the story; in addition, i-Ready successfully predicted proficiency on this Common Core-based assessment for 85% of students. In other words, before the actual state assessment, i-Ready is able to identify how students are likely to perform. (Refer to the i-Ready NY Validity Study for further details).

Why it matters

By uncovering specific Common Core needs early, you'll be able to match instructional priorities to those needs—months before students have to take the state assessment in the spring.

% of students whose proficiency on the 2013 Common Core-based state assessments were correctly predicted by i-Ready



Using Assessment Data to Meet Individual Student Needs

The adaptive logic enables a deep, customized evaluation of every student, tracking student growth consistently and continuously over a student's entire K–12 career and identifying gaps from prior years and areas for further enrichment.

The Diagnostic results directly drive instantaneous reports that detail each student's proficiency levels and areas of need, highlighting immediate next steps for instruction and enabling individualized learning programs. The reports (i.e., Student Profile Report pictured below) provide teachers with an action plan to make targeted, differentiated instruction a reality. The system also provides the tools to deliver that instruction in any style learning environment —including both online lessons and teacher-led instruction.

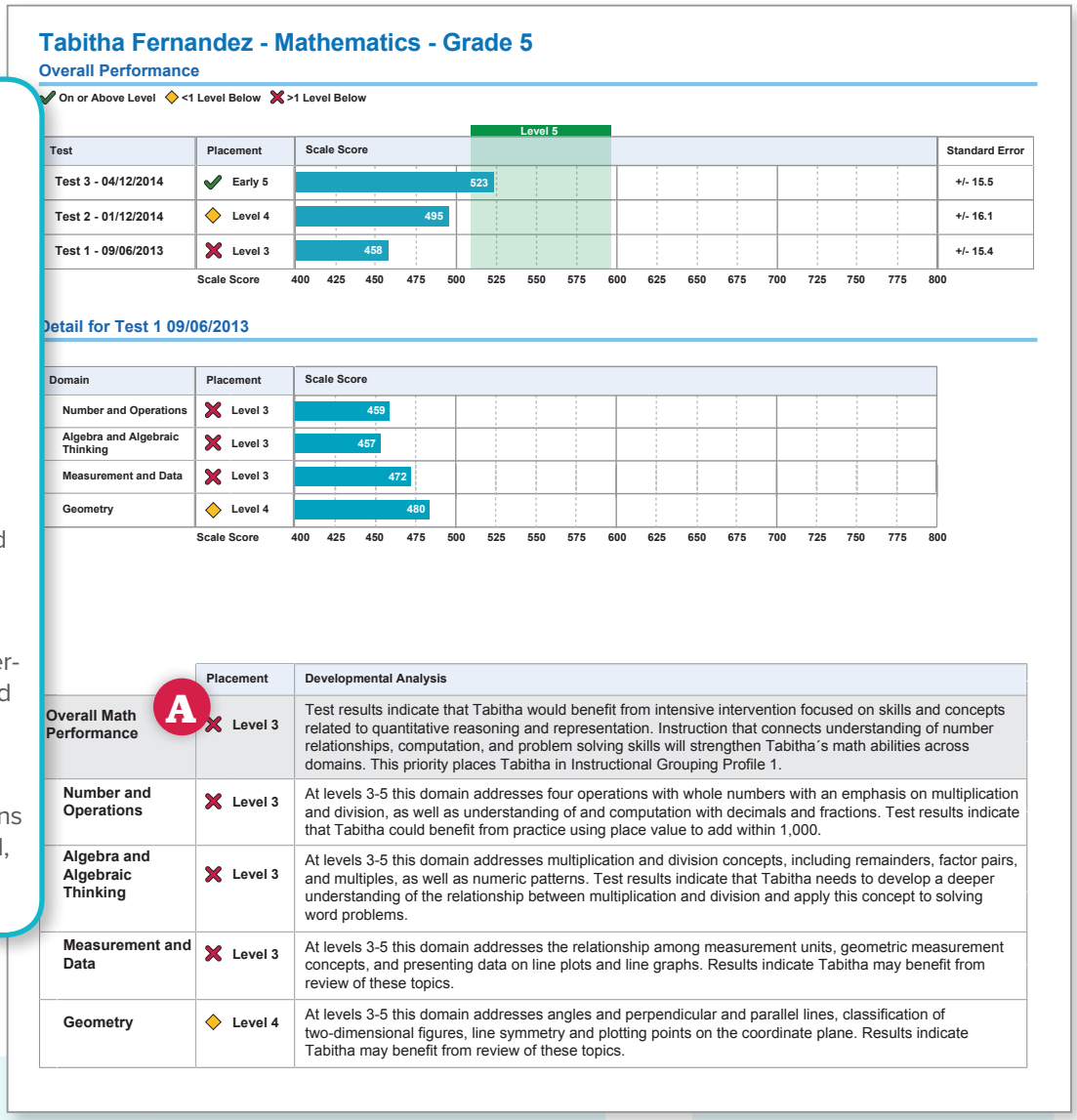
A Developmental Analysis:
Gives teachers more insight into the domain placement levels. It is used to inform instructional decisions by highlighting the most important areas of need.

B What (Student) Can Do:
Identifies student strengths in each domain.

C Student Next Steps:
Identifies the skills that should be prioritized for instruction.

D Tools for Instruction:
Delivers downloadable teacher-led lessons on areas identified in Next Steps.

E Recommended Products:
Points directly to specific lessons in print resources for targeted, data-driven instruction.



Tabitha Fernandez - Mathematics - Grade 5

Overall Performance

✓ On or Above Level ◆ <1 Level Below ✗ >1 Level Below

Level 5

Test	Placement	Scale Score	Standard Error
Test 3 - 04/12/2014	✓ Early 5	523	+/- 15.5
Test 2 - 01/12/2014	◆ Level 4	495	+/- 16.1
Test 1 - 09/06/2013	✗ Level 3	458	+/- 15.4

Scale Score 400 425 450 475 500 525 550 575 600 625 650 675 700 725 750 775 800

Detail for Test 1 09/06/2013

Domain	Placement	Scale Score
Number and Operations	✗ Level 3	459
Algebra and Algebraic Thinking	✗ Level 3	457
Measurement and Data	✗ Level 3	472
Geometry	◆ Level 4	480

Scale Score 400 425 450 475 500 525 550 575 600 625 650 675 700 725 750 775 800

	Placement	Developmental Analysis
Overall Math Performance (A)	✗ Level 3	Test results indicate that Tabitha would benefit from intensive intervention focused on skills and concepts related to quantitative reasoning and representation. Instruction that connects understanding of number relationships, computation, and problem solving skills will strengthen Tabitha´s math abilities across domains. This priority places Tabitha in Instructional Grouping Profile 1.
Number and Operations	✗ Level 3	At levels 3-5 this domain addresses four operations with whole numbers with an emphasis on multiplication and division, as well as understanding of and computation with decimals and fractions. Test results indicate that Tabitha could benefit from practice using place value to add within 1,000.
Algebra and Algebraic Thinking	✗ Level 3	At levels 3-5 this domain addresses multiplication and division concepts, including remainders, factor pairs, and multiples, as well as numeric patterns. Test results indicate that Tabitha needs to develop a deeper understanding of the relationship between multiplication and division and apply this concept to solving word problems.
Measurement and Data	✗ Level 3	At levels 3-5 this domain addresses the relationship among measurement units, geometric measurement concepts, and presenting data on line plots and line graphs. Results indicate Tabitha may benefit from review of these topics.
Geometry	◆ Level 4	At levels 3-5 this domain addresses angles and perpendicular and parallel lines, classification of two-dimensional figures, line symmetry and plotting points on the coordinate plane. Results indicate Tabitha may benefit from review of these topics.

Tabitha Fernandez – Mathematics – Grade 5

Test 1-09/06/2013	Placement	Scale Score
Number and Operations	Level 3	458

Scale Score 0 50 100 150 200 250 300 350 400 450 500 550 600 650 700 750 800

Building Number and Operations Skills

Number and Operations in grades K-8 focuses on representing, comparing, and performing operations with numbers. As in the CCSS, this domain includes whole numbers, decimals, fractions, integers, and irrational numbers, and emphasizes both conceptual understanding and computation. In grades 3-5, students gain an understanding of fractions and decimals and develop fluency with all four operations involving whole numbers, fractions, and decimals.

What Tabitha Can Do (B)

Results indicate that Tabitha can likely do the skills shown below.

Base Ten

- CC Model three-digit numbers.
- CC Compare and order three-digit numbers.
- CC Know multiplication facts through 9 x 9.

Fractions

- CC Identify fractions ($^1/_2$, $^1/_4$, $^3/_4$) as parts of a whole using pictures.
- CC Identify fractions that name part of a whole (denominators of 2, 3, 4, 5, 6, 8, 10, 12).

Next Steps for Instruction (C)

Results indicate that Tabitha will benefit from instruction and practice in the skills shown below.

Base Ten
Know division facts through 81 ÷ 9.
Add multi-digit numbers.
Subtract multi-digit numbers.
Multiply two-digit numbers by one-digit numbers.

Fractions
Identify fractions shown on a number line.
Use models to find equivalent fractions.
Write equivalent fractions, including fractions in simplest form.
Express fractions with denominators of 10 or 100 as decimals.
Decompose a fraction into a sum of fractions with like denominators.
Add and subtract fractions with like denominators.

Tools for Instruction (D)

- Know Division Facts
- Subtract Multi-Digit Numbers
- Fractions on the Number Line
- Find Equivalent Fractions

Recommended Print Products (E)

If you have this product...	Use...
Ready® Common Core	**Grade 3** Lesson 4: Understand the Meaning of Division, p. 30 Lesson 5: Understand How Multiplication and Division Are Connected, p. 36 Lesson 6: Multiplication and Division Facts, p. 42 Lesson 9 Use Place Value to Add and Subtract, p. 72 Lesson 10: Use Place Value to Multiply, p. 84 Lesson 15: Understand Fractions on a Number Line, p. 138 Lesson 16: Understand Equivalent Fractions, p. 144

Learn More

Development Led by Expert Advisors

Technical Advisory Committee Members

- **Dr. Richard Brown** | *Founder and CEO of West Coast Analytics*
  - Former Associate Professor, Psychometrician Rossier School of Education of the University of Southern California
  - Former Director of National Center for Research on Evaluation, Standards, and Student Testing (CRESST) at UCLA
  - One of the primary psychometricians for i-Ready Diagnostic since its inception

- **Dr. Anne Collins** | *Director of the mathematics programs at Lesley College and the Lesley School of Education, and the Achievement Center for Mathematics*
  - Past president of both the Association of Teachers of Mathematics in New England and the Association of Teachers of Mathematics in Massachusetts
  - Served as an elected member of the Board of Directors for the National Council of Teachers of Mathematics (NCTM)
  - Active member of the Association of State Supervisors of Mathematics
  - Elected into the Massachusetts Mathematics Educators Hall of Fame in 2005

- **Dr. James W. Cunningham** | *Professor Emeritus of Literacy Studies, University of North Carolina at Chapel Hill*
  - Known for his research on text complexity and reading comprehension. His work has been featured in many prominent publications, including the Journal of Literacy Research and Reading Research Quarterly
  - Member of the IRA Reading Hall of Fame

- **Dr. Roger Farr** | *President and Founder of the Educational Research Institute of America (ERIA)*
  - Over 50 years of experience in the educational field
  - Author of numerous publications and a former president of IRA

- **Dr. Andrew Ho** | *Professor of Education at the Harvard Graduate School of Education*
  - His research critiques and clarifies educational accountability metrics, including proficiency, growth, achievement gaps, and value-added
  - Member of the National Assessment Governing Board and a recipient of the Jason Millman Promising Measurement Scholar Award from the National Council on Measurement in Education

- **Dr. Mark Pomplun** | *Executive Director of Assessment and Accountability, St. Charles, Il Community Unit School District 303*
  - Previously served as Principal Research Scientist at Riverside Publishing Company

- **Dr. Stephen Sireci** | *Professor of Education Policy, Research, and Administration, and Director of the Center for Educational Assessment in the College of Education at the University of Massachusetts Amherst*
  - His research focuses primarily on educational test development and evaluation, particularly issues of validity, cross-lingual assessment, standard setting, and computer-based testing
  - Worked with Curriculum Associates on conducting an audit of i-Ready Diagnostic's adherence to the Standards, as well as reviewing growth models and placements

Expert Curriculum Advisors

- **Dr. Richard Bisk** | *Chair and Professor of Mathematics at Worcester State University*
  - Advisor to the Massachusetts Department of Education in the development of the Guidelines for the Mathematical Preparation of Elementary Teachers
  - Expert on Singaporean mathematics education

- **Dr. David Chard** | *Dean of the Annette Caldwell Simmons School of Education and Human Development at Southern Methodist University*
  - Research review panelist at both state and national levels, including panels of the National Science Foundation and U.S. Department of Education
  - Awarded more than $11 million in deferral, state, and private grants since 1993

- **Dr. Cathy Seeley** | *Senior Fellow at the Charles A. Dana Center at the University of Texas at Austin*
  - Veteran mathematics educator and change facilitator with 35 years of experience at the local, state, and national levels; works on state and national policy and improvement efforts in mathematics education
  - Prior president of the National Council of Teachers of Mathematics (NCTM) from 2004 through 2006, and currently an active member of the council

- **Dr. Lori Helman** | *Associate Professor in the Department of Curriculum and Instruction at the University of Minnesota*
  - Many years of bilingual teaching experience at the early grades; leads new teacher induction programs
  - Co-Director of the Minnesota Center for Reading Research

Conclusion

In summary, i-Ready Diagnostic is a computer-delivered, adaptive assessment in Reading and Mathematics for students in Kindergarten through High School. This assessment was developed to serve several purposes:

- Accurately and efficiently assess student knowledge by adapting to each student's ability for the content strands within each subject. Offer an accurate assessment of student knowledge, which can be monitored over a period of time to measure student growth
- Provide valid and reliable information on skills students are likely to have mastered and the recommended next steps for instruction
- Link assessment results to instructional advice and student placement decisions

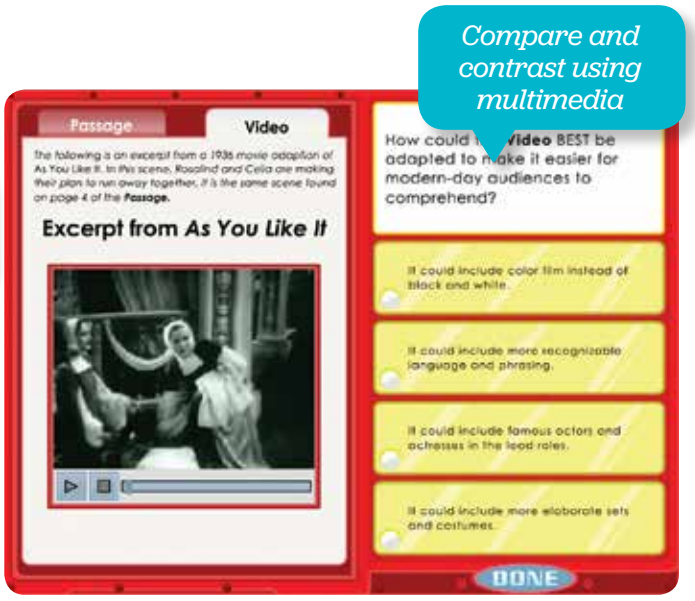
APPENDIX I

Sample Diagnostic Items

All items within the Diagnostic were specifically built to assess students against key Common Core skill areas. Below are sample Diagnostic items from both Reading and Math, across multiple grades. Features technology-enhanced items as recommended by SBAC and PARCC.

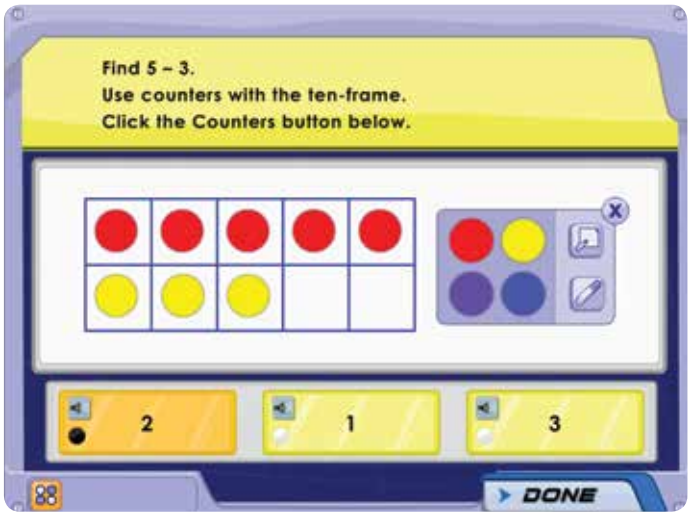
Reading

Level 3 – Reading Comprehension

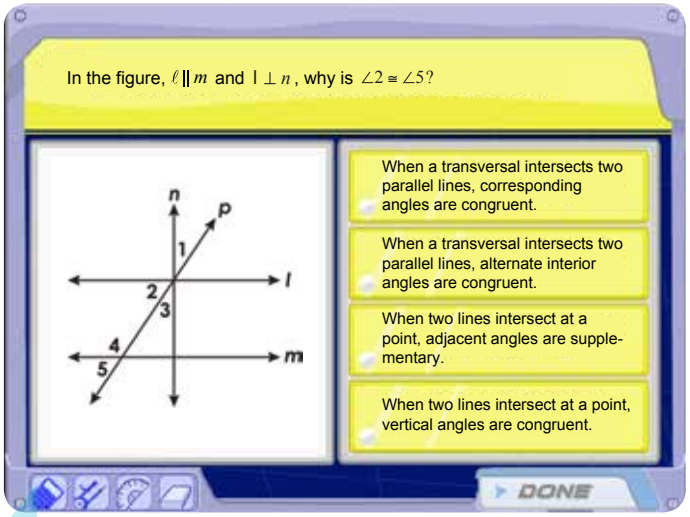


Level 12 – Reading Comprehension

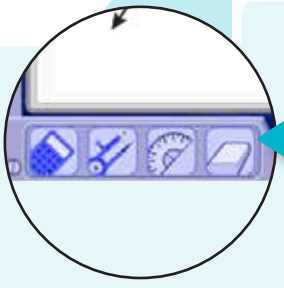
Mathematics



Level K – Number and Operations



Level 10 – Geometry

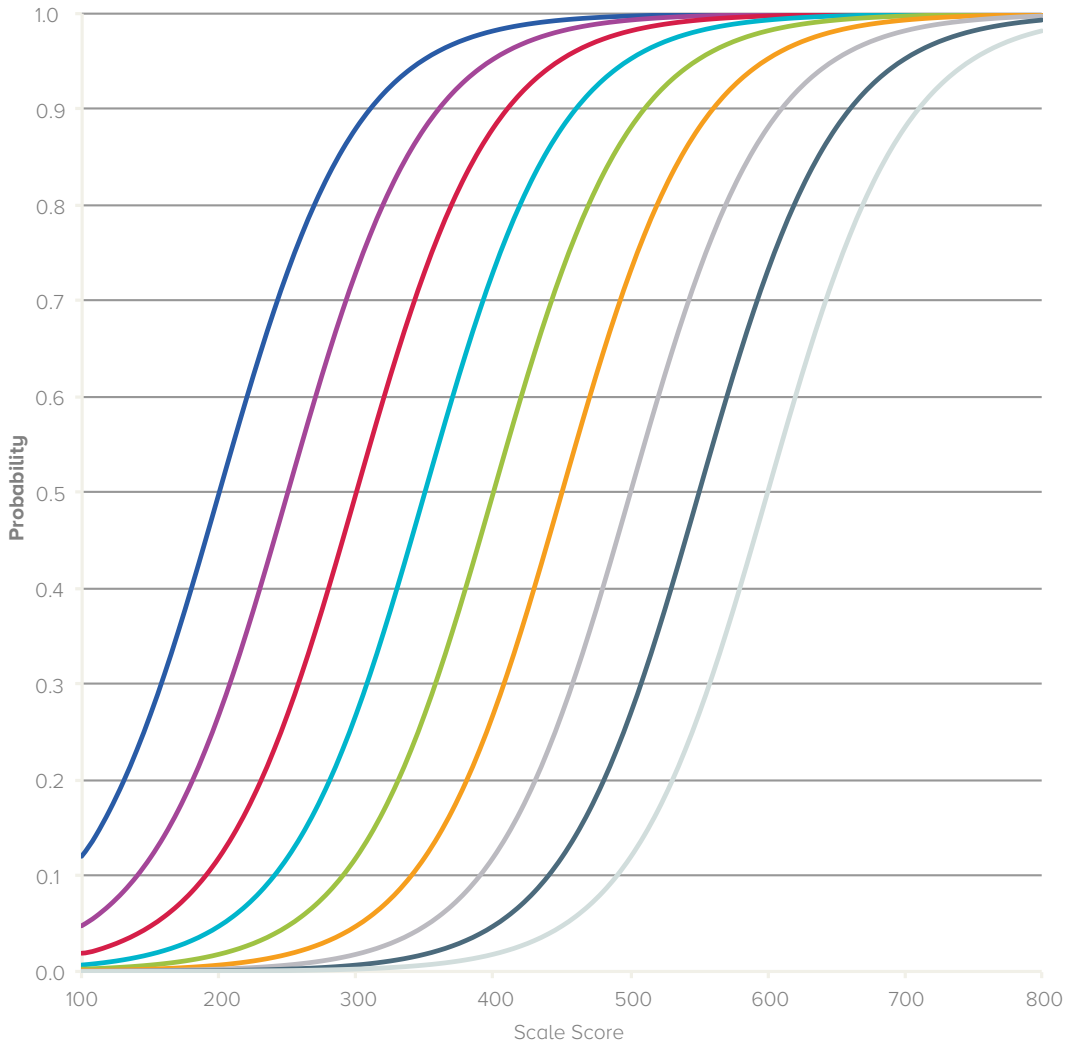


Interactive tools such as calculator, compass, and protractor are available to help students solve items.

APPENDIX II

A Deeper Dive into How i-Ready Diagnostic Works

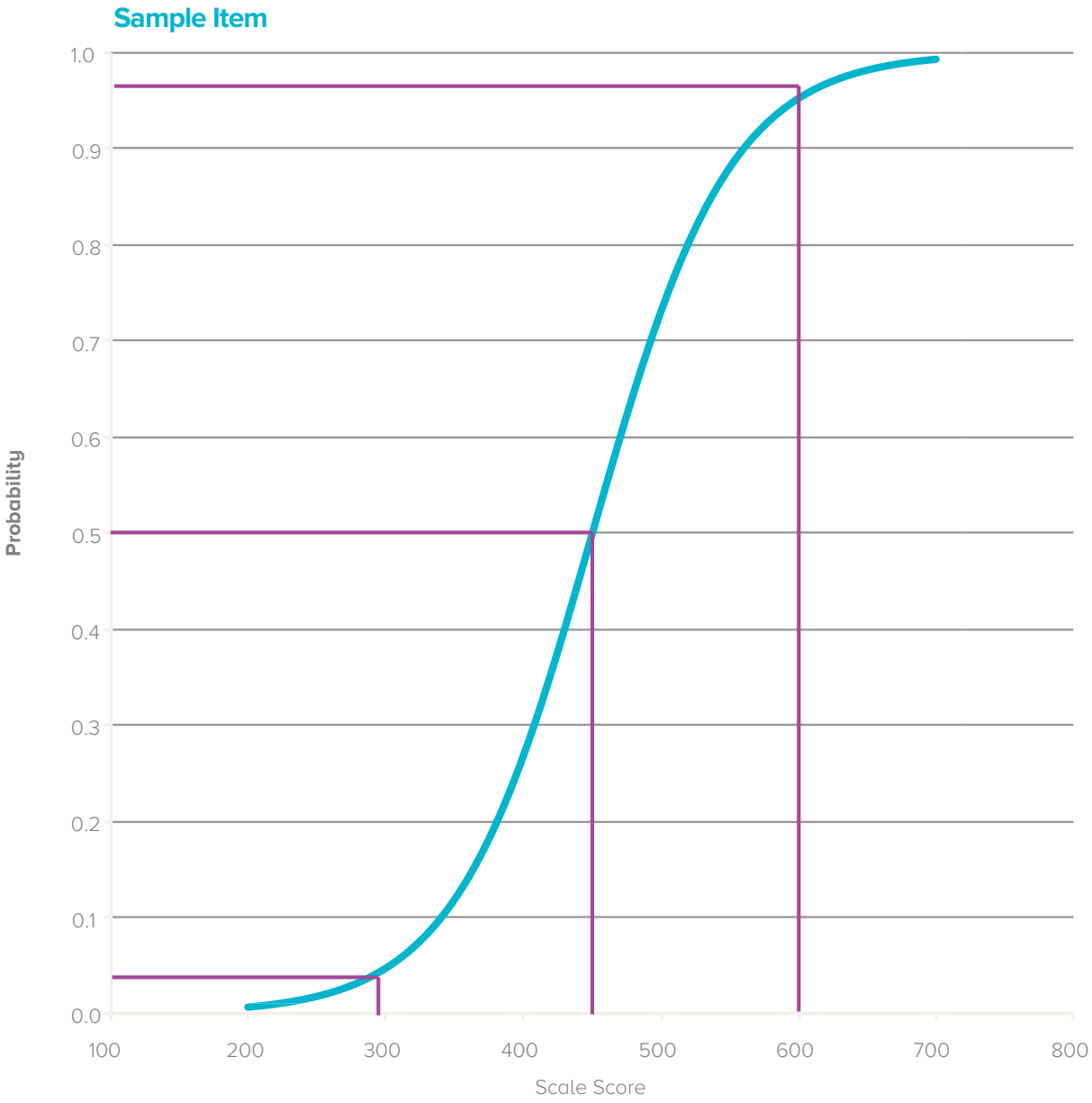
As previously mentioned, Item Response Theory (IRT) was employed as the adaptive theoretical foundation for i-Ready Diagnostic, and i-Ready specifically uses a probabilistic model known as the Rasch Model. Item characteristic curves provide information linking the probability of success on an item with the ability level of the student. In Rasch modeling, all item curves have a common slope. The location of the item curves differ, however, based on the difficulty of the items. The following figure provides an illustration of sample item characteristic curves for i-Ready Diagnostic.



The scale score on the horizontal axis represents the student's estimated ability level. Each curve represents an item of different difficulty, which is also calculated on the same scale as the student's ability level. The y-axis represents the probability of success students will have with the items. If the student's ability level matches the item's difficulty level, then the student will have a 50% probability of answering the item correctly—this is when an item has the highest level of differentiation of students' skill levels. This kind of matching allows for the most information to be derived from each student.

APPENDIX II (continued)

The following item characteristic curve is for a hypothetical two-digit multiplication item that has a difficulty level of 450.

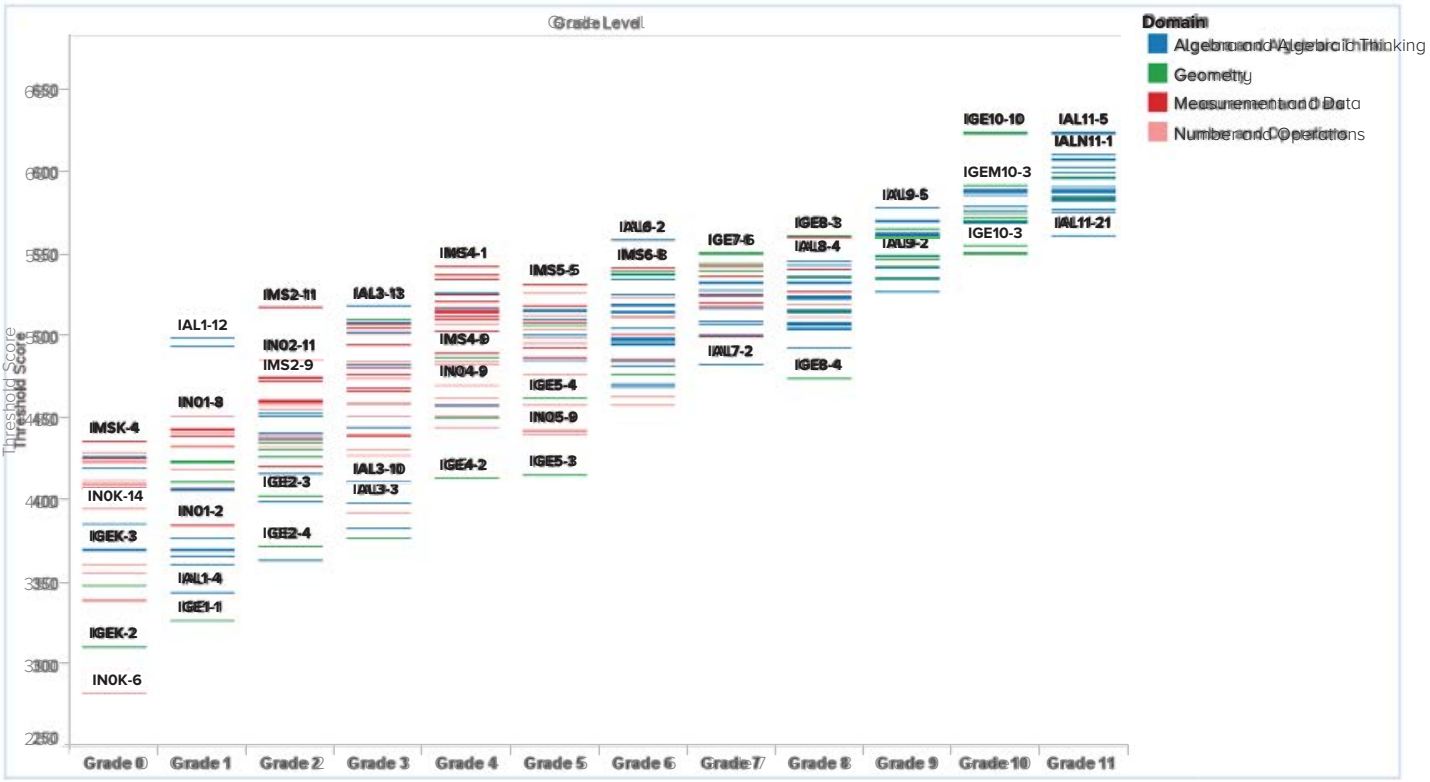


Note: If a student with an ability level of 300 were to get this item, then the student would have less than a 10% probability of answering it correctly. This would be like the student who has not mastered simple addition receiving a two-digit multiplication problem. We know the student struggles with simple addition, so we can deduce, probabilistically, that the student likely won't be able to answer the two-digit multiplication item correctly. Therefore, giving this item to that student will not provide that much information. On the other end of the spectrum, if we give this item to a student who has an ability level of 600, that would be similar to giving a student who we know has mastered long division an item that measures whether the student can multiply. Probabilistically, we can conclude that if the student has mastered long division, the student has also very likely mastered two-digit multiplication.

The Diagnostic hones in on the ability level of the student, and based on the student's scale scores, we can determine the probability of that student's success with other skills. The assessment measures proficiency on i-Ready Diagnostic indicators, which are based on the Common Core State Standards. The difficulty of an indicator is determined by the items associated with it. Hence, a look at the coverage and progression of ability measured by indicators across grade levels provides a good overview of the range of ability i-Ready measures.

The information regarding the difficulty of the indicators is modeled by a tabular Wright Map, which shows the difficulty level of the skills on the vertical axis and the grades on the horizontal axis. There is overlap from grade to grade, which is to be expected, as some skills in Grade 4, such as long division, may be more difficult than some of the simpler skills in Grade 5, such as multiplying by powers of 10.

The tabular Wright Map shows i-Ready indicators for Mathematics



In summary, as evidenced by the tabular Wright Map, i-Ready Diagnostic is built on a large, solid bank of items, well-structured and defined to assess the levels of skills across K–12. The adaptive algorithm selects items dynamically based on each student's ability level—only items providing the most information about the student are presented; below- or above-grade level items are available when the student is performing off level from his/her own grade. i-Ready provides an accurate and tailored testing experience for each student and, as a result, offers meaningful instructional information for educators. Refer to the i-Ready Technical Manual for additional information about i-Ready's technical properties.

APPENDIX III

Lexile® Linking Study

In the spring of 2012, MetaMetrics and Curriculum Associates partnered to conduct a Lexile Linking Study against i-Ready Diagnostic Reading. The purpose of this study was threefold:

1. Gather evidence of external (concurrent) validity.
2. Allow reporting of Lexile scores based on given i-Ready Diagnostic Reading scores.
3. Provide administrators, teachers, and parents information on appropriate reading materials for students.

MetaMetrics constructed Lexile Linking Tests for this study. A national sample of 3,280 students at grades 1, 3, 5, and 7, from 35 schools in 27 districts across 10 states completed both the Lexile Linking Test and i-Ready Diagnostic within 0–95 days with 97% of the students completing both tests within one month. About 60% of the students took the i-Ready Reading Diagnostic first and 40% took the Lexile Linking Test first. Table 1 shows the N counts of the target sample and the final sample by grade.

Table 1. Lexile Linking Study – Assessments Administered and Final Linking Sample by Grade

	1	3	5	7	Total
i-Ready Reading Diagnostic	1406	1724	1285	826	5241
Lexile Linking Test	1437	1781	1381	1038	5637
Final Linking Sample	840	1091	814	535	3280

The correlations between the i-Ready Reading Diagnostic Overall Score and Lexile Linking Text Lexile measure range from .88 to .89 across the four grades. These correlations support strong external validity with the Lexile measure. Linear linking models were created and incorporated into the i-Ready system to provide the i-Ready Reading Lexile measure along with links to appropriate reading materials given an i-Ready Reading Diagnostic score.

Quantile® Linking Study

A linking study between i-Ready Diagnostic Mathematics and the Quantile Measure was conducted in the spring of 2013 with similar purposes as the Lexile linking study referenced above:

1. Provide external validity information on i-Ready Diagnostic Mathematics.
2. Provide teachers linkage to resources from the Quantile Framework.

Detailed information about the Lexile and Quantile linking studies and validation of the linking results are available upon request.

Norm Research

During the third full-year implementation of i-Ready Diagnostic, the program served over a million students across the United States. From this large pool, Curriculum Associates collected i-Ready data from a nationally representative sample of students to create i-Ready's own National Norms. A technical report, available upon request, provides information about the development of the National Norms for grades K–8 during the spring of 2014.

For more information:
800-225-0248
www.i-Ready.com/Tour

To listen to a live webinar recording:
www.i-Ready.com/AdaptiveWebinar

NOTES:

NOTES:

NOTES:

9/14 10K
15917.0

Curriculum Associates

Highly rated for progress monitoring by the National Center on Intensive Intervention!

THE RESEARCH FOUNDATION FOR

STAR Assessments™

The Science of STAR

RENAISSANCE LEARNING

Accelerati learni ll

National Center on
INTENSIVE INTERVENTION
at American Institutes for Research

STAR Early Literacy™, **STAR Math™**, and **STAR Reading™** are highly rated for progress monitoring by the National Center on Intensive Intervention.

STAR Early Literacy™ is highly rated for screening and progress monitoring by the National Center on Response to Intervention.

STAR Reading™ and **STAR Math™** received the highest possible ratings for screening and progress monitoring from the National Center on Response to Intervention, with perfect scores in all categories.

Reports are regularly reviewed and may vary from those shown as enhancements are made.

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

1/14

Contents

Figures

Tables

Dear Educator,

Renaissance Learning is the world's leading provider of computer-based assessment technology, with products in use worldwide in grades pre-K–12. Renaissance Learning tools have a research base unmatched by makers of other educational products and have met the highest review standards set by reputable organizations such as the National Center on Intensive Intervention, the National Center on Response to Intervention, National Center on Student Progress Monitoring, the National Dropout Prevention Center, the Promising Practices Network, and the What Works Clearinghouse.

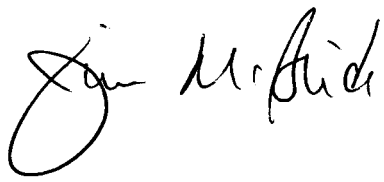
All Renaissance Learning tools are designed to accomplish our mission—"accelerating learning for all." A key educational principle supporting this mission is the notion that "the initial step in accelerating learning is to measure its occurrence." Our assessments—STAR Early Literacy Enterprise, STAR Reading Enterprise, and STAR Math Enterprise—do just that.

There is a reason approximately 18,000 schools worldwide use at least one STAR Enterprise assessment. They quickly gain favor with educators because of their ease of use, quick administration times, and ability to provide teachers with highly valid and reliable data upon completion of each test. The computer-based STAR assessment system is a multipurpose tool. STAR is used for screening and progress monitoring, and also includes resources that target instruction for all kinds of learners. Students who are most at risk can be identified quickly. No time is wasted in diagnosing their needs, allowing intervention to begin immediately.

Read on to learn more about STAR Enterprise assessments. I'm confident you'll see rather quickly why teachers using STAR Enterprise accelerate learning, get more satisfaction from teaching, and help their students achieve higher scores on state and national tests. The stakes are high. We must help all students in all schools be prepared for college or careers by the time they graduate from high school.

For additional information, full technical manuals are available for each STAR assessment by contacting Renaissance Learning at research@renlearn.com

Sincerely,



James R. McBride, Ph.D.
Vice President & Chief Psychometrician
Renaissance Learning, Inc.

James R. McBride, Ph.D., is vice president and chief psychometrician for Renaissance Learning. He was a leader of the pioneering work related to computerized adaptive testing (CAT) conducted by the Department of Defense. McBride has been instrumental in the practical application of item response theory (IRT) and since 1976 has conducted test development and personnel research for a variety of organizations. At Renaissance Learning, he has contributed to the psychometric research and development of STAR Math, STAR Reading, and STAR Early Literacy. McBride is co-editor of a leading book on the development of CAT and has authored numerous journal articles, professional papers, book chapters, and technical reports.

Introduction

STAR Enterprise assessments are designed to help teachers assess students quickly, accurately, and efficiently. STAR provides teachers with reliable and valid data instantly so that they can target instruction, monitor progress, provide students with the most appropriate instructional materials, and intervene with at-risk students. Administrators use real-time data from STAR to make decisions about curriculum, assessment, and instruction at the classroom, school, and district levels.

Three STAR Enterprise assessments measure student achievement in four areas:

- STAR Early Literacy Enterprise assesses early literacy and early numeracy skills (grades pre-K–3)
- STAR Reading Enterprise assesses reading skills (grades K–12)
- STAR Math Enterprise assesses math skills (grades K–12)

All STAR Enterprise assessments include skills-based test items, the Core Progress learning progressions for instructional planning, and in-depth reports. Operating on the Renaissance Place hosted platform, STAR Enterprise is a comprehensive assessment system for data-driven schools. The assessments provide accurate data in a short amount of time by combining computer-adaptive technology with a specialized psychometric test design that utilizes item response theory (IRT).

Students take STAR Enterprise assessments on individual computers or iPads. The software delivers multiple-choice items one by one, and a student selects answers with a mouse, keyboard, or touchscreen. After an assessment is completed, the software calculates the student's score. Teachers and administrators then select reports to provide results for an individual student, class, grade, school, or district.

STAR Assessments have been favorably reviewed as reliable, valid, and efficient by various independent groups, including the National Center on Intensive Intervention, the National Center on Response to Intervention, and the National Center on Student Progress Monitoring. STAR also has a significant research base as shown in Table 1.

Table 1: Research Support for STAR Assessments™

Assessment	Total Research Publications	Independent Research Publications
STAR Early Literacy	21	14
STAR Reading	76	22
STAR Math	65	21

STAR Assessments™ Overview

STAR Early Literacy Enterprise™ Assessment

The importance of assessing skills early in a child's schooling cannot be overstated. Research supports successful early intervention as the best single predictor for future academic success, particularly in the critical areas of reading and language acquisition.

Students are expected to develop a variety of early literacy as they progress from pre-kindergarten through third grade on their way to becoming readers. This progression reflects both the home literacy environment and educational interventions. The development of these skills, however, is not continuously upward. Students typically learn a skill, forget it, and then relearn it. Many well-established tests assess a student at a particular point in time. STAR Early Literacy Enterprise is designed to repeatedly assess a child's status at different stages throughout this important growth period.

STAR Early Literacy Enterprise measures early literacy and early numeracy skills throughout the early primary grades (pre-K–3).[1] Information from the assessment enables teachers to intervene immediately at the beginning of a student's formal learning process. This is particularly critical for students who enter school already lacking in experiences or the foundational skills necessary for early literacy and early numeracy development to take root.

STAR Early Literacy Enterprise is a standards-based test that measures student performance in key early literacy and early numeracy skills, providing valuable information regarding the acquisition of ability along a continuum of expectations. Table 2 breaks down the STAR Early Literacy Enterprise item bank by overall size, number of items administered per testing event, and average administration time.

Table 2: Summary of STAR Early Literacy Enterprise™ Item Bank Size and Administration Details

	STAR Early Literacy Enterprise
Item Bank Size	More than 2,500 items
Items Administered per testing event	27 items
Average Administration Time	About 10 minutes

For teachers, STAR Early Literacy Enterprise provides a simple way to monitor progress based on the specific needs of each student. It is especially helpful in identifying students who may be at risk for later reading failure. Data from the assessment is used for goal setting and outcome assessment as well as for planning instruction and intervention. A student's scaled score from STAR Early Literacy is also mapped to the empirically validated Core Progress learning progression. This score represents an entry point onto Core Progress, and using this tool, teachers can clearly see the skills students have likely mastered and the ones they are ready to develop next (for more about Core Progress, see pp. 10 and 31). Although STAR Early Literacy Enterprise is designed for students in grades pre-K–3, it can be used with older students, such as struggling readers, nonreaders, special education students, or English learners.

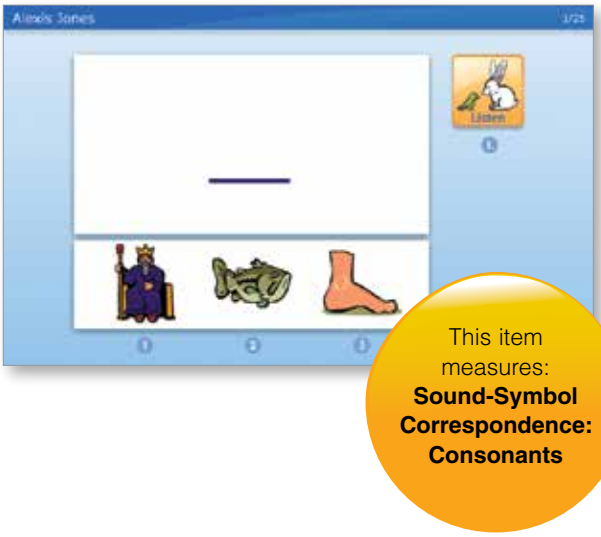
[1] STAR Early Literacy Enterprise is specifically designed for students who do not yet read. Students who have established a 100-sight-word vocabulary, or have reached the Probable Reader stage of literacy development in STAR Early Literacy Enterprise, typically are ready to take a STAR Reading Enterprise assessment.

STAR Early Literacy Enterprise is distinguished from other assessments of early literacy in three ways:

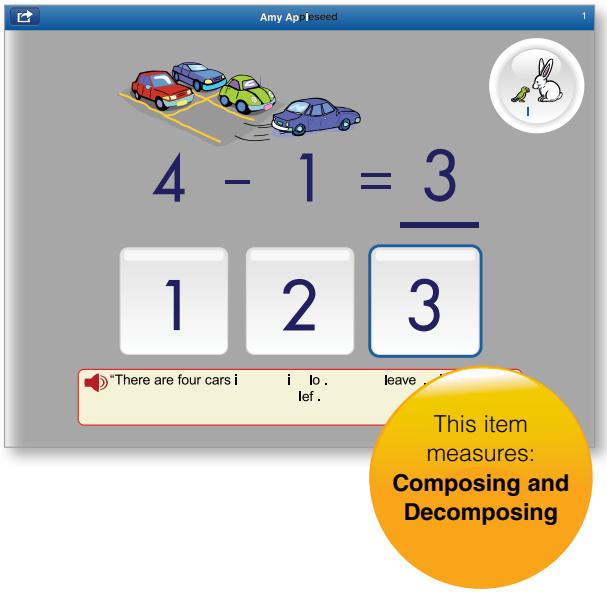
1. It is computer-administered, using graphics, audio instructions, and automatic dictation of instructions and test questions, so that most children can take the test without teacher assistance.
2. It is computer adaptive, which means the content and difficulty level of each test administration is tailored to each student's performance.
3. It is brief, administering 27 items (including five early numeracy items) in about 10 minutes. Despite its brevity, the assessment correlates highly with a wide range of more time-intensive standardized measures of early literacy, reading, and other readiness skills. Figure 1 shows sample assessment items.

Figure 1: STAR Early Literacy Enterprise™ Sample Assessment Items

Early Literacy Item



Early Numeracy Item



STAR Reading Enterprise™ Assessment

STAR Reading Enterprise is a challenging, interactive, and brief (about 15 minutes) assessment, consisting of 34 questions per test, that evaluates a breadth of reading skills appropriate for grades K–12.[2] The assessment's repeatability and flexibility in administration provide specific advantages for everyone responsible for the education of students:

- Teachers use results from STAR Reading Enterprise to facilitate individualized instruction and identify students who most need remediation or enrichment.
- Principals access assessment information through browser-based management and regular, accurate reports on performance at the individual, class, building, and district level.
- Administrators and assessment specialists apply reliable and timely information on reading growth at each school and districtwide, which serves as a valid basis for comparing data across schools, grades, and special student populations.

[2] Although STAR Reading Enterprise is normed for grades 1–12, kindergarten students may take the assessment with teacher discretion. Students with a 100-sight-word vocabulary, or who have reached the Probable Reader stage of literacy development in STAR Early Literacy Enterprise, are typically ready to take the assessment.

STAR Reading Enterprise is a standards-based test that measures student performance in key reading skills, providing valuable information regarding the acquisition of reading ability along a continuum of literary expectations. Table 3 breaks down the STAR Reading Enterprise item bank by overall size, number and types of items administered per testing event, and average administration time.

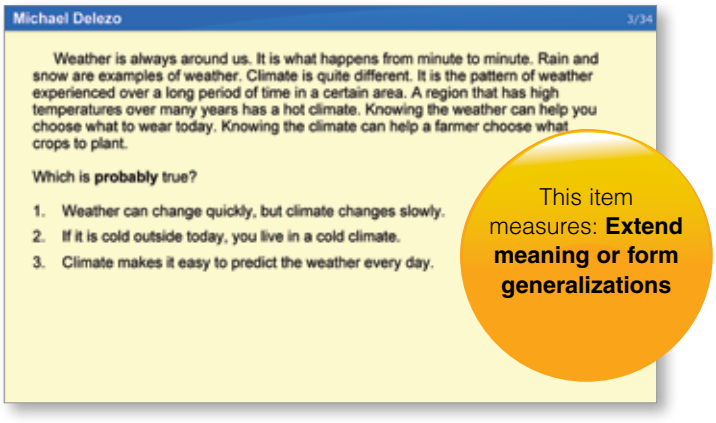
Table 3: Summary of STAR Reading Enterprise™ Item Bank Size and Administration Details

	STAR Reading Enterprise
Item Bank Size	More than 5,000
Items Administered per testing event	34 items
Average Administration Time	About 15 minutes

Renaissance Learning has conducted extensive research and consulted heavily with reading and assessment experts to arrive at the skills most appropriate for assessing reading development. Several publications have been studied, including the 2010 Common Core State Standards; the Reading Framework for the 2009 National Assessment of Educational Progress; the National Council of Teachers of English (2006) *Principles of Adolescent Literacy Reform* policy brief; and the Alliance for Excellent Education's (2004) *Reading Next* report. External advisors include Margaret Heritage, Ph.D., National Center for Research on Evaluation, Standards, and Student Testing at UCLA; Karin Hess, Ed.D., Center for Assessment (NCIEA); Thomas P. Hogan, Ph.D., University of Scranton; James Milgram, Ph.D., Stanford University; Michael Milone, Ph.D., research psychologist; R. Sharif M. Shakrani, Ph.D., private consultant; Amanda M. VanDerHeyden, Ph.D., private consultant; and James Ysseldyke, Ph.D., University of Minnesota.

Students with a 100-sight-word vocabulary, or who have reached the Probable Reader stage of literacy development in STAR Early Literacy Enterprise, are typically ready to take a STAR Reading Enterprise assessment. STAR Reading Enterprise serves three purposes of particular interest to school and district administrators: (1) to give teachers quick and accurate estimates of students' reading achievement levels, (2) to assess reading achievement relative to national norms, and (3) to provide a means for monitoring growth in a consistent manner longitudinally for all students. Figure 2 shows a sample assessment item.

Figure 2: STAR Reading Enterprise™ Sample Assessment Item



Teachers who use STAR Reading Enterprise can monitor progress toward college- and career-ready standards, such as the Common Core State Standards, as well as predict proficiency on state tests. After a STAR Enterprise assessment is taken, the software uses the resulting scaled score to locate the student's entry point onto the Core Progress learning progression, helping educators learn more about how the student is performing relative to grade-level expectations. Core Progress provides a road map of skills, spanning from emergent reading to the level of competence required for college and careers, displaying both prerequisite skills students have typically mastered and skills they are ready to develop next.

The learning progression, however, is not a straight trajectory. Because students develop at different rates and in different ways, STAR software includes additional resources for targeted instruction, intervention, and enrichment, including Worked Examples, Skill Probes, and Performance Tasks. Additional content will be continuously developed as a means to probe more deeply into students' understandings and skills development (for more about Core Progress, see pp. 10 and 31).

STAR Math Enterprise™ Assessment

STAR Math Enterprise is a challenging, interactive, and brief (about 20 minutes) assessment, consisting of 34 items per test, that evaluates students' mathematical abilities in grades K–12.[3] Like STAR Reading Enterprise, its repeatability and flexibility in administration provide specific advantages for educators:

- Teachers use results from STAR Math Enterprise to facilitate individualized instruction and identify students who most need remediation or enrichment.
- Principals access assessment information through browser-based management and regular, accurate reports on performance at the individual, class, building, and district level.
- Administrators and assessment specialists apply reliable and timely information on mathematical growth at each school and districtwide, which serves as a valid basis for comparing data across schools, grades, and special student populations.

STAR Math Enterprise is a skills-based assessment of math achievement. Table 4 breaks down the STAR Math Enterprise item bank by overall size, number of items administered per testing event, and average administration time.

Table 4: Summary of STAR Math Enterprise™ Item Bank Size and Administration Details

	STAR Math Enterprise
Item Bank Size	More than 5,000
Items Administered per testing event	34 items
Average Administration Time	About 20 minutes

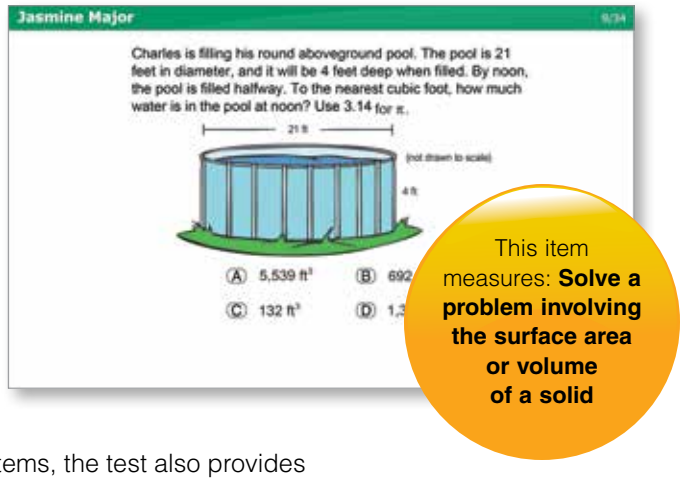
STAR Math Enterprise provides a reliable and valid method for measuring progress towards achievable goals in mathematics. Teachers, principals, literacy coaches, assessment directors, and district-level administrators can use the assessment data for instructional planning, growth measurement, and program evaluation. At an individual student level, STAR can be used for a variety of purposes, including screening, formative assessment, progress monitoring, calculating growth, and outcomes assessment. By using the assessment on a regular basis, such as quarterly or monthly, teachers can monitor progress and make appropriate adjustments to instruction. Research firmly supports progress monitoring, which has shown to be successful in a variety of educational settings.

[3] Although STAR Math Enterprise is normed for grades 1–12, kindergarten students may take the assessment with teacher discretion.

As with STAR Reading Enterprise, teachers who use STAR Math Enterprise can monitor progress toward college- and career-ready standards, such as those found in the Common Core State Standards, as well as predict proficiency on state tests. After a STAR Enterprise assessment is taken, the software uses the resulting scaled score to locate the student's entry point onto the Core Progress learning progression, helping educators learn more about how the student is performing relative to grade-level expectations. Core Progress provides a road map of skills, spanning from early numeracy to the level of competence required for college and careers, displaying both prerequisite skills students have typically mastered and skills they are ready to develop next.

The learning progression, however, is not a straight trajectory. Because students develop at different rates and in different ways, the software includes additional resources for targeted instruction, intervention, and enrichment, including Worked Examples, Skill Probes, Performance Tasks, and links to third-party educational resources. Additional content will be continuously developed as a means to probe more deeply into students' understandings and skills development (for more about Core Progress, see pp. 10 and 31).

Figure 3: STAR Math Enterprise™ Sample Assessment Item



Students taking a STAR Math Enterprise assessment follow a protocol in which they use blank work paper and pencils during the test administration. As warranted for specific assessment items, the test also provides an onscreen calculator and/or reference sheet. Figure 3 shows a sample assessment item.

Test Design

Computer adaptive testing (CAT)

STAR Enterprise assessments are computer adaptive tests (CATs). CATs continually adjust the difficulty of each student's test by selecting each assessment item based on the student's previous performance. CATs shorten testing time as well as spare students both the frustration of items that are too difficult and the boredom of items that are too easy.

> "STAR measures offer an important and potentially valuable contribution to RTI."
>
> Shapiro, 2012, p. 20

Decades of research have shown that CATs can be considerably more efficient than conventional tests, which present all students with the same test questions (e.g., Lord, 1980; McBride & Martin, 1983). A well-designed CAT is often two or more times as efficient as a conventional test. For example, to equal the reliability of a 50-item conventional test, a well-designed CAT may use only 25 items to yield the same information in half the time. As noted by Weiss (2004), "Early evidence of improved measurement precision (reliability) and validity (e.g., Johnson & Weiss, 1980; Kingsbury & Weiss, 1980) and large reductions in the number of items administered (typically 50% or more) without having an impact on the psychometric characteristics of test scores for CAT have been confirmed in a number of recent studies (e.g., Mardberg & Carlstedt, 1998; Moreno & Segall, 1997)" (pp. 77–78).

A new line of research suggests that CATs are a sound choice for progress monitoring student performance in response to intervention (RTI) settings. "RTI is a process of providing high quality interventions that are matched to student need, and uses frequent progress monitoring of student response to interventions to assist in making important educational decisions" (Bray & Kehle, 2011, p. 616). Progress monitoring feedback is key to RTI as it tells educators which interventions are helping students most. Thus, "progress-monitoring measures must be frequent, sensitive to instructional change over a short period of time, predictive of overall success as measured by the benchmark assessment, and able to drive instructional decisions" (Shapiro, 2012, p. 9). "STAR measures offer an important and potentially valuable contribution to RTI" (Shapiro, p. 20) in the following ways:

- Frequency of administration—STAR Aassessments were designed to provide educators with flexibility in administering the assessments at the frequency most fitting their needs, whether it be three times per school year for screening, monthly to better understand how student progress is unfolding during the school year with enough time to change the growth trajectory, or as often as weekly for progress monitoring students in tiers 2 and 3 of an RTI framework.

- Sensitivity—STAR Assessments meet all criteria set by the National Center on Intensive Intervention (NCII) (2012a, 2012b, 2012c) and the National Center on Response to Intervention (NCRTI) (2010a, 2010b, 2010c) for "Sensitivity to Student Improvement."

- Predictive power—STAR Assessments meet all criteria set by the NCII and the NCRTI for "Predictive Validity of the Slope of Improvement," as well as criteria set by the NCRTI for "Classification Accuracy." In addition, a (2012) study found that STAR Math "was the single best predictor of PSSA scores across grades" (Shapiro & Gebhardt, p. 303) when compared to CBM measures. For additional predictive validity evidence for each STAR, see Psychometric Properties, p. 19.

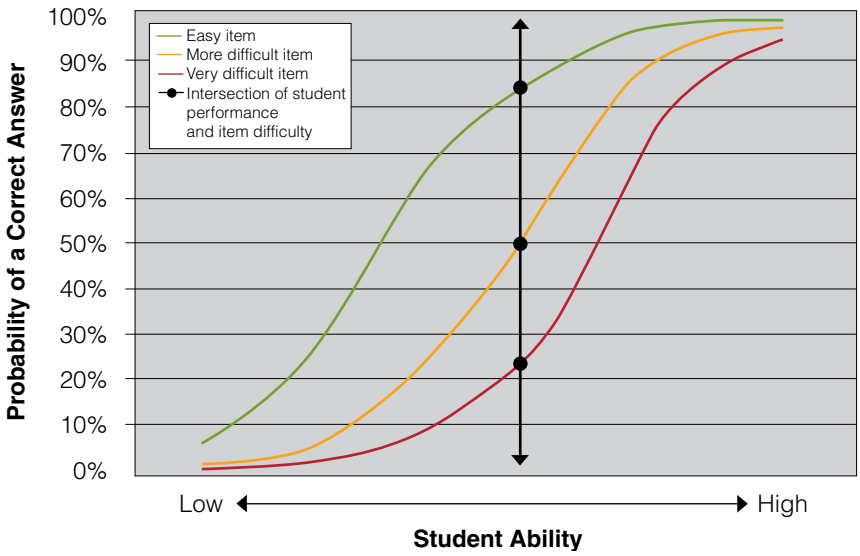
- Impact on instructional decisions—STAR Assessments meet all criteria set by NCII and NCRTI for both "Decision Rules for Changing Instruction" and "Decision Rules for Increasing Goals." Core Progress learning progressions—which place students scores within a progression of learning—make the data from STAR Assessments immediately actionable and facilitate instructional planning (for more information, see Instructional planning with Core Progress, p. 31).

Item response theory and its role in CAT

Tailoring item difficulty to match a student's knowledge or skill level can be done in a number of different ways; however, most CATs use item response theory (IRT) as the basis for both adaptive item selection and test scoring. IRT puts student performance and item difficulty on the same scale and offers a means to estimate the probability that a student will answer a given test item correctly. IRT models provide a way to measure each item's degree of difficulty and to estimate each student's achievement level from the pattern of correct and incorrect responses to items.

With item response theory, scientists can calculate the probability of a correct response to an item as a function of student ability. As student ability increases, so does the probability the student will answer correctly. Additionally, because some test items are harder than others, the probability trend differs from one item to another. Figure 4 shows the probability functions for three test items: one that's easy, one that's moderately difficult, and one that's very difficult.

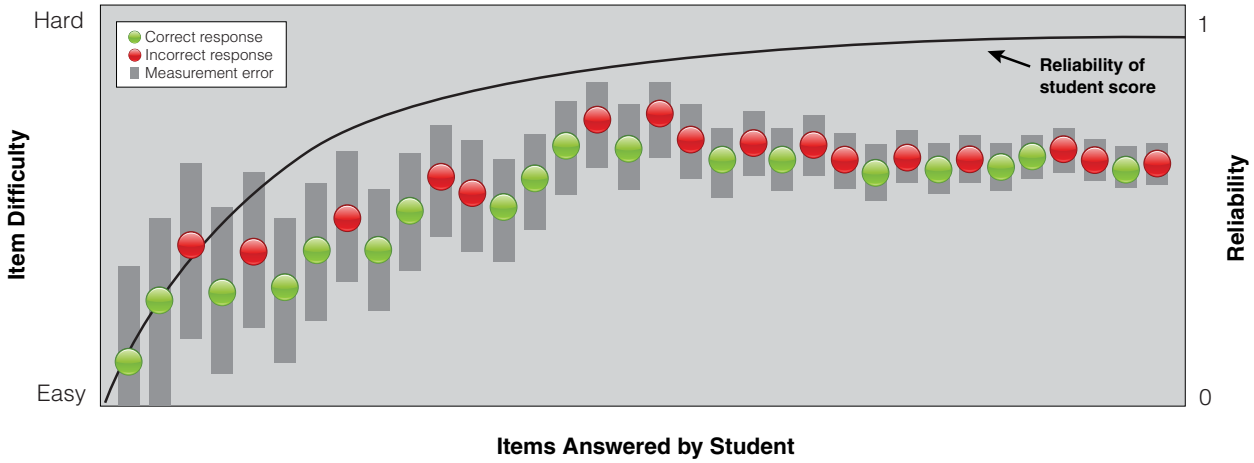
Figure 4: Illustration of a Student's Reactions to Three Test Items of Varying Difficulty



During a STAR Enterprise assessment administration, the software automatically moves up or down the item scale to select questions based on a student's answers. If the student answers a question correctly, the next question will be more difficult. If the student answers incorrectly, the next question will be less difficult. Unlike manual paper-and-pencil assessments, STAR Enterprise assessments dynamically adjust to each student's unique responses. As a result, STAR pinpoints student achievement levels quickly and efficiently.

Figure 5 displays an example progression of less difficult and more challenging items based on a student's previous item responses during a CAT administration. It also shows how selecting items tailored to a student's ability helps to reduce measurement error as the test progresses.

Figure 5: How Computer-Adaptive Technology Works



Core Progress™ Learning Progressions—The Bridge Between Assessment and Instruction

A learning progression is a continuum of expected learning, beginning with emergent reading or early numeracy skills and progressing to the level of competence required for college and careers. The skills are interconnected and related, formed from requisites and prerequisites, and represent how students typically advance their learning in a subject area. According to Heritage (2008), "Learning progressions that clearly articulate a progression of learning in a domain can provide the big picture of what is to be learned, support instructional planning, and act as a touchstone for formative assessment" (p. 1).

In July 2013, Renaissance Learning released two new learning progressions built specifically for the Common Core.

Skills in a learning progression are not meant to be taught sequentially; rather, a student's placement on a learning progression begins with a student's score from a standardized test of achievement. This information helps orient student and teacher to where the student has been, where the student is headed, and the skills with which they may need guidance in order to arrive at their destination successfully.

Evolution of Core Progress™

To build a bridge between assessment and instruction, Renaissance Learning created the Core Progress for Reading and Core Progress for Math learning progressions. Members of the Renaissance Learning standards team rigorously developed, tested, and validated Core Progress. For both reading and math, standards experts identified the initial order of item difficulty by researching reading and math theory, examining widely accepted frameworks such as state standards, reviewing the Common Core State Standards (CCSS), and consulting nationally recognized reading and math experts.

The road map of skills in Core Progress helps teachers monitor progress toward college- and career-ready standards. Using a student's STAR scaled score, Core Progress displays student progress in skills relative to grade-level expectations.

All students follow individual paths to achieve personalized goals. Because students develop reading and math ability at different rates and in different ways, a student's progression through Core Progress does not follow a straight trajectory. Additional resources, such as Worked Examples, Skill Probes, Performance Tasks, and links to third-party educational resources, help teachers meet students at their individual achievement levels for targeted instruction, intervention, and enrichment.

Built for the Common Core State Standards

As the majority of states implemented the Common Core State Standards, Renaissance Learning recognized a need for learning progressions created expressly for these new standards. In July 2013, Renaissance Learning released two new learning progressions built specifically for the CCSS:

- **Core Progress™ Learning Progression for Reading—Built for the Common Core State Standards**
- **Core Progress™ Learning Progression for Math—Built for the Common Core State Standards**

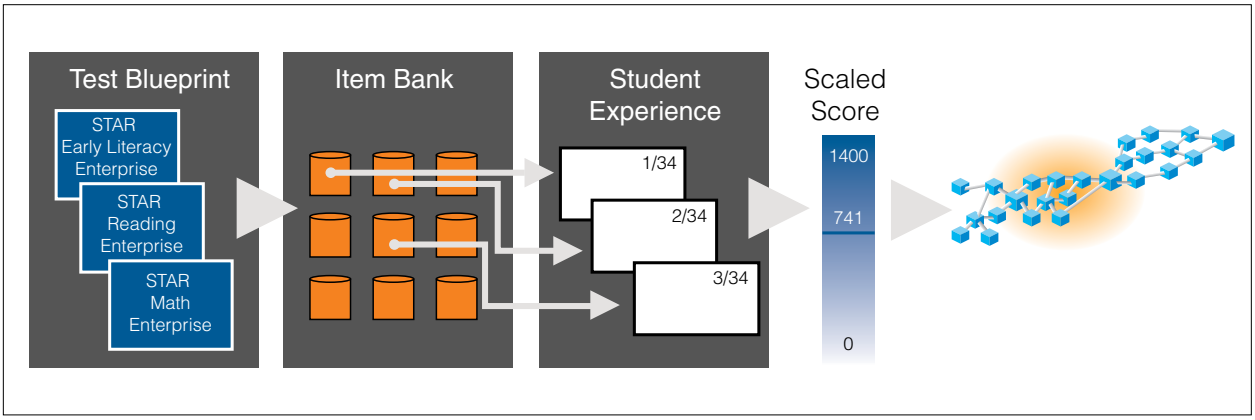
Like the original Core Progress, the new CCSS-specific learning progressions present a continuum of skills from emergent reading and early numeracy through the level of knowledge required for college and careers, as well as display both prerequisite skills students have mastered and skills they are ready to develop next. The new learning progressions are different in that they were built, from the ground up, specifically for the Common Core State Standards.

Renaissance Learning standards experts began this process with a close analysis of the CCSS, identifying each standard's inherent skills, intent, and key terminology. They also immersed themselves in the literature and resources available regarding the CCSS to determine how the standards were being interpreted and implemented by states and relevant consortia. All of this ensured that the new learning progressions included incremental steps of learning to fulfill the intent of the standards and ultimately culminate in college and career readiness.

Path from test blueprint to learning progression

Empirical testing has found a strong statistical link between the progression of skills in Core Progress and the assessed difficulty level of STAR Enterprise test items, meaning educators can use scores from the assessments to identify both what a student knows and what they need to work on. As Figure 6 shows, a STAR assessment's blueprint working in tandem with CAT technology ultimately dictates which items are presented to each student. While each STAR test event is unique, the blueprint ensures that a certain number of items from the domains and skill sets are presented to each student.

Figure 6: How it Works: From STAR™ Test Blueprint to Core Progress™ Learning Progression



Depending on the state in which you reside, you will either have access to the original Core Progress learning progression or the Core Progress Learning Progression—Built for the Common Core State Standards.

After a student takes a STAR Enterprise assessment, the software uses the resulting scaled score to find the student's entry point onto the Core Progress learning progression and then reports the skills the student has likely mastered in prior grades and those the student is ready to develop next, helping teachers to focus instruction. For more information about how Core Progress helps tailor student instruction, see Instructional planning with Core Progress, p. 31.[4]

[4] For more in-depth information, please see:
Core Progress for Reading: Empirically Validated Learning Progressions (http://doc.renlearn.com/KMNet/R0053985FA6D567F.pdf)
Core Progress for Math: Empirically Validated Learning Progressions (http://doc.renlearn.com/KMNet/R00552482161352C.pdf)

Skills in Core Progress™ Learning Progression–Built for the Common Core State Standards
The order of skills presented in the new learning progressions built for the CCSS emerged from Renaissance Learning content experts' deep study of the standards.

Figure 7 displays the organization of the domains and skill areas in the learning progression for early literacy.

Figure 7: Core Progress™ Learning Progression for Reading—Built for the Common Core State Standards: Domains and Skill Areas (Early Literacy)

Foundational Skills

Print Concepts
- Directionality
- Letters and Words
- Word Length
- Word Borders
- Visual Discrimination / Alphabetic Principle
- Alphabetic Sequence
- Print Features

Phonological Awareness
- Rhyming and Word Families
- Blending, Counting, and Segmenting Syllables
- Blending and Segmenting
- Distinguishing between Long and Short Vowel Sounds
- Isolating Initial, Final, and Medial Phonemes
- Adding/Substituting Phonemes

Phonics and Word Recognition
- Spelling-Sound Correspondences: Consonants
- Spelling-Sound Correspondences: Vowels
- Regular and Irregular Spellings / High-Frequency words
- Inflectional Endings / Affixes
- Syllables

Fluency
- Purpose of Reading / Reading with Comprehension
- Reading Rate WCPM
- Prosody
- Repair Strategies

Language

Vocabulary Acquisition and Use
- Real-Life Word Connections and Applications
- Word Reference Materials
- Antonyms
- Synonyms
- Structural Analysis
- Word Relationships
- Context Clues
- Vocabulary in Context
- Multiple-Meaning Words
- Figures of Speech
- Connotation

As Figure 8 shows, for reading, the organization of the learning progression reflects the CCSS with four domains: (1) Foundational Skills, (2) Language, (3) Literature, and (4) Informational Text (which reflects the emphasis on nonfiction text in the standards).

Figure 8: Core Progress™ Learning Progression for Reading—Built for the Common Core State Standards: Domains and Skill Areas

Foundational Skills

Print Concepts
- Directionality
- Letters and Words
- Word Length
- Word Borders
- Visual Discrimination / Alphabetic Principle
- Alphabetic Sequence
- Print Features

Phonological Awareness
- Rhyming and Word Families
- Blending, Counting, and Segmenting Syllables
- Blending and Segmenting
- Distinguishing between Long and Short Vowel Sounds
- Isolating Initial, Final, and Medial Phonemes
- Adding/Substituting Phonemes

Phonics and Word Recognition
- Spelling-Sound Correspondences: Consonants
- Spelling-Sound Correspondences: Vowels
- Regular and Irregular Spellings / High-Frequency words
- Inflectional Endings / Affixes
- Syllables

Fluency
- Purpose of Reading / Reading with Comprehension
- Reading Rate WCPM
- Prosody
- Repair Strategies

Language

Vocabulary Acquisition and Use
- Real-Life Word Connections and Applications
- Word Reference Materials
- Antonyms
- Synonyms
- Structural Analysis
- Word Relationships
- Context Clues
- Vocabulary in Context
- Multiple-Meaning Words
- Figures of Speech
- Connotation

Literature

Key Ideas and Details
- Character
- Setting
- Plot
- Theme
- Summary
- Inference and Evidence

Craft and Structure
- Point of View
- Structure of Literary Text
- Word Meaning
- Author's Word Choice and Figurative Language
- Connotation

Integration of Knowledge and Ideas
- Modes of Representation
- Analysis and Comparison

Range of Reading and Level of Text Complexity
- Range of Reading
- Development of Independence

Informational Text

Key Ideas and Details
- Main Idea and Details
- Inference and Evidence
- Prediction
- Sequence
- Compare and Contrast
- Cause and Effect
- Summary
- Connections and Relationships

Craft and Structure
- Text Features
- Author's Purpose and Perspective
- Word Meaning
- Connotation
- Organization
- Author's Word Choice and Figurative Language

Integration of Knowledge and Ideas
- Modes of Representation
- Argumentation
- Analysis and Comparison

Range of Reading and Level of Text Complexity
- Range of Reading
- Development of Independence

In Figures 9 and 10, the organization of the learning progression for math is identical to the CCSS framework for grades K–8 and high school.

Figure 9: Core Progress™ Learning Progression for Math—Built for the Common Core State Standards: Domains and Skill Areas (K–8)

Counting and Cardinality

- Whole Numbers: Counting, Comparing, and Ordering

Operations and Algebraic Thinking

- Algebraic Thinking
- Evaluate Numerical Expressions
- Whole Numbers: Addition and Subtraction
- Whole Numbers: Counting, Comparing, and Ordering
- Whole Numbers: Multiplication and Division

Number and Operations in Base Ten

- Decimal Concepts and Operations
- Powers, Roots, and Radicals
- Whole Numbers: Addition and Subtraction
- Whole Numbers: Counting, Comparing, and Ordering
- Whole Numbers: Multiplication and Division
- Whole Numbers: Place Value

Number and Operations — Fractions

- Decimal Concepts and Operations
- Fraction Concepts and Operations

Ratios and Proportional Relationships

- Percents, Ratios, and Proportions

The Number System

- Coordinate Geometry
- Decimal Concepts and Operations
- Fraction Concepts and Operations
- Integers
- Whole Numbers: Multiplication and Division

Expressions and Equations

- Evaluate and Use Variable Expressions
- Evaluate Numerical Expressions
- Linear Equations and Inequalities
- Powers, Roots, and Radicals
- Quadratic and Nonlinear Equations and Inequalities
- Systems of Equations and Inequalities

Functions

- Relations and Functions

Geometry

- Angles, Segments, and Lines
- Congruence and Similarity
- Coordinate Geometry

- Fraction Concepts and Operations
- Geometry: Three-Dimensional Shapes and Attributes
- Geometry: Two-Dimensional Shapes and Attributes
- Perimeter, Circumference, and Area
- Right Triangles and Trigonometry
- Surface Area and Volume
- Transformations

Measurement and Data

- Angles, Segments, and Lines
- Data Representation and Analysis
- Geometry: Two-Dimensional Shapes and Attributes
- Measurement
- Money
- Perimeter, Circumference, and Area
- Surface Area and Volume
- Time
- Whole Numbers: Addition and Subtraction
- Whole Numbers: Counting, Comparing, and Ordering

Statistics and Probability

- Combinatorics and Probability
- Data Representation and Analysis

Figure 10: Core Progress™ Learning Progression for Math—Built for the Common Core State Standards: Domains and Skill Areas (High School)

The Real Number System
- Fraction Concepts and Operations
- Powers, Roots, and Radicals

Quantities
- Data Representation and Analysis

Seeing Structure in Expressions
- Algebra of Polynomials
- Linear Equations and Inequalities
- Quadratic and Nonlinear Equations and Inequalities
- Relations and Functions

Arithmetic with Polynomials and Rational Expressions
- Algebra of Polynomials

Creating Equations
- Linear Equations and Inequalities

Reasoning with Equations and Inequalities
- Linear Equations and Inequalities
- Quadratic and Nonlinear Equations and Inequalities
- Relations and Functions
- Systems of Equations and Inequalities

Interpreting Functions
- Relations and Functions

Building Functions
- Relations and Functions

Linear, Quadratic, and Exponential Models
- Linear Equations and Inequalities
- Quadratic and Nonlinear Equations and Inequalities

The Complex Number System
- Algebra of Polynomials
- Complex Numbers

Trigonometric Functions
- Right Triangles and Trigonometry

Congruence
- Angles, Segments, and Lines
- Congruence and Similarity
- Geometry: Two-Dimensional Shapes and Attributes
- Polygons and Circles
- Transformations

Similarity, Right Triangles, and Trigonometry
- Congruence and Similarity
- Right Triangles and Trigonometry
- Transformations

Circles
- Polygons and Circles

Expressing Geometric Properties with Equations
- Coordinate Geometry
- Polygons and Circles

Geometric Measure and Dimension
- Geometry: Three-Dimensional Shapes and Attributes
- Perimeter, Circumference, and Area
- Surface Area and Volume

Modeling with Geometry
- Coordinate Geometry
- Geometry: Three-Dimensional Shapes and Attributes
- Perimeter, Circumference, and Area
- Polygons and Circles
- Right Triangles and Trigonometry
- Surface Area and Volume

Conditional Probability and the Rules of Probability
- Combinatorics and Probability

Using Probability to Make Decisions
- Combinatorics and Probability

Interpreting Categorical and Quantitative Data
- Data Representation and Analysis

Making Inferences and Justifying Conclusions
- Data Representation and Analysis

Skills in original Core Progress™ Learning Progression

Development of the original Core Progress learning progressions for reading and math took into account research as well as state and other standards.

Figure 11 shows the organization of the early literacy and early numeracy skills in the learning progression within three key domains: (1) Word Knowledge and Skills, (2) Comprehension Strategies and Constructing Meaning, and (3) Numbers and Operations.

Figure 11: Core Progress™ for Reading Learning Progression: Domains and Skill Sets (Early Literacy)

Word Knowledge and Skills

Alphabetic Principle
- Alphabetic Knowledge
- Alphabetic Sequence
- Letter Sounds

Concept of Word
- Print Concepts: Word Length
- Print Concepts: Word Borders
- Print Concepts: Letters and Words

Visual Discrimination
- Letters
- Identification and Word Matching

Phonemic Awareness
- Rhyming and Word Families
- Blending Word Parts
- Blending Phonemes
- Initial and Final Phonemes
- Consonant Blends (PA)
- Medial Phoneme Discrimination
- Phoneme Segmentation
- Phoneme Isolation/Manipulation

Phonics
- Short Vowel Sounds
- Initial Consonant Sounds
- Final Consonant Sounds
- Long Vowel Sounds
- Variant Vowel Sounds
- Consonant Blends (PH)
- Consonant Digraphs
- Other Vowel Sounds
- Sound-Symbol Correspondence: Consonants
- Word Building
- Sound-Symbol Correspondence: Vowels
- Word Families/Rhyming

Structural Analysis
- Words with Affixes
- Syllabification
- Compound Words

Vocabulary
- Word Facility
- Synonyms
- Antonyms

Comprehension Strategies and Constructing Meaning

Sentence-level Comprehension
- Comprehension at the Sentence Level

Paragraph-level Comprehension
- Comprehension of Paragraphs

Numbers and Operations

Early Numeracy
- Number Naming and Number Identification
- Number Object Correspondence
- Sequence Completion
- Composing and Decomposing
- Measurement

In Figure 12, for reading, the learning progression is organized by five domains: (1) Word Knowledge and Skills, (2) Comprehension Strategies and Constructing Meaning, (3) Understanding Author's Craft, (4) Analyzing Literary Text, and (5) Analyzing Argument and Evaluating Text.

Figure 12: Core Progress™ for Reading Learning Progression: Domains and Skills

Word Knowledge and Skills

Vocabulary Strategies
- Use context clues
- Use structural analysis

Vocabulary Knowledge
- Recognize and understand synonyms
- Recognize and understand homonyms and multi-meaning words
- Recognize connotation and denotation
- Understand idioms
- Understand analogies

Analyzing Literary Text

Literary Elements
- Identify and understand elements of plot
- Identify and understand setting
- Identify characters and understand characterization
- Identify and understand theme
- Identify the narrator and point of view

Genre Characteristics
- Identify fiction and nonfiction, reality and fantasy
- Identify and understand characteristics of genres

Understanding Author's Craft

Author's Choices
- Understand figurative language
- Understand literary devices
- Identify sensory detail

Comprehension Strategies and Constructing Meaning

Reading Process Skills
- Make predictions
- Identify author's purpose
- Identify and understand text features
- Recognize an accurate summary of text
- Use repair strategies

Constructing Meaning
- Understand vocabulary in context
- Draw conclusions
- Identify and understand main ideas
- Identify details
- Extend meaning or form generalizations
- Identify and differentiate fact and opinion

Organizational Structure
- Identify organizational structure
- Understand cause and effect
- Understand comparison and contrast
- Identify and understand sequence

Analyzing Argument and Evaluating Text

Analysis
- Identify bias and analyze text for logical fallacies
- Identify and understand persuasion

Evaluation
- Evaluate reasoning and support
- Evaluate credibility

Figure 13 shows the math learning progression's organization within four domains: (1) Numbers and Operations, (2) Algebra (3) Geometry and Measurement, (4) Data Analysis, Statistics, and Probability.

Figure 13: Core Progress™ for Math Learning Progression: Domains and Skill Sets

Numbers and Operations

- Count with objects and numbers
- Identify odd and even numbers
- Relate place and value to a whole number
- Add and subtract whole numbers without regrouping
- Add and subtract whole numbers with regrouping
- Multiply whole numbers
- Divide whole numbers without a remainder in the quotient
- Divide whole numbers with a remainder in the quotient
- Identify, compare, and order fractions
- Add and subtract fractions with like denominators
- Find prime factors, common factors, and common multiples
- Add and subtract fractions with unlike denominators
- Convert between an improper fraction and a mixed number
- Relate a decimal to a fraction
- Relate place and value to a decimal number
- Add or subtract decimal numbers
- Divide a whole number resulting in a decimal quotient
- Multiply and divide with fractions
- Multiply and divide with decimals
- Relate a decimal number to a percent
- Solve a proportion, rate, or ratio
- Evaluate a numerical expression
- Perform operations with integers
- Determine a square root
- Solve a problem involving percents

Data Analysis, Statistics, and Probability

- Read or answer a question about charts, tables, or graphs
- Use a chart, table, or graph to represent data
- Determine a measure of central tendency
- Use a proportion to make an estimate
- Determine the probability of one or more events

Algebra

- Relate a rule to a pattern
- Determine the operation given a situation
- Graph on a coordinate plane
- Evaluate an algebraic expression or function
- Solve a linear equation
- Determine a linear equation
- Identify characteristics of a linear equation or function
- Solve a system of linear equations
- Determine a system of linear equations
- Simplify an algebraic expression
- Solve a linear inequality
- Solve a nonlinear equation
- Graph a 1-variable inequality

Geometry and Measurement

- Relate money to symbols, words, and amounts
- Use the vocabulary of geometry and measurement
- Determine a missing figure in a pattern
- Determine a measurement
- Tell time
- Calculate elapsed time
- Solve a problem involving the perimeter of a shape
- Solve a problem involving the area of a shape
- Identify congruence and similarity of geometric shapes
- Solve a problem involving the surface area or volume of a solid
- Determine a missing measure or dimension of a shape

Psychometric Properties

The computer-adaptive STAR Assessments are highly rated for reliability and validity by key federal groups, such as the National Center on Intensive Intervention, the National Center on Response to Intervention, and the National Center on Student Progress Monitoring.

In 2012, STAR Assessments were highly rated for progress monitoring by the federally funded National Center on Intensive Intervention (NCII), whose mission is "to build state and district capacity to support educators in using data-based individualization to effectively implement intensive interventions in reading, mathematics, and behavior in Grades K–12" (http://www.intensiveintervention.org), in the organization's first review of progress-monitoring tools.

Earlier, in 2009, the U.S. Department of Education began funding the National Center on Response to Intervention (NCRTI), whose mission is "to provide technical assistance to states and districts and building the capacity of states to assist districts in implementing proven models for RTI/EIS" (www.rti4success.org). That same year, STAR Early Literacy, STAR Reading, and STAR Math were among the first assessments highly rated by the NCRTI for screening and progress monitoring. In subsequent reviews, STAR Assessments have maintained strong ratings, meaning they fulfill both these key elements of a school's RTI framework. For information on using STAR Enterprise assessments in intervention settings, see Pupose and Frequency, p. 28.

STAR Assessments are highly rated for reliability and validity by key federal groups, such as the National Center on Intensive Intervention, the National Center on Response to Intervention, and the National Center on Student Progress Monitoring.

STAR Assessments have received high marks as tools for Response to Intervention since 2006 when the NCRTI's predecessor, the National Center on Student Progress Monitoring, first deemed STAR Early Literacy, STAR Reading, and STAR Math reliable and valid for progress monitoring (http://www.studentprogress.org/chart/docs/print_chart122007.pdf).

Each STAR assessment followed a unique path to determine reliability and validity, which is explained below along with lists of the wide range of assessments to which each STAR assessment relates.

Reliability and validity of STAR Early Literacy Enterprise™

Reliability

Test reliability is often described as a measure of the consistency of test scores; tests must yield somewhat consistent results in order to be useful. Two kinds of consistency are of concern when evaluating a test's measurement precision: internal consistency and the consistency of the scores obtained when an assessment is given two or more times.

The internal consistency of STAR Early Literacy Enterprise assessments has been calculated using a method referred to as *generic reliability*, which uses the conditional measurement error of individual students' tests to estimate what percentage of the variation in STAR test scores is attributable to the attribute the test is intended to measure. Consistency of scores across multiple administrations of the assessment to the same students is measured by *retest reliability*, which is the coefficient of correlation between pairs of test scores earned by the same students on different occasions.

The generic estimates of internal consistency reliability were calculated from analyes of the test scores and their estimated conditional measurement error in a balanced random sample of 10,000 students in each grade, pre-K through 3, who took STAR Early Literacy Enterprise in fall 2012. Another random sample of students who took SEL Enterprise two or more times within a 2-week period across the same school year was analyzed in to order to calculate retest reliability. Table 5 displays both the internal consistency and the retest reliability estimates, by grade and for the five grades combined. The combined-grades reliability coefficients are 0.85 for internal consistency, and 0.79 for consistency on retest.

Table 5: Internal Consistency and Retest Reliability of STAR Early Literacy Enterprise™
Assessments Taken Between June 2012 and June 2013

	Internal Consistency		Retest Reliability	
Grade	**Students**	**Reliability Coefficient**	**Students**	**Reliability Coefficient**
All	3,083,334	0.85	25,000	0.79
Pre-K	54,144	0.81	5,000	0.59
K	1,427,660	0.80	5,000	0.50
1	1,187,216	0.82	5,000	0.47
2	340,912	0.85	5,000	0.64
3	73,402	0.89	5,000	0.74

Validity

Evidence of the validity of any educational assessment has a number of facets that, in aggregate, constitute empirical support for the use of the assessments for specific purposes, and for the inferences that are to be made on the basis of students' test scores. A crucial facet is the content of the tests; content-related evidence of validity lies in the degree of correspondence, or alignment, between the knowledge and skills measured by an assessment's test items and the knowledge and skills intended to be taught and learned in a given curriculum at a given grade level or levels. STAR Early Literacy Enterprise content is aligned to curriculum standards at the state and national levels—including the Common Core State Standards (see Core Progress Learning Progressions—The Bridge Between Assessment and Instruction, p. 10).

It could be argued that solid evidence of psychometric reliability, combined with a high degree of alignment of test content to curriculum standards, is evidence enough of an assessment's validity. However, a number of other measures complement or corroborate those two facets and serve to further strengthen an assessment's claims of validity: cumulative evidence of criterion-related validity, convergent and discriminant validity evidence, demonstrated accuracy of screening and diagnostic classifications, among others. Altogether, these are among the components of *construct validity*, in other words, evidence the assessments measure specific attributes as claimed and are appropriate for specific uses and inferences. Construct validity evidence is cumulative in nature; when first released, an assessment may have sound evidence that is consistent with construct validity, but over time additional evidence may and should be accumulated and documented.

To support, STAR Early Literacy as a measure of literacy skills, Renaissance Learning knew it was necessary that its scores correlate highly with other measures of reading, literacy, and readiness. To evaluate this, Renaissance Learning performed a multifaceted validity research study of STAR Early Literacy prior to the assessment's initial release to assess reliability, criterion-related validity, and score distributions by age and grade. The participating school districts, specific schools, and individual students were

approximately representative of the U.S. school population in terms of geographic region, school system and per-grade district enrollment, and socioeconomic status. The final study sample included approximately 11,000 students from 84 schools in the U.S. and Canada.

Renaissance Learning asked teachers participating in the study to submit student scores from other assessments of reading, early literacy, readiness, and social skills. Scores were received for more than 2,400 students. The resulting correlation estimates were substantial and reflect well on the concurrent validity of STAR Early Literacy as a tool for assessing early literacy skills. Subsequent to the original validity study, a number of additional studies, including both concurrent and predictive correlational studies, studies of classification accuracy, and others, have been conducted. Table 6 summarizes the results of more than 80 concurrent and predictive validity studies conducted for STAR Early Literacy. The average correlations observed in these studies range from 0.52 to 0.77; correlations in that range are considered moderate to strong. Below the table is a list of major assessments of early literacy skills that have been found to correlate well with scores on STAR Early Literacy.

Table 6: Summary of STAR Early Literacy™ Validity Studies

	Predictive			Concurrent		
Grade	**Studies**	**Students**	**Average Correlation**	**Studies**	**Students**	**Average Correlation**
K	15	30,423	0.52	6	198	0.64
1	15	24,525	0.62	7	281	0.68
2	15	5,370	0.67	12	513	0.52
3	2	558	0.67	9	384	0.57

STAR Early Literacy™ relates to several assessments of early literacy skills
Studies have been conducted with STAR Early Literacy and the following assessments to correlate the tests:

- AIMSweb
- Alabama Early Learning Inventory
- Brigance K & 1 Screen for Kindergarten and First Grade Children
- Canadian Achievement Test
- Child Observation Record (COR)
- Developing Skills Checklist (DSC)
- Developmental Indicators for the Assessment of Learning (DIAL-3)
- Dynamic Indicators of Basic Early Literacy Skills (DIBELS)
- easyCBM
- Florida Comprehensive Assessment Test (FCAT)
- Gates-MacGinitie Reading Test (GMRT)
- Group Reading Assessment and Diagnostic Evaluation (GRADE)
- Indiana Statewide Testing for Educational Progress (ISTEP)
- Iowa Test of Basic Skills (ITBS)
- Kaufman Survey of Early Academic and Language Skills (K-SEALS)
- Metropolitan Early Childhood Assessment Program (MKIDS)
- Metropolitan Readiness Test (MRT)
- Michigan Literacy Progress Profile (MLPP)
- NWEA Levels Test
- Running Records

- Stanford Achievement Test (SAT-9)
- Stanford Test of Academic Skills
- TerraNova
- Test of Phonological Awareness (TOPA)
- Texas Primary Reading Inventory (TPRI)
- Woodcock Reading Mastery Tests-Revised/Normative Update

Reliability and validity of STAR Reading Enterprise™

Reliability

The reliability of STAR Reading Enterprise assessments was estimated using two methods, internal consistency (generic reliability coefficients) and test-retest correlation coefficients, in a random national sample of more than 1.2 million STAR Reading Enterprise tests administered between September 2012 and June 2013. The retest correlation coefficients were based on samples of 5,000 students per grade, from the same dataset. Results are displayed in Table 7. The internal consistency reliability estimates were very high, equaling or exceeding those of most major published assessments. Over all grades combined, the reliability was 0.97; it ranged from 0.93 to 0.95 within grades. Retest reliability estimates were 0.90 for all grades combined, and ranged from 0.54 to 0.85 within grades.

Table 7: Internal Consistency and Retest Reliability of STAR Reading Enterprise™

Assessments Taken Between June 2012 and June 2013

	Internal Consistency		Retest Reliability	
	Students	Reliability Coefficient	Students	Reliability Coefficient
All	1,227,915	0.97	60,000	0.90
1	100,000	0.95	5,000	0.54
2	100,000	0.94	5,000	0.66
3	100,000	0.94	5,000	0.75
4	100,000	0.93	5,000	0.77
5	100,000	0.93	5,000	0.78
6	100,000	0.93	5,000	0.83
7	100,000	0.94	5,000	0.82
8	100,000	0.94	5,000	0.83
9	95,171	0.94	5,000	0.85
10	94,624	0.95	5,000	0.85
11	93,118	0.95	5,000	0.85
12	89,031	0.95	5,000	0.85

Validity

As noted in the discussion of STAR Early Literacy validity, content is a crucial facet of test validity; content-related evidence of validity lies in the degree of correspondence, or alignment, between the knowledge and skills measured by an assessment's test items and the knowledge and skills intended to be taught and learned in a given curriculum at a given grade level or levels. STAR Reading Enterprise content is aligned to curriculum standards at the state and national levels—including the Common Core State Standards (see Core Progress Learning Progressions—The Bridge Between Assessment and Instruction, p. 10).

Psychometric reliability, combined with a high degree of alignment of test content to curriculum standards may be evidence enough of an assessment's validity. However, other measures complement or corroborate those two facets and serve to further strengthen an assessment's claims of validity.

To support STAR Reading Enterprise as a measure of both reading comprehension and a broad range of other reading skills, Renaissance Learning has collected a wide range of correlations between scores on STAR Reading and scores on other recognized, established measures of different aspects of reading achievement, such as survey achievement tests, diagnostic reading measures, and state accountability tests, among others. Table 8 summarizes the results of more than 400 concurrent and predictive validity studies conducted for STAR Reading, involving a total of more than 1 million students. The average correlations observed in these studies range from 0.60 to 0.87; correlations in that range are considered strong. Below the table is a list of state assessments that have been found to correlate well with scores on STAR Reading.

Table 8: Summary of STAR Reading™ Validity Studies

	Predictive			Concurrent and Other External Validity		
Grade	**Studies**	**Students**	**Average Correlation**	**Studies**	**Students**	**Average Correlation**
1	6	74,77	.68	15	1,135	.77
2	10	184,434	.78	32	4,142	.72
3	30	200,929	.80	44	4,051	.75
4	25	185,528	.82	41	5,409	.75
5	29	126,029	.82	40	3,588	.75
6	23	82,189	.82	37	2,728	.71
7	23	64,978	.81	33	3,294	.70
8	25	34,764	.81	29	2,148	.72
9	8	9,567	.83	15	949	.72
10	9	7,021	.85	11	566	.61
11	6	6,653	.86	6	324	.70
12	2	3,107	.86	4	165	.74

STAR Reading™ relates to several state assessments

Studies have been conducted with STAR Reading and the following assessments to statistically link[5] the tests:

- ACT EXPLORE
- Alabama Reading and Mathematics Test+ (ARMT+)
- Alaska's Standards Based Assessment (SBA)
- Arizona's Instrument to Measure Standards (AIMS)
- Arkansas Augmented Benchmark Examinations (AABE)
- California Standards Tests (CST)
- Colorado—Transitional Colorado Assessment Program (TCAP)
- Connecticut Mastery Test (CMT4)
- Delaware Comprehensive Assessment System (DCAS)
- Florida Comprehensive Assessment Test 2.0 (FCAT 2.0)

[5] Statistical linking studies are continuously underway and Performance Reports for new states are released on a regular basis. For the most up-to-date list of state assessments with statistical links to STAR Assessments, email research@renlearn.com. Technical manuals are also available upon request and include information on the assessments that correlate with STAR Assessments.

- Georgia's Criterion-Referenced Competency Tests (CRCT)
- Idaho Standards Achievement Tests (ISAT)
- Illinois Standards Achievement Test (ISAT)
- Indiana Reading Evaluation and Determination (IREAD-3)
- Indiana Statewide Testing for Education Progress-Plus (ISTEP+) Assessments
- Iowa Assessment (IA)
- Kansas State Assessment Program (KSAP)
- Kentucky Performance Rating for Educational Progress (K-PREP) Tests
- Louisiana Educational Assessment Program (LEAP) and Integrated Educational Assessment Program (iLEAP) Assessments
- Maine—New England Common Assessment Program (NECAP)
- Massachusetts Comprehensive Assessment System (MCAS)
- Michigan Educational Assessment Program (MEAP)
- Minnesota Comprehensive Assessments (MCAs)
- Mississippi Curriculum Test, Second Edition (MCT2)
- Missouri Assessment Program (MAP) Grade-Level Assessments
- Montana's Criterion-Referenced Test (CRT)
- Nebraska State Accountability (NeSA) Reading Test
- Nevada's Criterion-Referenced Test (CRT)
- New Hampshire—New England Common Assessment Program (NECAP)
- New Jersey Assessment of Skills and Knowledge (NJ ASK)
- New Mexico Standards Based Assessments (SBA)
- New York State Assessment Program (NYSTP)
- North Carolina End-of-Grade (NC EOG) Tests
- North Dakota State Assessment (NDSA)
- Ohio Achievement Assessments (OAA)
- Oklahoma Core Curriculum Tests (OCCT)
- Pennsylvania's System of School Assessment (PSSA)
- Rhode Island—New England Common Assessment Program (NECAP)
- South Carolina Palmetto Assessment of State Standards (SCPASS)
- South Dakota State Test of Educational Progress (DSTEP)
- Tennessee Comprehensive Assessment Program (TCAP)
- Texas—State of Texas Assessments of Academic Readiness (STAAR)
- Utah's Criterion-Referenced Test for English Language Arts
- Vermont—New England Common Assessment Program (NECAP)
- Virginia Standards of Learning (SOL)
- Washington—Measurements of Student Progress (MSP)
- West Virginia Educational Standards Test 2 (WESTEST 2)
- Wisconsin Knowledge and Concepts Examination (WKCE)
- Wyoming—Proficiency Assessments for Wyoming Students (PAWS)

Reliability and validity of STAR Math Enterprise™

Reliability

The reliability of STAR Math Enterprise assessments was estimated using two methods, internal consistency (generic reliability coefficients) and test-retest correlation coefficients, in a national sample of more than 9 million STAR Math Enterprise tests administered between September 2012 and June 2013. The retest correlation coefficients were based on random samples of 5,000 students per grade from the same dataset. Results are displayed in Table 9. The internal consistency reliability estimates were very high, equaling or exceeding those of most major published assessments. Over all grades combined, the reliability was 0.97; it ranged from 0.90 to 0.95 within grades. Retest reliability estimates were 0.93 for all grades combined, and ranged from 0.76 to 0.84 within grades.

Table 9: Internal Consistency and Retest Reliability of STAR Math Enterprise™
Assessments Taken Between June 2012 and June 2013

Validity

Grade	Internal Consistency		Retest Reliability	
	Students	Reliability Coefficient	Students	Reliability Coefficient
All	9,311,595	0.97	60,000	0.93
1	805,980	0.90	5,000	0.76
2	1,254,611	0.91	5,000	0.80
3	1,330,600	0.92	5,000	0.81
4	1,306,386	0.92	5,000	0.83
5	1,227,139	0.93	5,000	0.83
6	968,367	0.93	5,000	0.84
7	785,789	0.94	5,000	0.82
8	721,994	0.94	5,000	0.83
9	327,455	0.93	5,000	0.83
10	241,728	0.94	5,000	0.82
11	167,902	0.94	5,000	0.83
12	108,492	0.95	5,000	0.80

As noted in the discussion of STAR Early Literacy and STAR Reading validity, content is a crucial facet of test validity; content-related evidence of validity lies in the degree of correspondence, or alignment, between the knowledge and skills measured by an assessment's test items and the knowledge and skills intended to be taught and learned in a given curriculum at a given grade level or levels. STAR Math Enterprise content is aligned to curriculum standards at the state and national levels—including the Common Core State Standards (see Core Progress Learning Progressions—The Bridge Between Assessment and Instruction, p. 10).

Psychometric reliability, combined with a high degree of alignment of test content to curriculum standards may be evidence enough of an assessment's validity. However, other measures complement or corroborate those two facets and serve to further strengthen an assessment's claims of validity.

To support STAR Math Enterprise as a measure of a broad range of mathematics skills, Renaissance Learning has collected a wide range of correlations between scores on STAR Math and scores on other recognized, established measures of different aspects of mathematics achievement, such as survey achievement tests, diagnostic math measures, and state accountability tests, among others. Table 10 summarizes the results of

more than 400 concurrent and predictive validity studies conducted for STAR Math, involving a total of more than 400,000 students. The average correlations observed in these studies range from 0.55 to 0.80; correlations in that range are considered moderate to strong. Below the table is a list of state assessments that have been found to correlate well with scores on STAR Math.

Table 10: Summary of STAR Math™ Validity Studies

	Predictive			Concurrent		
Grade	**Studies**	**Students**	**Average Correlation**	**Studies**	**Students**	**Average Correlation**
1	6	11,880	.55	6	179	.58
2	10	33,076	.63	17	987	.61
3	30	52,604	.66	49	6,400	.61
4	23	55,285	.69	49	5,823	.59
5	29	39,869	.70	58	6,873	.64
6	13	27,663	.73	37	4,202	.66
7	15	18,919	.75	29	3,361	.64
8	11	12,780	.76	29	3,713	.65
9	6	2,545	.78	13	665	.57
10	6	2,236	.79	10	334	.60
11	6	1,921	.80	10	495	.68
12	2	885	.77	9	233	.68

STAR Math™ relates to several state assessments

Studies have been conducted with STAR Math and the following assessments to statistically link[6] the tests:

- ACT EXPLORE
- Alabama Reading and Mathematics Test+ (ARMT+)
- Arizona's Instrument to Measure Standards (AIMS)
- Arkansas Augmented Benchmark Examinations (AABE)
- California Standards Tests (CST)
- Colorado—Transitional Colorado Assessment Program (TCAP)
- Connecticut Mastery Test (CMT4)
- Delaware Comprehensive Assessment System (DCAS)
- Florida Comprehensive Assessment Test 2.0 (FCAT 2.0)
- Georgia's Criterion-Referenced Competency Tests (CRCT)
- Idaho Standards Achievement Test (ISAT)
- Illinois Standards Achievement Test (ISAT)
- Indiana Statewide Testing for Education Progress-Plus (ISTEP+) Assessments
- Iowa Assessment (IA)
- Kansas State Assessment Program (KSAP)
- Kentucky Performance Rating for Educational Progress (K-PREP)
- Louisiana Educational Assessment Program (LEAP) and Integrated Educational Assessment Program (iLEAP) Assessments

[6] Statistical linking studies are continuously underway and Performance Reports for new states are released on a regular basis. For the most up-to-date list of state assessments with statistical links to STAR Reading or STAR Math, email research@renlearn.com. Technical manuals are also available upon request and include information on the assessments that correlate with STAR Assessments.

- Maine—New England Common Assessment Program (NECAP)
- Massachusetts Comprehensive Assessment System (MCAS)
- Michigan Educational Assessment Program (MEAP)
- Minnesota Comprehensive Assessments (MCAs)
- Mississippi Curriculum Test, Second Edition (MCT2)
- Missouri Assessment Program (MAP) Grade-Level Assessments
- Montana's Criterion-Referenced Test (CRT)
- Nevada's Criterion-Referenced Test (CRT)
- New Hampshire—New England Common Assessment Program (NECAP)
- New Jersey Assessment of Skills and Knowledge (NJ ASK)
- New Mexico Standards Based Assessments (SBA)
- New York State Assessment Program (NYSTP)
- North Carolina End-of-Grade (NC EOG) Test
- Ohio Achievement Assessments (OAA)
- Oklahoma Core Curriculum Tests (OCCT)
- Pennsylvania's System of School Assessment (PSSA)
- Rhode Island—New England Common Assessment Program (NECAP)
- South Carolina Palmetto Assessment of State Standards (SCPASS)
- South Dakota State Test of Educational Progress (DSTEP)
- Tennessee Comprehensive Assessment Program (TCAP)
- Texas—State of Texas Assessments of Academic Readiness (STAAR)
- Vermont—New England Common Assessment Program (NECAP)
- Virginia Standards of Learning (SOL)
- Washington—Measurements of Student Progress (MSP)
- West Virginia Educational Standards Test 2 (WESTEST 2)
- Wisconsin Knowledge and Concepts Examination (WKCE)
- Wyoming—Proficiency Assessments for Wyoming Students (PAWS)

Purpose and Frequency

Most schools administer STAR Enterprise assessments to all students in the fall, winter, and spring for screening purposes. If educators want to establish a trend line for students (visible in reports of STAR results) to forecast proficiency on state tests or mastery of standards, they must administer an additional test in late fall. This way, after the winter screening, three data points have been established so the software can chart students' growth trajectories.

Teachers who monitor progress more closely for specific students, in an intervention or other setting, or for instructional planning, typically test more frequently. Although STAR Assessments can be administered as often as weekly, an important general guideline is to administer assessments to students only when educators are prepared to act upon the resulting data.

Response to Intervention screening and progress monitoring

Response to Intervention (RTI)—also known as a Multi-Tier System of Supports (MTSS)— is a framework for making instructional decisions based on data, in order to accelerate learning for all students. Interim assessments play a key role in RTI, helping to provide data to inform and improve instruction. Interim assessments are generally used for screening/benchmarking or progress monitoring. STAR Enterprise assessments are used for both of these purposes:

- **Screening and benchmarking** periodic assessment, typically administered two to four times per year to monitor growth of a group toward a proficiency target, which also may provide information about the standards students have likely mastered.

- **Progress-monitoring assessment**—defined as measures of academic performance by the National Center on Response to Intervention—administered more frequently than annually, but as often as monthly to monitor students' growth trajectories or weekly in intervention situations to measure individual student progress. Progress-monitoring assessments measure growth during the year and longitudinally over two or more years. Also included in this category are diagnostic assessments administered as needed to help identify specific areas of weakness. (For more information, see Computer Adaptive Testing, p. 7.)

Growth measurement: Scaled score, growth norms, and student growth percentile

Because changes in student achievement do not happen overnight, measuring growth is essential to understanding the effects of instruction. Renaissance Learning has unique insight into how students grow through ongoing study of data from the millions of tests taken by students at thousands of schools. During the 2012–2013 school year alone, more than 45 million STAR tests were taken. With this wealth of data, we are able to calculate growth norms. We can approximate how much growth is typical for students of different achievement levels in different grades from one time period to another.

In addition to screening students to forecast proficiency on end-of-year summative tests and progress monitoring their growth throughout the year, teachers can use STAR Enterprise assessments to capture a picture of each student's overall growth from the beginning of the school year to the end, or in semester increments.

During the 2012–2013 school year alone, more than 45 million STAR tests were taken.

Scaled score

STAR Enterprise assessments generate a scaled score (SS), which is useful for comparing student performance over time. The same range is used for all students, so scaled scores help to compare student

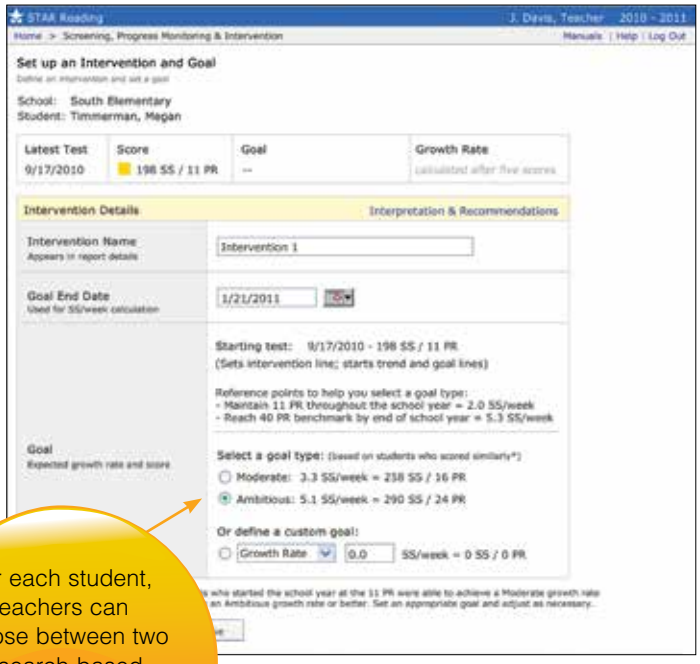
performance across grade levels. Any scaled score increase indicates that a student has experienced growth. STAR Reading Enterprise and STAR Math Enterprise scaled scores range from 0–1400, while STAR Early Literacy Enterprise scaled scores range from 300–900 and relate directly to specific literacy classifications (Emergent Reader, Transitional Reader, and Probable Reader).

Growth norms

Just as meteorologists use statistical models to predict the weather, educational researchers use growth models to identify patterns in student growth. Renaissance Learning has developed such a model based on study of the growth patterns for millions of students. Growth norms indicate typical rates of growth per week and are differentiated by subject, grade, and starting score. These norms are updated every year, as more and more students take STAR Enterprise assessments, to ensure the growth rates reflect the most up-to-date data possible.

Currently, the STAR Reading Enterprise data set includes more than 3.5 million students, STAR Math Enterprise includes more than 2.2 million students, and STAR Early Literacy Enterprise includes more than 400,000 students. Using this information, STAR software is able to provide a projected scaled score for the end of the year, based on a growth rate achieved by 50 percent of students with a similar percentile rank as the student for whom you are setting goals. This information appears on STAR State Standards Reports and provides educators with key information about how students grow over time.

Figure 14: Goal-Setting Wizard



For each student, teachers can choose between two research-based recommendations for goal setting or determine a custom goal.

STAR growth norms also drive the Goal-Setting Wizard (see Figure 14), which helps educators set challenging, but reasonable, progress-monitoring goals personalized to each student.

Student growth percentile

Student growth percentile (SGP) was first developed by Dr. Damian Betebenner from the National Center for the Improvement of Educational Assessment, in partnership with the Colorado Department of Education. Dr. Dan Bolt, at the University of Wisconsin-Madison, assisted Renaissance Learning in adapting SGP for STAR Assessments.

SGP compares a student's growth to that of his/her *academic peers* nationwide and helps educators understand student growth.

SGP compares a student's growth to that of his/her *academic peers* nationwide and helps educators understand student growth. A student's academic peers are students at the same grade level and at similar achievement levels as that student.

An advantage of SGP is that it gives a clear picture of whether a student's growth is more or less than can be expected. A student must take at least two STAR Enterprise assessments during a school year, within specific testing windows (fall to winter, winter to spring, or fall to spring), in order to generate an SGP score and measure growth. This score helps educators at the classroom, school, and district level address important questions via tools such as reports and the Growth Proficiency Chart (see Figure 15).

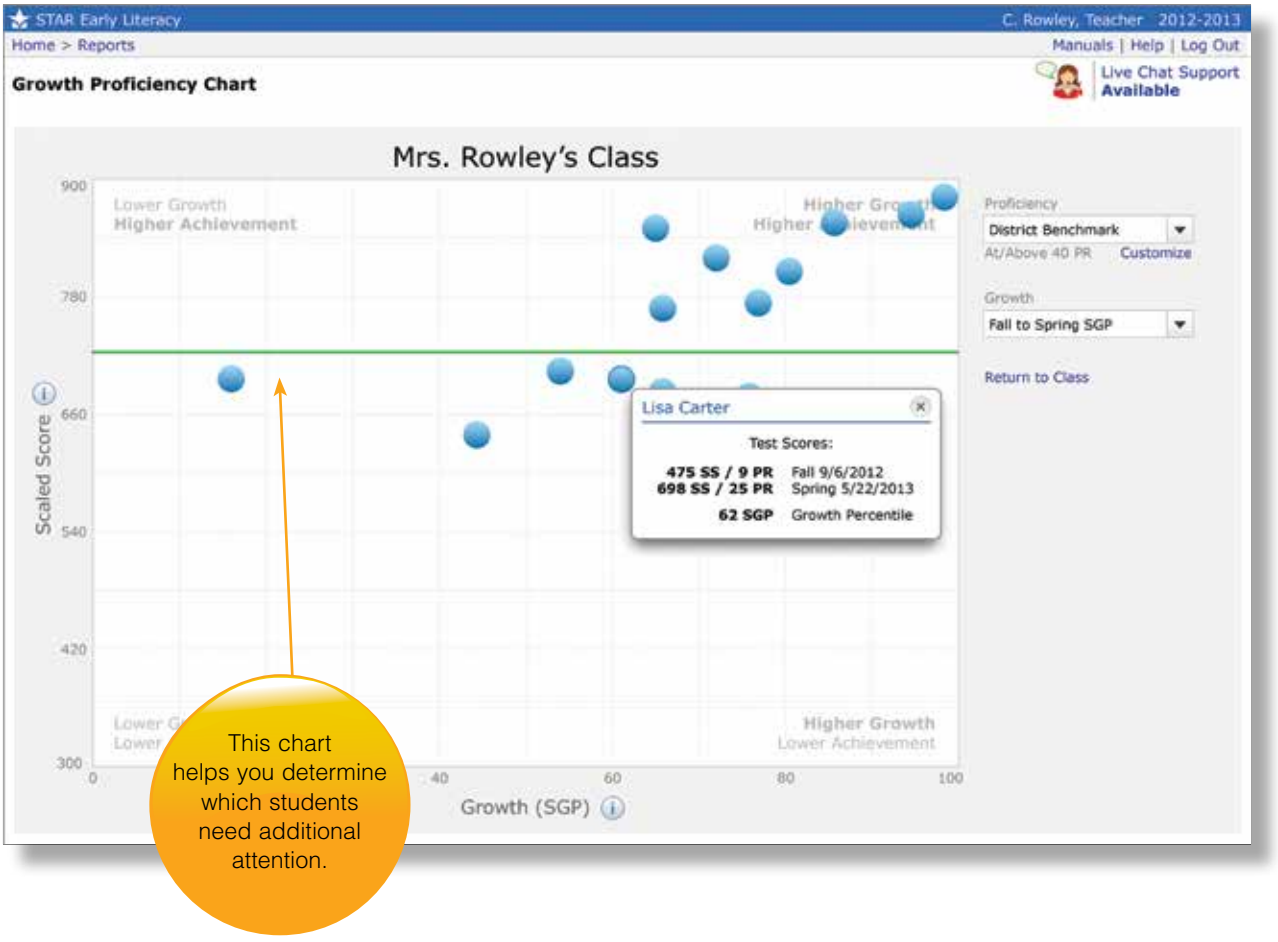
For teachers:

- Did students grow from one testing period to the next?
- Did students of all abilities grow?
- Did students grow as much as expected? More? Less?
- Did students in intervention grow more than their peers nationwide?
- Did my intervention strategies lead to greater growth?

For administrators:

- How much did all students in my district grow?
- Did students of all abilities grow?
- Did students grow as much as expected? More? Less?
- Did students in some classes, grades, or schools grow more than others? What does that say about our core curriculum, intervention strategies, and programs and/or professional development needs?

Figure 15: Growth Proficiency Chart



Instructional planning with Core Progress™

As mentioned, after a student takes a STAR Enterprise assessment, the software places the resulting scaled score on the Core Progress learning progression, which reports skills the student has likely mastered, those they are ready to develop next, and suggestions for the teacher to focus instruction. In essence, Core Progress serves as a road map to help teachers understand both where students have been and where they need to go to become college and career ready.

Instructional Planning Reports from STAR Enterprise provide teachers lists of skills individual students—and at the class level, groups of students—are ready to develop next (see student example, Figure 16). Within Core Progress, teachers can search for the skills and domains listed on the reports to further focus next steps for students.

Figure 17 shows a visual of the Core Progress software. Within each domain, headings match those outlined in the CCSS, and under each heading, grade-level domain expectations are identified. The software also provides resources for instruction, including Worked Examples, Skill Probes, Performance Tasks, and links to third-party educational resources.

The Record Book is another pathway to see suggested skills with which students need additional practice. This resource is especially helpful for teachers of students who need intervention, in that it suggests skills for differentiated instruction and allows teachers to create instructional groups designed for specific student needs.

Figure 16: Instructional Planning Report

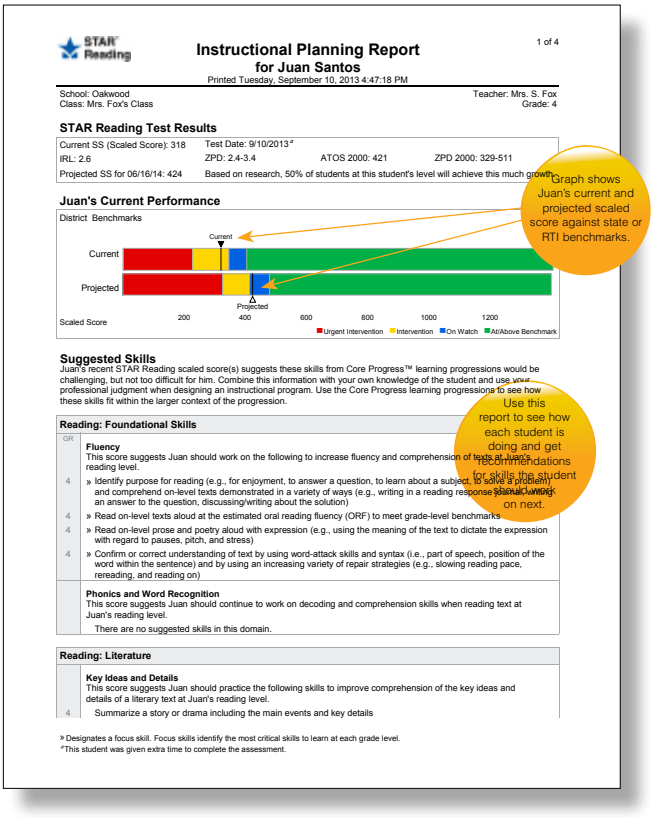
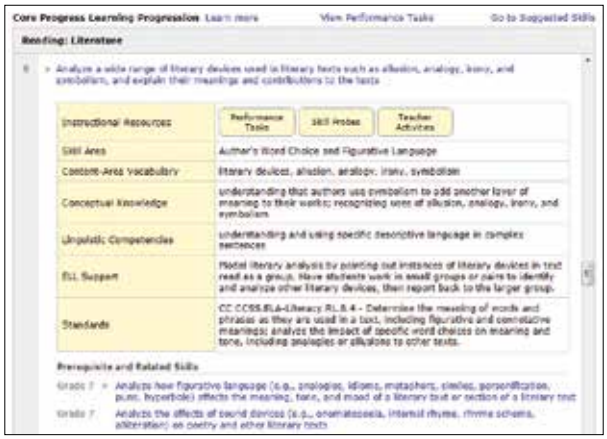


Figure 17: Core Progress™ Learning Progression for Reading—Built for the Common Core State Standards Example Screen



Predicting achievement: Linking studies and performance reporting

Will my students perform well on the state test? is one of the most serious and challenging questions teachers and administrators face. STAR Enterprise assessments are integral tools for educators to use to evaluate student progress toward proficiency.

Because STAR Assessments are computerized, achievement data for millions of students nationwide is collected each year. The Research Department at Renaissance Learning has analyzed this data and linked student performance on STAR Reading and STAR Math to student performance on several summative end-of-year state tests. (For a full list of state assessments to which STAR Reading and STAR Math have been linked, see Psychometric Properties, pp. 23, 26.)

The linking studies combined with the Renaissance Learning growth model (see Growth Norms, p. 29), which is based on STAR test results from millions of students, drive the information displayed in STAR Enterprise State Performance Reports. With versions available at the student, class, and district levels, these reports are used to monitor proficiency not only periodically, but also, more importantly, early. This way, educators know whether students are on track to achieve proficiency on the state test, and if not, they can make key instructional decisions while there is still time to do so.

Two of the reports are specifically for teachers (see Figures 18 and 19):

- **State Performance Report—Student**: Graphs a student's STAR Reading or STAR Math scores and trend line (indicates projected growth) for easy comparison with the pathway to proficiency on state reading and math tests.

- **State Performance Report—Class**: Provides a trend line at the class level depicting the average STAR Reading or STAR Math scaled score, making group progress available at a glance. Also lists individual student scores and categorizes performance as Below or On the pathway.

Figure 18: State Performance Report—Student

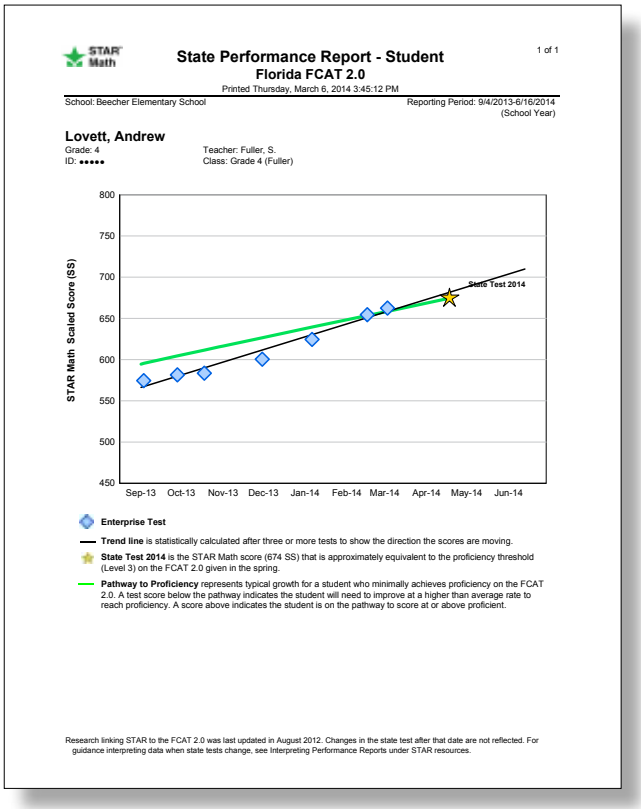
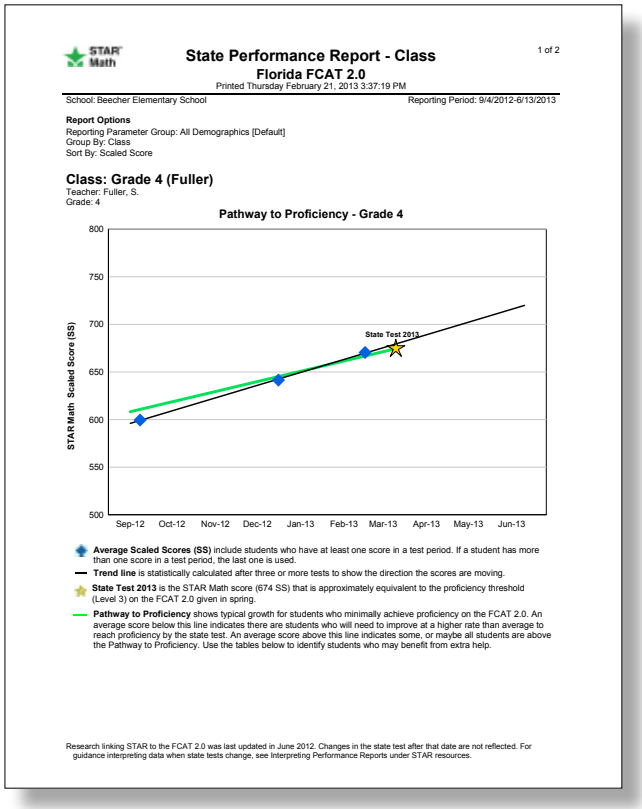


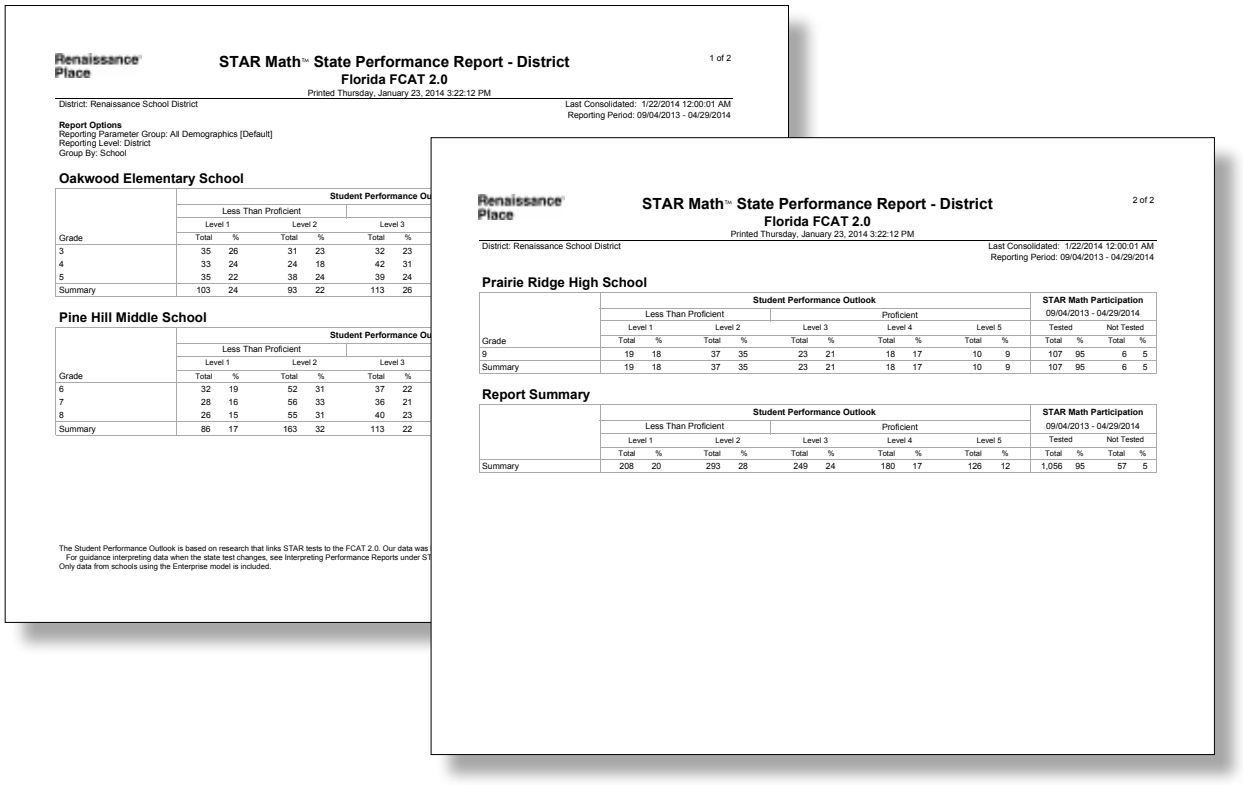
Figure 19: State Performance Report—Class



The third report is geared toward administrators (see Figure 20):

- **State Performance Report—District**: Provides a high-level performance view during the specified reporting period for each state performance level.

Figure 20: State Performance Report—District



Renaissance Place

STAR Math™ State Performance Report - District
Florida FCAT 2.0
Printed Thursday, January 23, 2014 3:22:12 PM

1 of 2

District: Renaissance School District

Last Consolidated: 1/22/2014 12:00:01 AM
Reporting Period: 09/04/2013 - 04/29/2014

Report Options
Reporting Parameter Group: All Demographics [Default]
Reporting Level: District
Group By: School

Oakwood Elementary School

	Student Performance Ou					
	Less Than Proficient					
	Level 1		Level 2		Level 3	
Grade	Total	%	Total	%	Total	%
3	35	26	31	23	32	23
4	33	24	24	18	42	31
5	35	22	38	24	39	24
Summary	103	24	93	22	113	26

Pine Hill Middle School

	Student Performance Ou					
	Less Than Proficient					
	Level 1		Level 2		Level 3	
Grade	Total	%	Total	%	Total	%
6	32	19	52	31	37	22
7	28	16	56	33	36	21
8	26	15	55	31	40	23
Summary	86	17	163	32	113	22

The Student Performance Outlook is based on research that links STAR tests to the FCAT 2.0. Our data was
For guidance interpreting data when the state test changes, see Interpreting Performance Reports under ST
Only data from schools using the Enterprise model is included.

Renaissance Place

STAR Math™ State Performance Report - District
Florida FCAT 2.0
Printed Thursday, January 23, 2014 3:22:12 PM

2 of 2

District: Renaissance School District

Last Consolidated: 1/22/2014 12:00:01 AM
Reporting Period: 09/04/2013 - 04/29/2014

Prairie Ridge High School

	Student Performance Outlook										STAR Math Participation 09/04/2013 - 04/29/2014			
	Less Than Proficient				Proficient									
	Level 1		Level 2		Level 3		Level 4		Level 5		Tested		Not Tested	
Grade	Total	%	Total	%	Total	%	Total	%	Total	%	Total	%	Total	%
9	19	18	37	35	23	21	18	17	10	9	107	95	6	5
Summary	19	18	37	35	23	21	18	17	10	9	107	95	6	5

Report Summary

	Student Performance Outlook										STAR Math Participation 09/04/2013 - 04/29/2014			
	Less Than Proficient				Proficient									
	Level 1		Level 2		Level 3		Level 4		Level 5		Tested		Not Tested	
	Total	%	Total	%	Total	%	Total	%	Total	%	Total	%	Total	%
Summary	208	20	293	28	249	24	180	17	126	12	1,056	95	57	5

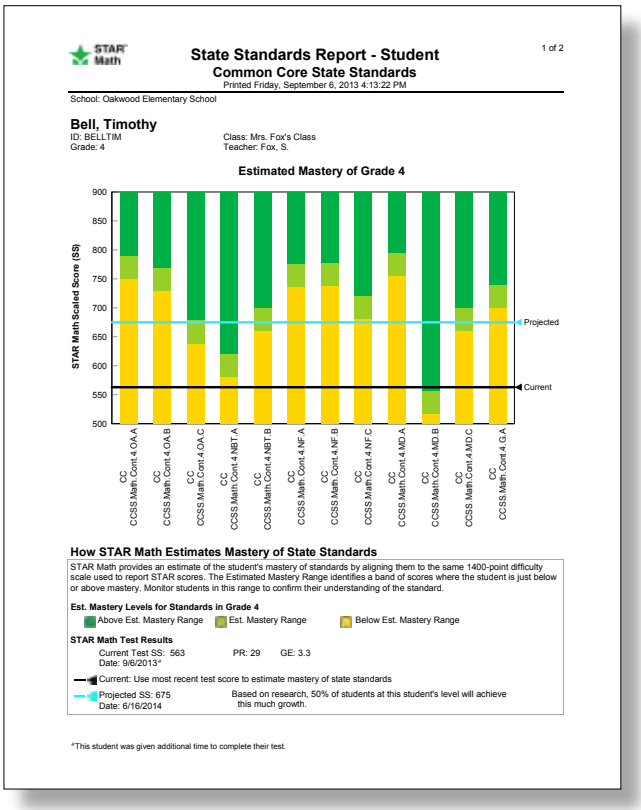
Standards alignment and reporting with the Common Core and other state standards

The Renaissance Learning standards team actively follows best practices in standards research and alignment, as well as maintains ongoing relationships in research and consultation with leading educational organizations, such as Mid-continent Research for Education and Learning (McREL) and the Northwest Regional Educational Laboratory (NWREL). This team rigorously developed, tested, and validated the original Core Progress learning progressions, and in 2013, they fully immersed themselves in the Common Core State Standards, as well as literature, resources, interpretations, and implementation information surrounding them to create another set of learning progressions built specifically for the Common Core.

STAR State Standards Reports (see example, Figure 21, next page), generated by the STAR Enterprise software, help educators estimate a student, class, or district's level of mastery on the Common Core State Standards or individual state standards (for those states that have not adopted the CCSS). To develop these reports, the standards team used both empirical data and content-area expert review, similar to the method used by states to place their standards on state test scales. Standards were aligned with the STAR scale using the following rigorous five-step approach:

1. Identify standards
2. Identify STAR skills and the items for those skills that assess the standard's expectations for the skills and concepts.
3. Review the calibrated (research-based) difficulty level of STAR items associated with the skills and concepts embedded in the standard.
4. Assign a difficulty level to the standard based on the review of empirical data. Equate the difficulty level to a scaled score on the STAR scale.
5. Review of assigned STAR scaled score by a content-area expert who analyzes the assigned score in relation to the composite standard to ensure the placement is accurate and appropriate.

Figure 21: State Standards Report—Student (Common Core State Standards)



High stakes purposes

Educators use assessments for different purposes. Some assessments can only be used for a single purpose, while others, such as STAR Enterprise, can meet various needs. Many of the uses of STAR described in this document are instructional—helping teachers understand what students know and what they are ready to learn next, how much they are growing, or whether they are responding adequately to instruction. Yet as educators well know, states and districts have been using assessment results for other, higher stakes decisions. STAR Enterprise assessments are approved by many states and districts for such purposes, typically as one of multiple measures or data points. These purposes include, but are not limited to, the following:

- Serving as an indicator of student growth in educator evaluation formulas
- Grade promotion
- Gifted & Talented identification

Uses of STAR Assessments for these purposes depend on specific state and district policies, but one commonality among them is that they demand assessments show evidence of strong technical adequacy, including reliability, validity, and predictive accuracy. The fact that STAR Assessments are often approved for these uses provides further reinforcement that the assessments meet high technical standards.

Test Content

Large item banks

STAR Assessments have large item banks to allow multiple administrations without risk of item overexposure. The STAR Early Literacy Enterprise item bank contains more than 2,500 items, while the STAR Reading Enterprise and STAR Math Enterprise items banks each number more than 5,000 items. Renaissance Learning continually develops new high-quality assessment items that are added to the banks to support frequent testing and to achieve an even distribution of items across the difficulty levels of each STAR assessment.

STAR Enterprise assessments are fixed-length tests, which mean item count is the sole criterion for ending an administration. STAR Early Literacy Enterprise administers 27 items per test event, and STAR Reading Enterprise and STAR Math Enterprise each administer 34 items. The tests were developed to provide precise measurement of student achievement in early literacy (and early numeracy), reading, and math, and to do so efficiently. Because the assessments are computer adaptive tests (CATs), they save teachers time by automating administration and scoring. Even more importantly, CATs allow students to be assessed on a larger and more varied range of skills using fewer items, which results in students spending less time completing the assessment (for more information, see Test Design, p. 7).

CATs allow students to be assessed on a larger and more varied range of skills using fewer items, which results in students spending less time completing the assessment.

Multiple-choice format

Renaissance Learning examined, researched, discussed, and prototyped several item-response formats and ultimately chose to use multiple-choice test items. Much research supports the use of this item type, also referred to as selected-response format. As noted by Stiggins (2005):

> [Selected-response] tests are efficient in that we can administer large numbers of multiple-choice or true/false test items per unit of testing time. Thus, they permit us to sample widely and draw relatively confident generalizations from the content sampled. For this reason, when the target is knowledge mastery, selected-response formats fit nicely into the resource realities of most classrooms. (p. 70)

The multiple-choice format lends itself well to computerized scoring, which automates the testing process and saves teachers time in collecting and scoring results (Nicol, 2007). A large number of multiple-choice test items can be administered in a short amount of time, and a key factor in the measurement precision of any test is the number of items each student must answer. According to Haladyna and Downing (1989), “the use of multiple-choice formats generally leads to more content-valid test score interpretations.”

Renaissance Learning constructs multiple-choice items to represent a balanced range of cognitive complexity. Item specifications require verifying the accuracy of all content; using grade-level-appropriate cognitive load, vocabulary, syntax, and readability; including only essential text and graphics to avoid wordiness and visual clutter; and employing standards for bias, fairness, and sensitivity.

Research has shown that well-designed multiple-choice questions can assess an array of skills (Cassels & Johnstone, 1984; Popham, 2008; Russell, Fischer, Fischer, & Premo, 2003) at higher levels of student learning (Cox, 1976; Johnstone & Arnbusaidi, 2000; Mattimore, 2009; Osterlind, 1998; Popham, 2003).

Item-development process

Item development is of critical concern to Renaissance Learning. The care in developing items is reflected in the high ratings STAR Assessments have garnered from several key federal groups, such as the National Center on Intensive Intervention, the National Center on Response to Intervention, and then National Center on Student Progress Monitoring (for more information, see Psychometric Properties, p. 19).

Professional designers, writers, and editors—with education backgrounds and content-area expertise—develop all content for Renaissance Learning products, including STAR Enterprise assessments. These experts follow research-based practices for developing assessment items, and rigorously adhere to the following process to ensure quality item creation:

1. Analyze standards to be assessed in the categories of skill, action, vocabulary, and context; refer to national or state resources for appropriate standard and grade-level expectation interpretation.
2. Write item specifications and provide training on their use to item writers and editors.
3. Establish item metadata to guide development, including standards-related and item-related data.
4. Use a multistep, recursive writing and editing process that ensures adherence to specifications and alignment to standards and item metadata.
5. Post items for calibration and acquire student-response data dynamic calibration (see below).
6. Examine psychometricians' analyses of item-testing results.
7. Add successful items to the operational assessment item bank.

Experts also receive ongoing item-writing training, which includes bias-and-fairness criteria to avoid stereotypes and characterizations of people or events that could be construed as demeaning, patronizing, or otherwise insensitive. Content-development tools track and report attributes such as gender, age, ethnicity, subject matter, and regional references. Individual attributes, as well as the intersection of multiple attributes, are tracked throughout the development process to ensure that final content is demographically balanced and free of bias.

In addition, 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.

Rules for item retention

Following these steps, all information pertaining to each test item—including traditional- and 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 any items for elimination, the recommendations are combined and a second review is conducted to resolve any issues.

Dynamic calibration

To maintain and update the large item banks for each STAR assessment, Renaissance Learning continually develops and calibrates new test items using a special feature called dynamic calibration. Each new STAR assessment item goes through calibration to determine its exact point on the STAR difficulty scale.

In dynamic calibration, one or more new items are embedded at random points in a STAR test. The items are administered to large samples of students, so that Renaissance Learning psychometricians can collect student-response and other data on the item, and then perform a statistical analysis of the response data to determine the scale values.

These items do not count toward students' scores on the STAR assessment. Students, on average, receive two or three additional items per test when calibration is turned on, and testing time is increased by approximately one minute. Norming, reliability, and validity studies take place after items successfully pass through calibration.

Appendix: STAR Assessments™ Score Definitions

STAR Early Literacy Enterprise™ scores

Literacy Classifications are the stages of literacy development measured in STAR Early Literacy and associated with scaled scores. They are an easy way to monitor student progress:

Emergent Reader (300–674):

An *Early Emergent Reader (300–487)* is beginning to understand that printed text has meaning. The student is learning that reading involves printed words and sentences and that print flows from left to right and from top to bottom of a page. The student is also beginning to identify colors, shapes, numbers, and letters.

A *Late Emergent Reader (488–674)* can identify most of the letters of the alphabet and match most of the letters to sounds. The student is beginning to "read" picture books and familiar words around home. Through repeated reading of favorite books with an adult, a student at this stage is building vocabulary, listening skills, and understanding of print.

A **Transitional Reader (675–774**) has mastered alphabet skills and letter-sound relationships. The student can identify many beginning and ending consonant sounds as well as long and short vowel sounds. The student is probably able to blend sounds and word parts to read simple words and is likely using a variety of strategies to figure out words, such as pictures, story patterns, and phonics.

A **Probable Reader (775–900)** is becoming proficient at recognizing many words, both in and out of context, and spends less time identifying and sounding out words and more time understanding what was read. A probable reader can blend sounds and word parts to read words and sentences more quickly, smoothly, and independently than students in other stages of development.

Literacy Domain Score, ranging from 0–100, is criterion-referenced and represents the percentage of items a student would be expected to answer correctly within the assessment's domains, which include key early literacy sub-domains comprised of skill sets.

Sub-Domain and **Skill Set Scores**, ranging from 0–100, are criterion-referenced and represent the percent of mastery of specific skills within the assessment's domains, sub-domains, and skill sets.

Estimated Oral Reading Fluency (Est. ORF), reported in correct words per minute, is an estimate of a student's ability to read words quickly and accurately in order to comprehend text efficiently. Students with oral reading fluency demonstrate accurate decoding, automatic word recognition, and appropriate use of the rhythmic aspects of language (e.g., intonation, phrasing, pitch, emphasis). Est. ORF is based on a known relationship between STAR Early Literacy Enterprise performance and oral reading fluency and is reported for grades 1–4.

Growth Norms characterize typical student growth within a given grade and achievement level.

Scaled Score (SS) is useful in comparing student performance over time and in identifying student performance in relation to a vertical scale and all criteria associated with that scale. Because the same range is used for all students, scaled scores are also useful for comparing student performance across grade levels. STAR Early Literacy Enterprise scaled scores range from 300–900.

Student Growth Percentile (SGP) is a measure of growth between a pre- and posttest, relative to the growth made by other students in the same grade with the same pretest score. It is a simple and effective way for educators to interpret student growth rate relative to that of his or her academic peers nationwide. SGPs, which were derived from growth norms, range from 1–99, with lower numbers representing lower relative growth and high numbers representing higher relative growth.

STAR Reading Enterprise™ scores

ATOS 2000 is the STAR scaled score converted to Renaissance Learning's 2000-point scale, based on an extensive research study correlating STAR to the Lexile scale. While it is not a Lexile score, it is intended to provide a score that can be used in place of a Lexile score as a close approximation.

Domain and **Skill Set Scores**, ranging from 0–100, are criterion-referenced and estimate a student's percent of mastery of specific skills within the assessment's domains and skill sets.

Estimated Oral Reading Fluency (Est. ORF), reported in correct words per minute, is an estimate of a student's ability to read words quickly and accurately in order to comprehend text efficiently. Students with oral reading fluency demonstrate accurate decoding, automatic word recognition, and appropriate use of the rhythmic aspects of language (e.g., intonation, phrasing, pitch, emphasis). Est. ORF is based on a known relationship between STAR Reading Enterprise performance and oral reading fluency and is reported for grades 1–4.

Grade Equivalent (GE) score, ranging 0.0–12.9+, is norm-referenced and represents how a student's test performance compares with other students nationally. For example, a fifth-grade student with a GE of 7.6 performed as well as a typical seventh-grader in the sixth month of the school year. This does not mean the student is necessarily capable of reading seventh-grade material—rather, it indicates that the student's reading skills are well above average for fifth grade.

Growth Norms characterize typical student growth within a given grade and achievement level.

Instructional Reading Level (IRL) is a criterion-referenced score that is the highest reading level at which a student is 80% proficient (or higher) at comprehending material with assistance (Gickling & Thompson, 2001). Research has found that this level of comprehension corresponds to being at least 90–98% proficient at recognizing words (Gickling & Havertape, 1981; Johnson, Kress, & Pikulski, 1987; McCormick, 1999).[7] IRL scores are PP (Pre-Primer), P (Primer, grades 0.1–0.9), grades 1.0 through 12.9, and PHS (Post-High School, grades 13.0+).

Normal Curve Equivalent (NCE) score, ranging from 1–99, is norm-referenced and similar to the percentile rank score but based on an equal interval scale. This means the difference between any two successive scores on the NCE scale has the same meaning throughout the scale. Mostly used for research, NCEs are useful in making comparisons between different achievement tests and in statistical computations—for example, determining an average score for a group of students.

Percentile Rank (PR) score, ranging from 1–99, is norm-referenced and provides the best measure of a student's reading achievement level compared to other students in the same grade nationally. The score indicates the percentage of a student's peers whose scores were equal to or lower than the score of that student—for example, a student with a PR score of 85 performed as well as or better than 85 percent of students in the same grade.

Scaled Score (SS) is useful in comparing student performance over time and in identifying student performance in relation to a vertical scale and all criteria and norms associated with that scale. Because the same range is used for all students, scaled scores are also useful for comparing student performance across grade levels. STAR Reading Enterprise scaled scores range from 0–1400.

Student Growth Percentile (SGP) is a measure of growth between a pre- and posttest relative to the growth made by other students in the same grade with the same pretest score. It is a simple and effective way for educators to interpret student growth rate relative to that of his or her academic peers nationwide. SGPs, which were derived from growth norms, range from 1–99, with lower numbers representing lower relative growth and high numbers representing higher relative growth.

Zone of Proximal Development (ZPD) is an individualized range of readability levels based on a student's results from a STAR Reading Enterprise assessment. Books students choose to read within their ZPD range will be neither too difficult nor too easy and should allow students to experience optimal growth.

STAR Math Enterprise™ scores

Accelerated Math Library Recommendation helps educators place a student in the Accelerated Math library that will be of the most benefit, based on that student's achievement level per the results of a STAR Math Enterprise assessment.

Algebra Readiness Indicator is based solely on skills associated with algebra readiness. The math concepts and skills learned in elementary through middle school provide the foundation for high school level algebra. The Student Instructional Planning Report in STAR Math Enterprise provides an Algebra Readiness Indicator to help teachers identify student progress through these foundational skills to ensure the student is on track to be ready for algebra.

Domain and **Skill Set Scores**, ranging from 0–100, are criterion-referenced and estimate a student's percentage of mastery of specific skills within the assessment's domains and skill sets.

Grade Equivalent (GE) score, ranging 0.0–12.9+, is norm-referenced and represents how a student's test performance compares with other students nationally. For example, a fifth-grade student with a GE of 7.6 performed as well as a typical seventh-grader in the sixth month of the school year. This does not mean the student is necessarily capable of doing seventh-grade math—rather, it indicates that the student's math skills are well above average for fifth grade.

Growth Norms characterize typical student growth within a given grade and achievement level.

Normal Curve Equivalent (NCE) score, ranging from 1–99, is norm-referenced and similar to the percentile rank score but based on an equal interval scale. This means the difference between any two successive scores on the NCE scale has the same meaning throughout the scale. Mostly used for research, NCEs are useful in making comparisons between different achievement tests and in statistical computations—for example, determining an average score for a group of students.

---

[7] Gickling, E. E., & Havertape, S. (1981). *Curriculum-based assessment (CBA).* Minneapolis, MN: School Psychology Inservice Training Network.

Gickling, E. E., & Thompson, V. E. (2001). Putting the learning needs of children first. In B. Sornson (Ed.). *Preventing early learning failure.* Alexandria, VA: ASCD.

Johnson, M. S., Kress, R. A., & Pikulski, J. J. (1987). *Informal reading inventories.* Newark, DE: International Reading Association.

McCormick, S. (1999). *Instructing students who have literacy problems* (3rd Ed.). Englewood Cliffs, NJ: Prentice-Hall.

Percentile Rank (PR) score, ranging from 1–99, is norm-referenced and provides the best measure of a student's math achievement level compared to other students in the same grade nationally. The score indicates the percentage of a student's peers whose scores were equal to or lower than the score of that student—for example, a student with a PR score of 85 performed as well as or better than 85 percent of students in the same grade.

Scaled Score (SS) is useful in comparing student performance over time and in identifying student performance in relation to a vertical scale and all criteria and norms associated with that scale. Because the same range is used for all students, scaled scores are also useful for comparing student performance across grade levels. STAR Math Enterprise scaled scores range from 0–1400.

Student Growth Percentile (SGP) is a measure of growth between a pre- and posttest relative to the growth made by other students in the same grade with the same pretest score. It is a simple and effective way for educators to interpret student growth rate relative to that of his or her academic peers nationwide. SGPs, which were derived from growth norms, range from 1–99, with lower numbers representing lower relative growth and high numbers representing higher relative growth.

References

Bray, M. A. & Kehle, T. J., (Eds.) (2011). *The Oxford handbook of school psychology.* New York: Oxford University Press, Inc.

Cassels, J. R. T., & Johnstone, A. H. (1984). The effect of language on student performance on multiple choice tests in chemistry. *Journal of Chemistry Education, 61*, 613–615.

Cox, K. R. (1976). How did you guess? Or what do multiple choice questions measure? *Medical Journal of Australia, 1*, 884–886.

Haladyna, T. M., & Downing, S. M. (1989). The validity of a taxonomy of multiple-choice item-writing rules. *Applied Measurement in Education, 1*, 51–78.

Heritage, M. (2008). *Learning progressions: Supporting instruction and formative assessment.* Washington, DC: Council of Chief School Officers. Retrieved from http://www.ccsso.org/Documents/2008/Learning_Progressions_Supporting_2008.pdf

Johnson, M. J., & Weiss, D. J. (1980). Parallel forms reliability and measurement accuracy comparison of adaptive and conventional testing strategies. In D. J. Weiss (Ed.), *Proceedings of the 1979 Computerized Adaptive Testing Conference* (pp. 16–34). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, Computerized Adaptive Testing Laboratory. Retrieved from http://www.psych.umn.edu/psylabs/catcentral/pdf%20files/jo80-01.pdf

Johnstone, A. H., & Ambusaidi, A. (2000). Fixed response: What are we testing? *Chemistry Education: Research and Practice in Europe, 1*(3), 323–328.

Kingsbury, G. G., & Weiss, D. J. (1980). *An alternate-forms reliability and concurrent validity comparison of Bayesian adaptive and conventional ability tests* (Research Report 80-5). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, Computerized Adaptive Testing Laboratory. Retrieved from http://iacat.org/sites/default/files/biblio/ki80-05.pdf

Lord, F. M. (1980). *Applications of item response theory to practical testing problems* (pp. 158–159). Hillsdale, NJ: Lawrence Erlbaum Associates.

Mardberg, B., & Carlstedt, B. (1998). Swedish Enlistment Battery: Construct validity and latent variable estimation of cognitive abilities by the CAT-SEB. *International Journal of Selection and Assessment, 6*, 107–114. Retrieved from http://onlinelibrary.wiley.com/doi/10.1111/1468-2389.00079/abstract

Mattimore, P. (2009, February 5). *Why our children need national multiple choice tests.* Retrieved from http://www.opednews.com/articles/Why-Our-Children-Need-Nati-by-Patrick-Mattimore-090205-402.html

McBride, J., & Martin, J. T. (1983). Reliability and validity of adaptive ability tests. In D. J. Weiss (Ed.), *New horizons in testing: Latent trait test theory and computerized adaptive testing* (Chapter 11, pp. 224–225). New York: Academic Press.

Moreno, K. E., & Segall, O. D. (1997). Reliability and construct validity of CAT-ASVAB. In W. A. Sands, B. K. Waters, & J. R. McBride (Eds.), *Computerized adaptive testing: From inquiry to operation* (pp. 169–179). Washington, DC: American Psychological Association. Retrieved from http://psycnet.apa.org/index.cfm?fa=buy.optionToBuy&id=1997-36257-018

Nicol, D. (2007). E-assessment by design: Using multiple-choice tests to good effect. *Journal of Further and Higher Education, 31*(1), 53–64.

Osterlind, S. J. (1998). *Constructing test items: Multiple-choice, constructed-response, performance, and other formats* (2nd ed.). New York: Kluwer.

Popham, W. J. (2003). *Test better, teach better: The instructional role of assessment.* Alexandria, VA: Association for Supervision and Curriculum Development.

Popham, W. J. (2008). *Classroom assessment: What teachers need to know* (5th Ed.). Boston: Allyn and Bacon.

Renaissance Learning. (2013a). *STAR Early Literacy Enterprise: Technical manual.* Wisconsin Rapids, WI: Author. Available by request to research@renlearn.com

Renaissance Learning. (2013b). *STAR Math: Technical manual.* Wisconsin Rapids, WI: Author. Available by request to research@renlearn.com

Renaissance Learning. (2013c). *STAR Reading: Technical manual.* Wisconsin Rapids, WI: Author. Available by request to research@renlearn.com

Russell, M., Fischer, M. J., Fischer, C. M., & Premo, K. (2003). Exam question sequencing effects on marketing and management sciences student performance. *Journal of Advancement of Marketing Education, 3*, 1–10.

Shapiro, E. S. (2012). *New thinking in response to intervention: A comparison of computer-adaptive tests and curriculum-based measurement within RTI.* Wisconsin Rapids, WI: Renaissance Learning, Inc.

Shapiro, E. S., & Gebhardt, S. N. (2012). Comparing computer-adaptive and curriculum-based measurement methods of assessment. *School Psychology Review, 41*(3), 295-305.

Stiggins, R. J. (2005). *Student-involved classroom assessment for learning* (4th ed.). Upper Saddle River, New Jersey: Pearson/Merrill Prentice Hall.

U.S. Department of Education: National Center on Intensive Intervention. (2012a). *Review of progress-monitoring tools* [Review of STAR Early Literacy]. Washington, DC: Author. Available online from http://www.intensiveintervention.org/chart/progress-monitoring

U.S. Department of Education: National Center on Intensive Intervention. (2012b). *Review of progress-monitoring tools* [Review of STAR Math]. Washington, DC: Author. Available online from http://www.intensiveintervention.org/chart/progress-monitoring

U.S. Department of Education: National Center on Intensive Intervention. (2012c). *Review of progress-monitoring tools* [Review of STAR Reading]. Washington, DC: Author. Available online from http://www.intensiveintervention.org/chart/progress-monitoring

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

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

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

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

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

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

U.S. Department of Education: National Center on Student Progress Monitoring. (2006a). *Review of progress monitoring tools* [Review of STAR Early Literacy]. Washington, DC: Author. Available online from http://doc.renlearn.com/kmnet/R004100821GG6E03.pdf

U.S. Department of Education: National Center on Student Progress Monitoring. (2006b). *Review of progress monitoring tools* [Review of STAR Math]. Washington, DC: Author. Available online from http://doc.renlearn.com/kmnet/R004100821GG6E03.pdf

U.S. Department of Education: National Center on Student Progress Monitoring. (2006c). *Review of progress monitoring tools* [Review of STAR Reading]. Washington, DC: Author. Available online from http://doc.renlearn.com/kmnet/R004100821GG6E03.pdf

Weiss, D. J. (2004). Computerized adaptive testing for effective and efficient measurement in counseling and education. *Measurement and Evaluation in Counseling and Development, 37*, 70–84. Retrieved from http://www.psych.umn.edu/psylabs/catcentral/pdf%20files/we04070.pdf

Acknowledgements

The following experts have advised Renaissance Learning in the development of STAR Assessments.

Catherine N. Close, Ph.D., is a psychometrician for Renaissance Learning. She has extensive background in both classical and modern test theories and their application to computerized adaptive tests (CAT) and conventional non-adaptive tests. At Renaissance Learning, Close researches, develops, and evaluates the STAR Enterprise assessments. She is responsible for ensuring that the assessments meet professional standards for reliability and validity for use in grades K–12. Close has co-authored research articles and published in peer-reviewed journals of measurement.

Margaret Heritage, Ph.D., is assistant director for professional development at the National Center for Research on Evaluation, Standards, and Student Testing (CRESST) at UCLA. Her current work focuses on the development of academic language for EL students and formative assessment, including teachers' use of formative assessment evidence. She has made numerous presentations on these topics all over the United States, in Europe, and in India.

Karin Hess, Ed.D., Senior Associate with the Center for Assessment (NCIEA) since 2002, brings to the Center's work over 30 years of deep experience in curriculum, instruction, and assessment. Hess is recognized nationally for her research and work with learning progressions, text complexity, performance assessment, and cognitive rigor. In addition to experiences as a classroom teacher and school administrator, she has been a program evaluator for the Vermont Mathematics Project, a content specialist for development of the Vermont Science assessment, and a developer and editor of Science Exemplars K-8 performance assessments. Hess is the principal author of the content specifications for assessment of the CCSS ELA and Literacy standards and was a contributor to the mathematics content specifications for the *Smarter Balanced Assessment Consortium*.

Thomas P. Hogan, Ph.D., is a professor of psychology and a Distinguished University Fellow at the University of Scranton. He has more than 40 years of experience conducting reviews of mathematics curricular content, principally in connection with the preparation of a wide variety of educational tests, including the Stanford Diagnostic Mathematics Test, Stanford Modern Mathematics Test, and the Metropolitan Achievement Test. Hogan has published articles in the *Journal for Research in Mathematics Education* and *Mathematical Thinking and Learning*, and he has authored two textbooks and more than 100 scholarly publications in the areas of measurement and evaluation. He has also served as consultant to a wide variety of school systems, states, and other organizations on matters of educational assessment, program evaluation, and research design.

James R. McBride, Ph.D., is vice president and chief psychometrician for Renaissance Learning. He was a leader of the pioneering work related to computerized adaptive testing (CAT) conducted by the Department of Defense. McBride has been instrumental in the practical application of item response theory (IRT) and since 1976 has conducted test development and personnel research for a variety of organizations. At Renaissance Learning, he has contributed to the psychometric research and development of STAR Math, STAR Reading, and STAR Early Literacy. McBride is co-editor of a leading book on the development of CAT and has authored numerous journal articles, professional papers, book chapters, and technical reports.

R. James Milgram, Ph.D., is a professor of mathematics at Stanford University. His work in mathematics education includes consulting with several states on math standards, including California. Milgram has given lectures around the world and is a member of numerous boards and committees, including the National Board of Education Sciences, created by the Education Sciences Reform Act of 2002 "to advise and consult with the Director of the Institute of Education Sciences (IES) on agency policies," and the Human Capital Committee of the NASA Advisory Council, which "provides the NASA Administrator with counsel and advice on programs and issues of importance to the Agency." Milgram is author of "An Evaluation of CMP," "A Preliminary Analysis of SAT-I Mathematics Data for IMP Schools in California," and "Outcomes Analysis for Core Plus Students at Andover High School: One Year Later." Each of these papers identifies serious shortcomings in popular mathematics programs.

Michael Milone, Ph.D., is a research psychologist and an award-winning educational writer and consultant to publishers and school districts. He earned a Ph.D. in 1978 from The Ohio State University and has served in an adjunct capacity at Ohio State, the University of Arizona, Gallaudet University, and New Mexico State University. He has taught in regular and special education programs at all levels, holds a Master of Arts degree from Gallaudet University, and is fluent in American Sign Language. Milone served on the board of directors of the Association of Educational Publishers and was a member of the Literacy Assessment Committee and a past chair of the Technology and Literacy Committee of the International Reading Association. He has contributed to both readingonline.org and *Technology & Learning* magazine on a regular basis. Over the past 30 years, he has been involved in a broad range of publishing projects, including the SRA reading series, assessments developed for Academic Therapy Publications, and software published by The Learning Company and LeapFrog. He has completed 34 marathons and 2 Ironman races.

James B. Olsen, Ph.D., is a psychometrician for Renaissance Learning. He has professional skills in educational and professional assessment, job analysis, test design and development, psychometric and statistical analysis, research and evaluation, and professional certification and accreditation. He serves on the Technical Advisory Committees (TAC) for the Graduate Management Admissions Council, the Comprehensive Adult Student Assessment System and the Utah State Office of Education. He has served as a consultant for the National Assessment Governing Board, National Assessment of Educational Progress, and American National Standards Institute. He was principal investigator and program manager for multiple assessment projects with the U.S. Department of Defense, National Science Foundation, and U.S. Department of Education. For the Association of Test Publishers (ATP), he has served for two terms on the Board of Directors and for five years as Innovations Chair for the annual Innovations in Testing Conference.

Sharif M. Shakrani, Ph.D., is a private consultant and researcher specializing in measurement and quantitative methods. Shakrani is a former co-director of the Education Policy Center at Michigan State University and professor of measurement and quantitative methods in the Department of Counseling, Educational Psychology and Special Education. Before coming to Michigan State University, Shakrani served 8 years as the deputy executive director of the National Assessment Governing Board in the U.S. Department of Education. He was responsible for technical and policy direction for the National Assessment of Educational Programs (NAEP). He has also worked for the National Center for Education Statistics in the U.S. Department of Education where he guided the design and analysis of federal educational assessments. In his work in the Michigan Department of Education, Shakrani was responsible for K–12 general curriculum and assessment and was instrumental in revising the Michigan Educational Assessment Program (MEAP).

Amanda M. VanDerHeyden, Ph.D., is a private consultant and researcher, living in Fairhope, Alabama, who has previously held faculty positions and has worked as a researcher, consultant, and national trainer in a number of school districts. In 2006, VanDerHeyden was named to a National Center for Learning Disabilities advisory panel to provide guidance related to RTI and the diagnosis of specific learning disability. She is associate editor of *Assessment for Effective Intervention*, serves on the editorial boards of several journals including *School Psychology Quarterly* and *Journal of School Psychology*, and has recently co-authored *Essentials of Response to Intervention* (with Dr. Matthew Burns). VanDerHeyden received the 2006 Lightner Witmer Early Career Contributions Award from the APA for her scholarship on early intervention, RTI, and models of data-based decision making. She serves as research advisor to iSTEEP, has published measures of early numeracy for preschool and kindergarten children, and currently directs research projects in early numeracy with preschool/kindergarten children as well as evaluation of Tier 2 mathematics intervention.

James Ysseldyke, Ph.D., is Emma Birkmaier Professor of Educational Leadership in the Department of Educational Psychology at the University of Minnesota. He has been educating school psychologists and researchers for more than 35 years. Ysseldyke has served the University of Minnesota as director of the Minnesota Institute for Research on Learning Disabilities, director of the National School Psychology Network, director of the National Center on Educational Outcomes, director of the School Psychology Program, and associate dean for research. His research and writing have focused on enhancing the competence of individual students and enhancing the capacity of systems to meet students' needs. He is an author of major textbooks and more than 300 journal articles. Ysseldyke is conducting a set of investigations on the use of technology-enhanced progress-monitoring systems to track the performance and progress of students in urban environments. He chaired the task forces that produced the three *Blueprints on the Future of Training and Practice in School Psychology*, and he is former editor of *Exceptional Children*, the flagship journal of the Council for Exceptional Children. Ysseldyke has received awards for his research from the School Psychology Division of the American Psychological Association, the American Educational Research Association, and the Council for Exceptional Children. The University of Minnesota presented him a distinguished teaching award, and he received a distinguished alumni award from the University of Illinois.

Acknowledgements

The following experts have advised Renaissance Learning in the development of STAR Assessments.

Catherine N. Close, Ph.D., is a psychometrician for Renaissance Learning.

Michael Milone, Ph.D., is a research psychologist and an award-winning educational writer and consultant to publishers and school districts.

Margaret Heritage, Ph.D., is assistant director for professional development at the National Center for Research on Evaluation, Standards, and Student Testing (CRESST) at UCLA.

James B. Olsen, Ph.D., is a psychometrician for Renaissance Learning.

Karin Hess, Ed.D., is a Senior Associate with the Center for Assessment (NCIEA).

Sharif M. Shakrani, Ph.D., is a private consultant and researcher specializing in measurement and quantitative methods.

Thomas P. Hogan, Ph.D., is a professor of psychology and a Distinguished University Fellow at the University of Scranton.

Amanda M. VanDerHeyden, Ph.D., is a private consultant and researcher living in Fairhope, Alabama.

James R. McBride, Ph.D., is vice president and chief psychometrician for Renaissance Learning.

James Ysseldyke, Ph.D., is Emma Birkmaier Professor of Educational Leadership in the Department of Educational Psychology at the University of Minnesota.

R. James Milgram, Ph.D., is a professor of mathematics at Stanford University.

RENAISSANCE™ LEARNING
Accelerati learni ll

L1526.0114.RN.5M
R14807

www.SuccessMaker.com

SuccessMaker® Reading Evidence of Effectiveness

A Summary of the Randomized, Control Trial Conducted by Gatti Evaluation, Inc.

Released November 2011

SuccessMaker Reading Summative Research Overview

Pearson Digital Learning strongly believes that its programs should be proven through scientific research to increase student achievement. As such, it contracted with independent research group Gatti Evaluation, Inc., to conduct a randomized, control trial of its *SuccessMaker Reading* program. The study was conducted in 3rd, 5th, and 7th grade classrooms over the 2010-11 school year. This report summary presents the evaluation design and methods, an assessment of program usage and implementation, student performance results, and a discussion of findings.

Study Design and Research Questions

The purpose of this study was twofold. The primary goal to conduct rigorous research to support the assertion that *the SuccessMaker Reading* program effectively increases students' English language arts achievement, specifically vocabulary, comprehension, and fluency, as well as academic attitudes. The second goal of the study was to collect information on teacher and student attitudes toward specific features and aspects of the *SuccessMaker* program. The study employed an experimental randomized, control trial research design. That is, teachers within each research school were *randomly assigned* to either use the *SuccessMaker Reading* program with their students (also referred to as the "treatment" group) or to refrain from using the *SuccessMaker Reading* program (also referred to as the "comparison" condition). Teachers assigned to the comparison condition did not regularly use a computer-based adaptive reading program.

The study addressed the following overarching evaluation questions:

1. Do students using the *SuccessMaker Reading* program demonstrate a significant improvement in achievement over their non-*SuccessMaker* counterparts?
2. Do students using the *SuccessMaker Reading* program demonstrate more positive attitudes toward reading and reading instruction when compared to non-*SuccessMaker* counterparts?
3. How did teachers and students react to the *SuccessMaker Reading* program?
4. How was the *SuccessMaker Reading* program implemented, and how are teachers using program reporting to monitor progress and inform instruction?

Participants and Setting

Gatti Evaluation recruited eight school districts to participate in the study, including schools in Arizona, California, Indiana, Kansas, Michigan, Missouri and Texas. The final analytic sample was comprised of 80 classrooms and 1,711 students. The study schools were members of public school districts located in suburban and urban-fringe areas. The study sample demonstrated considerable variation in ethnicity, socioeconomic status as evidenced by eligibility for free or reduced lunch status, and English language learner status, as well as a wide range of reading achievement levels as evidenced by previous year state reading assessment data. Figure 1 presents the full study sample demographics broken out by school.

Figure 1. SuccessMaker Reading Study Sample Demographic Information

District	Group	Grade	Student Count*	% One Grade Equivalent Below**	% Not English Proficient	% Reduced Lunch	% Caucasian	% Hispanic/ Native American	% African American/ Caribbean	Other Ethnicity or No Information
Arizona District 1	SM	3	55 (97%)	15%	0%	27%	94%	4%	2%	0%
	Comparison		49 (91%)	18%	0%	16%	92%	4%	2%	2%
	SM	5	48 (84%)	6%	0%	35%	90%	4%	0%	6%
	Comparison		52 (88%)	8%	0%	35%	87%	12%	0%	1%
Arizona District 2	SM	7	58 (94%)	38%	3%	64%	40%	43%	14%	3%
	Comparison		31 (91%)	61%	19%	65%	29%	55%	13%	3%
California District	SM	3	48 (96%)	10%	4%	0%	50%	4%	2%	44%
	Comparison		25 (100%)	0%	0%	0%	40%	8%	0%	52%
	SM	5	54 (93%)	2%	0%	2%	46%	6%	0%	48%
	Comparison		55 (93%)	9%	0%	5%	42%	7%	5%	46%
Indiana District	SM	3	36 (100%)	19%	0%	47%	100%	0%	0%	0%
	Comparison		41 (100%)	27%	0%	24%	98%	0%	0%	2%
	SM	5	21 (96%)	38%	0%	43%	90%	0%	0%	10%
	Comparison		17 (85%)	29%	0%	47%	100%	0%	0%	0%
Kansas District	SM	3	36 (97%)	8%	0%	14%	81%	8%	3%	8%
	Comparison		37 (97%)	16%	0%	14%	89%	5%	0%	6%
	SM	5	45 (100%)	11%	0%	11%	87%	7%	0%	0%
	Comparison		24 (96%)	8%	0%	0%	96%	0%	0%	0%
	SM	7	34 (100%)	15%	0%	15%	85%	12%	3%	0%
	Comparison		34 (100%)	18%	0%	21%	88%	9%	0%	3%
Michigan District	SM	3	47 (89%)	62%	2%	74%	40%	2%	43%	15%
	Comparison		25 (93%)	56%	0%	60%	48%	0%	44%	8%
	SM	5	30 (100%)	43%	0%	87%	40%	0%	50%	10%
	Comparison		28 (93%)	46%	7%	67%	25%	0%	54%	21%
	SM	7	140 *83%)	33%	1%	67%	44%	1%	43%	12%
	Comparison		110 (73%)	41%	0%	69%	33%	2%	51%	14%

Figure 1. SuccessMaker Reading Study Sample Demographic Information (Cont'd.)

District	Group	Grade	Student Count*	% One Grade Equivalent Below**	% Not English Proficient	% Reduced Lunch	% Caucasian	% Hispanic/ Native American	% African American/ Caribbean	Other Ethnicity or No Information
Missouri District	SM	3	41 (95%)	12%	0%	0%	85%	5%	2%	8%
	Comparison		20 (95%)	10%	0%	0%	100%	0%	0%	0%
	SM	5	36 (92%)	8%	0%	0%	91%	3%	6%	0%
	Comparison		18 (90%)	0%	0%	0%	89%	0%	0%	11%
	SM	7	22 (88%)	5%	0%	0%	95%	0%	0%	5%
	Comparison		23 (93%)	0%	0%	0%	100%	0%	0%	0%
Texas District	SM	3	89 (87%)	35%	4%	69%	17%	67%	13%	3%
	Comparison		69 (92%)	23%	14%	78%	16%	68%	9%	7%
	SM	5	108 (83%)	41%	26%	87%	7%	88%	4%	1%
	Comparison		105 (91%)	23%	35%	83%	5%	89%	5%	1%

*Percents within parentheses next to student counts indicate the percent of students tested at baseline that were also tested at the end of the school year.

**Study sample was broken out by baseline GRADE national norm cutoff score for 1.0 grade equivalent below grade and month at the time of testing.

Measures

Multiple measures were used to assess student achievement, program implementation, and student attitudes.

Statistical analyses were performed on students' gain scores (i.e., end-of-year raw score minus beginning-of-year raw score) for the GRADE, AIMSweb and Reading Academic Attitude Survey at each grade level. Results compared the *SuccessMaker* users to the comparison group. Results were broken out and analyzed for separate levels of three key demographic variables (i.e., ethnicity, gender, meal status). Students' English language learner (i.e., ELL) status was also collected but results were not broken out and reported by ELL status because there were very few students designated as ELL (i.e., 5%).

Results were also calculated for a group of lower achieving students at each grade level. These students scored at baseline one grade equivalent below their current grade level. Further, the performance for the comparison group was compared to four blocks of program usage (i.e., block 1 = 1 to 9 hours, block 2 = 10 to 19 hours, block 3 = 20 to 29 hours, block 4 = 30 or more hours).

Students responded to self-report questions on a reading academic attitude survey regarding general reading attitude, confidence, motivation, and self-perceived aptitude. Further, students in *SuccessMaker* classrooms were surveyed as to their opinions on several aspects of the program.

In addition to the assessment battery, qualitative data collection methods, including program reports, teacher surveys, daily lesson logs, classroom

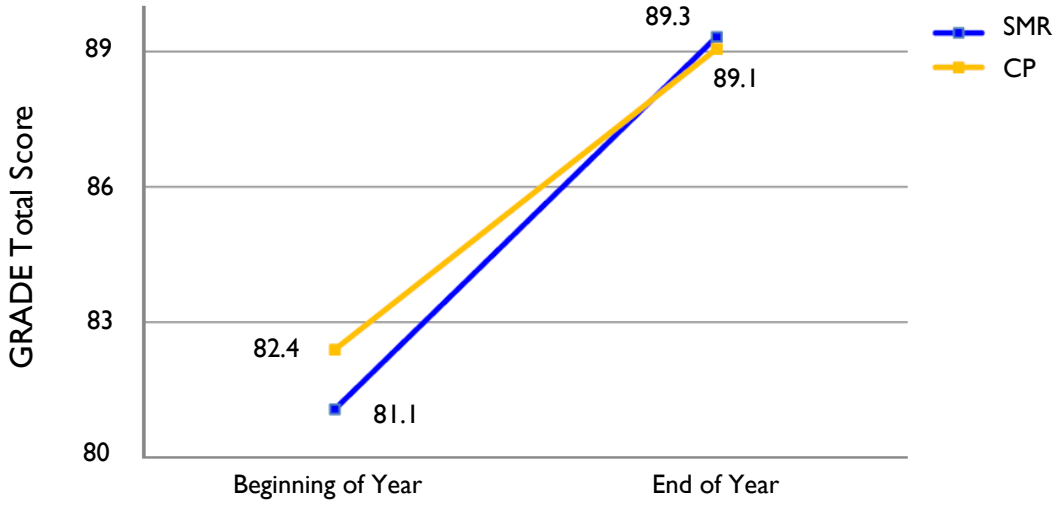
observations, as well as, teacher notes from electronic correspondences, were also employed. Teachers were routinely asked for their opinions throughout the school year. Weekly lesson notes were collected for both *SuccessMaker* and comparison classes. Cumulative usage reports and program implementation logs were regularly collected from *SuccessMaker* users. All study classrooms were observed twice during the school year teaching routine reading lessons and *SuccessMaker* teachers were further interviewed as to their opinions regarding the program. All this data was compiled and content analyzed to determine teacher attitudes and performance, as well as to illuminate the various ways teachers and students interact with the program.

Student Performance Results

Results for SuccessMaker Reading versus Comparison.

Evaluators conducted analyses to examine how *SuccessMaker Reading* students performed in comparison to students using print supplemental reading programs. Results showed positive effects of the *SuccessMaker Reading* program with program users statistically significantly outperforming the comparison group students on the GRADE in all three grade levels. In Figures 2 through 4, the mean gain scores on the GRADE for the three study grades is graphed after adjusting for differences in baseline student and classroom characteristics.

Figure 2. Third Grade Reading Achievement Gains



After adjusting for student & classroom characteristics, 3rd, 5th & 7th grade *SuccessMaker Math* users out scored their comparison group counterparts by 17.5% (SE=2.19%), 10.0% (SE=2.72%) and 9.8% (SE=2.23%) respectively.

Figure 3. Fifth Grade Reading Achievement Gains

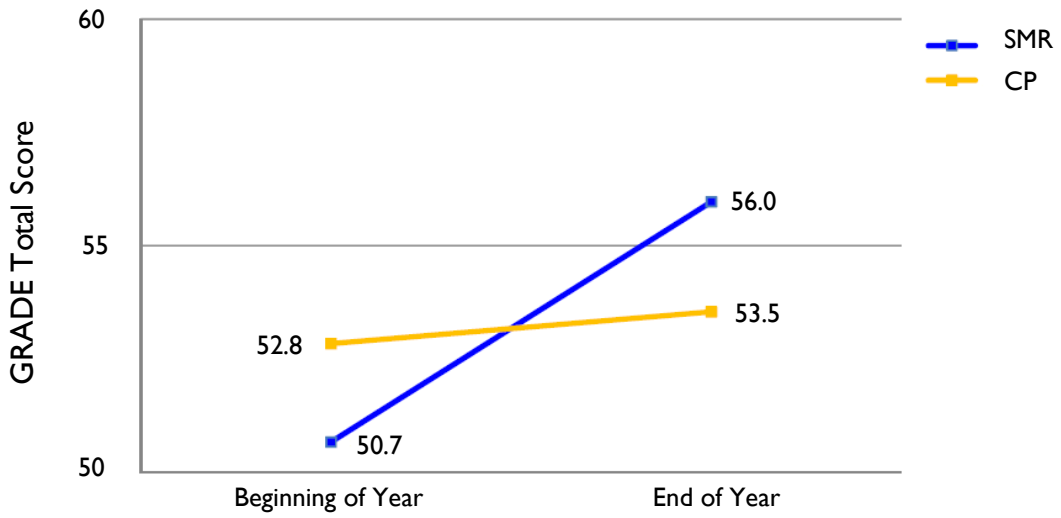
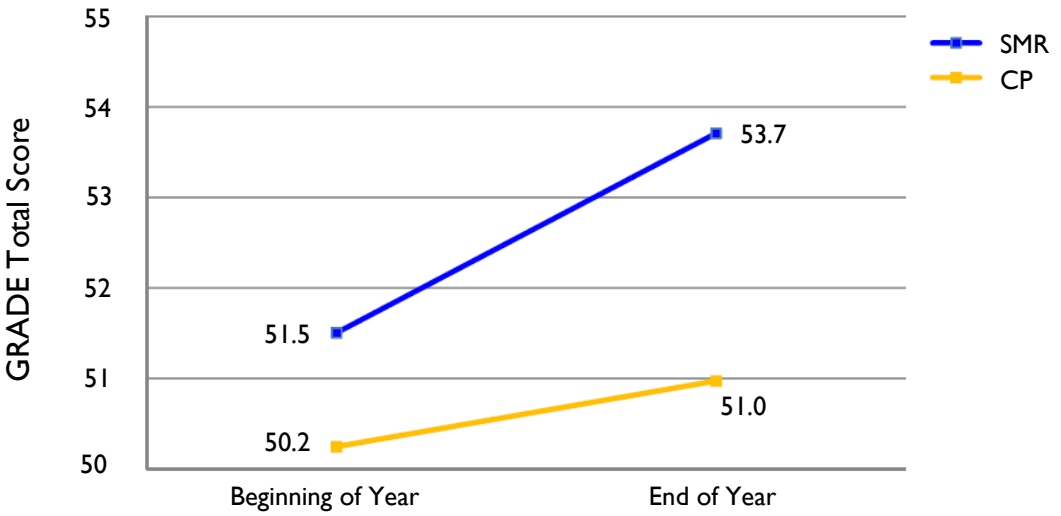


Figure 4. Seventh Grade Reading Achievement Gains



SuccessMaker students in 3rd, 5th, and 7th grade saw large statistically significant gains on the GRADE from the beginning to the end of the school year. In addition, *SuccessMaker* students in 3rd, 5th, and 7th grade statistically significantly outperformed their comparison group counterparts on the GRADE Total score.

Across all grade levels, *SuccessMaker Reading* students also had significantly larger gains on the subtests of sentence and passage comprehension. At 7th grade, a significant effect in favor of *SuccessMaker Reading* was also found on the vocabulary subtest. The only two instances when the comparison group outperformed the *SuccessMaker* students

was on the Word Reading subtest at 3rd grade, and the AIMSweb fluency scale at 5th grade.

The *SuccessMaker* students at 3rd grade did, however, outperform the comparison group on the AIMSweb fluency scale (i.e., words read correctly). It should be noted this basic measurement of accuracy and pacing for oral reading is most appropriate as an outcome for early elementary grades when judging the efficacy of the *SuccessMaker Reading* program. In the *SuccessMaker Reading* program, accuracy and pacing for oral reading is emphasized most in 2nd and 3rd grade. Accuracy and pacing is minimized at 5th grade and not included in the program at all in 6th through 8th grade.

Results for SuccessMaker Reading Users versus Comparison by Subpopulations

When the data was broken out for student subpopulations, results indicated that the subpopulations of Hispanic and low SES really benefited from SuccessMaker Reading. Specifically, 3rd grade Hispanic, male, and low SES *SuccessMaker* students all statistically significantly outperformed their comparison group peers on the GRADE. Similarly, 3rd grade lower-achieving, African American, female, and low SES *SuccessMaker* students saw moderate sized gains over the comparison group in fluency.

Hispanic and low SES *SuccessMaker* 5th grade students statistically significantly outperformed their comparison group peers on the GRADE. Conversely, for all subpopulations except African-American and lower-achieving students, the comparison group outgained the *SuccessMaker* students in fluency.

In all 7th grade subpopulations, *SuccessMaker* students statistically significantly outperformed their comparison group peers on the GRADE and saw moderate to large effects. In fluency, though the whole sample did not show a statistically significant difference, the African-American and Hispanic comparison students had somewhat greater gains.

Participant Feedback

Student Attitudes

SuccessMaker Reading students at 3rd and 5th grade demonstrated statistically greater gains in their academic attitudes than their comparison group counterparts. These effects were also seen in several at-risk populations.

When students were surveyed at the end of the school year as to their opinions on several aspects of the program, 96% of 3rd grade, 85% of 5th grade, and 78% of 7th grade students indicated they liked using the *SuccessMaker* program.

Similarly, the users found the learning activities and stories engaging. Ninety-five percent of 3rd grade students responded that they liked the

characters and animation. Finally, the majority of 5th (i.e., 73%) and 7th (i.e., 64%) grade users reported the video hosts as being helpful to them while using the program.

Teacher Attitudes

"SuccessMaker is a very engaging, interactive program that is differentiated for students' achievement levels."

– Seventh Grade Teacher

"It's easy to not give them [higher kids] the attention that they need, so I really appreciate when the program gets harder and they say that it is difficult because they are finally being challenged. This is great."

– Interventionist

"I don't have to re-assess students to find out where their gaps are, SuccessMaker already tells me."

– Seventh Grade Teacher

"I think the animation is great. It captures [the students'] attention right away. I see them laughing and it keeps them engaged."

– Interventionist

"There have been times when they run across something in SuccessMaker that I've already taught and there are other times when the program will teach them something that is totally new and they'll either bring it back to the classroom or when I go to teach that thing they'll say 'Oh, I already know this from Success Maker!' which is great."

– Third Grade Teacher

Opinions about the program were systematically collected from teachers during focus group sessions. Thirty-six of the 37 *SuccessMaker* teachers were available to participate in the focus group sessions providing extensive insight into teacher and student experiences with, and attitudes about, the *SuccessMaker Reading* program.

The teacher response to the program was overwhelmingly positive. Teachers felt that the program was a welcomed and successful addition to their print curriculum for many reasons including interactivity, differentiated content, immediate feedback, and student engagement. Most teachers felt the initial placement was satisfactory for the majority of students and that the adaptive motion through the content worked well. The occurrence of students being initially placed too high or too low was rare, and the custom course feature allowed teachers to easily rectify the situation.

The program's reporting feature was also well-received by the teachers. Most teachers expressed an appreciation for a program that explicitly demonstrates student progress such as the *SuccessMaker Reading* program. Teachers also liked the progress-reporting feature that monitored the students' motion through the program.

Teachers believe that their students like using the program. Teachers were tremendously positive about their students' interactions with the program and felt that the program successfully engaged and motivated students to become better readers. A majority of teachers felt that the program challenged both their special needs and higher achieving student population. Teachers also felt the *SuccessMaker* program was more engaging and challenging than previous printed and computer-based supplements, helpful for ELL students and struggling readers, and an overall good educational investment.

Additionally, teachers indicated that the content of the *SuccessMaker Reading* program is generally aligned with their current curriculum. Most teachers felt the program reinforces skills already learned in the classroom, and also teaches students reading skills or concepts they have not yet learned in class. Teachers felt that the program's scaffolding feature positively challenged their students to become better readers and that the *SuccessMaker Reading* program provided other valuable resources to assist users. Such assistance proved to be motivational for struggling readers and included the glossary, clip art, read-to-me and roll-over audio features, as well as custom courses.

Conclusion

The study sample included sizable portions of the type of at-risk students that would benefit from a well-conceived and implemented reading intervention, specifically; Hispanic, African American, low SES, and lower achieving. Teachers came up with creative solutions to get all students on the program each week, overcoming packed classroom lesson plans and filled computer lab schedules and firmly believed that their students liked using *SuccessMaker Reading* and felt that the program made the learning process more fun for students. A majority of students reported positive attitudes towards the program as well as more positive academic attitudes than non-users.

The achievement data implies that students using *SuccessMaker Reading*, including at-risk students, can be more successful in vocabulary, comprehension and fluency when receiving 16 hours or more on the program over their first school year using the program. Further, it appears users can be more successful the more they use the program.

About Gatti Evaluation, Inc.

Gatti Evaluation was founded in 2003 to provide assistance in researching current topics in education and biomed. Gatti has extensive experience managing and consulting larger research projects for Fortune 500 companies and major academic institutions. Gatti researchers hold advanced degrees in Research Methods and Education. They also collaborate with numerous hand-picked, world-renowned researchers, practitioners, and academic research centers. Learn more at www.GattiEval.com.

To learn more about the SuccessMaker program or to access the full report, visit us at www.SuccessMaker.com.

PEARSON SUCCESSMAKER MATH EFFICACY STUDY

2009-10 FINAL REPORT

September 15, 2010

Principal Investigator

Guido G. Gatti
Gatti Evaluation Inc.
162 Fairfax Rd.
Pittsburgh, PA 15221
(412) 371-9832
gggatti@gattieval.com

Co-Principal Investigator

Katya Petrochenkov
Gatti Evaluation Inc.

Primary Stakeholder

Funded By **Pearson**

For Information Please Contact:

Marcy Baughman
Director of Academic Research
(724) 863-1621
marcy.baughman@pearsoned.com

TABLE OF CONTENTS

EXECUTIVE SUMMARY

Pearson partnered with Gatti Evaluation to conduct rigorous research to support the assertion that the SuccessMaker Math computer based learning program effectively increases student mathematics achievement and attitudes. The program was evaluated in sixty-three diverse elementary and middle grade classrooms from ten schools in seven different states (i.e., AZ, AR, CA, IN, KS, NY, PA) during the 2009-10 school year. Students in classrooms randomly assigned to use SuccessMaker made regular use of the program while students in comparison classrooms received supplemental instruction from non-computerized supplemental mathematics programs. Four widely-used classroom mathematics programs were utilized by the sites at 3rd and 5th grade, and three different programs were utilized at 7th grade.

The study schools come from public school districts located in large cities or suburbs of large cities. The study schools show considerable variation in ethnicity, students eligible for reduced priced lunch, as well as a wide range of ability with respect to mathematics and reading achievement. The evaluation team sought out diversity in the study sample to ensure the program would be used by learners of all abilities and backgrounds, thus reflecting the reality that is today's elementary classrooms. Five schools began the study in the first month, three began in the third month, one in the fourth and the last in the fifth month of the school year. The final study sample was large, consisting of 505 3rd grade (i.e., SuccessMaker = 282, comparison = 223), 408 5th grade (i.e., SuccessMaker = 224, comparison = 184) and 273 7th grade (i.e., SuccessMaker = 136, comparison = 137) students.

A challenging assessment battery was group administered to students at baseline and again at the end of the school year. The assessment battery consisted of the *Group Mathematics Assessment and Diagnostic Evaluation* (GMADE), and the mathematics attitude survey developed by the principal investigator where students respond to self-report questions regarding general math attitude, confidence, motivation, and self-perceived aptitude. Comparisons on assessment outcomes were made between study groups using model adjusted end-of-year raw score group mean differences. Adjusted group mean differences are calculated holding the effects of confounding variables constant for both groups. The equating of confounding variables and the maintaining of consistent implementation ensures the outcomes may more confidently be attributed to the study conditions randomly assigned to these groups.

Results were broken out and analyzed separately for each GMADE subtest (i.e., Concepts and Communication, Operations and Computation, and Process and Applications). Results were also broken out and analyzed for separate levels of five key student populations (i.e., English proficiency, ethnicity, gender, meal status, math ability). Further, the performance for the comparison group was compared to four blocks of program usage (i.e., block 1 = 1 to 9 hours, block 2 = 10 to 19 hours, block 3 = 20 to 29 hours, block 4 = 30 or more hours).

RQ: How did teachers and students react to the SuccessMaker Math program?

Focus groups were conducted at each school during site visits between April and early June. These sessions provided the evaluators with insights into teacher and student experiences with the program. Teachers and students became quickly comfortable with the SuccessMaker program, and felt the program was a good educational investment. The teacher response to the

program was overwhelmingly positive, with 80% of the 646 recorded comments coded as positive in nature. Teachers appreciated the reporting system for informing classroom instruction, identifying students for remediation, monitoring student progress, and as a tool to share student progress with curriculum specialists and parents. A majority of teachers felt the initial placement and the adaptive motion of sequencing students through the program was effective. In addition, the learning activities were rated as well-differentiated and aligned to current curricula and state educational objectives, and the program challenged both lower and higher achieving student populations. Teachers reported rare minor technical issues (ex., logging in, activities loading), primarily the result of district and school infrastructure.

Teachers firmly believed that their students enjoyed using the math version of SuccessMaker, and felt that the program made the learning process more fun. When formally interviewed, teachers were overwhelmingly positive about their students' interactions with the program. Of the 170 recorded comments, 79% were positive in nature. When students were surveyed, 93% of 3rd grade, 79% of 5th grade, and 88% of 7th grade students indicated they liked using the SuccessMaker program. Third grade students responded most positively to the characters and animation, and found the learning activities engaging. Fifth and 7th grade students more often perceived the characters as immature and the animation sometimes excessive and distracting (i.e., only 9% of 3rd grade versus 28% of 5th grade and 35% of 7th grade students indicated they disliked the animation).

RQ: How was the SuccessMaker Math program utilized?

Students in the 3rd grade SuccessMaker classrooms used the math program a median of 19 hours, attempted a median 43 exercises every thirty minutes with a median success rate of 69%. Students in the 5th grade SuccessMaker classrooms used the math program approximately 18 hours, attempting 44 exercises every thirty minutes with a success rate of 68%. Students in the 7th grade SuccessMaker classrooms used the math program approximately 17 hours, attempting 38 exercises every thirty minutes with a success rate of 63%.

The majority of study teachers implemented SuccessMaker Math in a computer laboratory environment, typically implementing the program 2-3 days per week for an average of 24 minutes. Ten teachers implemented SuccessMaker in the lab more than three times a week. Three teachers utilized a joint-usage model, implementing SuccessMaker in the classroom for 30% to 40% of the total usage, and the remainder in the computer lab. One 3rd grade teacher chose not to utilize the computer laboratory after a couple months of implementation, and implemented SuccessMaker the remainder of the year in the classroom with laptop stations (accounting for 75% of total usage minutes in the classroom). SuccessMaker students in 3rd and 5th grade generally used the program in addition to their regular math block, while 7th grade students used SuccessMaker during their daily math block.

RQ: Do 3rd, 5th, and 7th grade students making regular use of the SuccessMaker Math program demonstrate higher mathematics achievement as compared to students that did not utilize SuccessMaker Math?

SuccessMaker students in 3rd, 5th, and 7th grades statistically significantly outperformed their comparison group counterparts on the GMADE Total score. The magnitude of the difference in performance observed at all three grades was remarkable, with standard deviations of 1.00, 0.53,

and 0.61 for 3rd, 5th, and 7th grade respectively. These effects were consistently large across usage levels. SuccessMaker students in 3rd, 5th, and 7th grade statistically significantly outperformed their comparison group counterparts on the Process and Applications subtest. The magnitude of the difference in performance observed at all three grades was again very large, with standard deviations of 1.32, 0.59, and 1.01 for 3rd, 5th, and 7th grade respectively. These effects were also consistently large across usage levels.

SuccessMaker students in 3rd and 5th grade statistically significantly outperformed their comparison group counterparts on the Operations and Computation subtest. The magnitude of the differences in performance observed at both grades were equivalently very large, 0.75 standard deviations. And yet again, these effects were consistently large across usage levels. The 7th grade SuccessMaker students outperformed their comparison group peers though not statistically so. Though the SuccessMaker students in 3rd and 7th grade performed statistically similar to the comparison group on the Concepts and Communications subtest, the 5th grade comparison group statistically significantly outperformed the SuccessMaker group on this subtest.

When the data was broken out for student subpopulations, 3rd grade Hispanic, low SES, non-English proficient, female, and lower-achieving SuccessMaker students all statistically significantly outperformed their comparison group peers on GMADE Total score (i.e., 0.50 to 1.31 standard deviations), as well as the Process and Applications (i.e., 0.91 to 1.65 standard deviations) and the Operations and Computation subtests (i.e., 0.49 to 1.19 standard deviations). Low SES, non-English proficient and female 5th grade SuccessMaker students statistically significantly outperformed their comparison group peers on GMADE Total score (i.e., 0.48 to 0.53 standard deviations), as well as, both the Process and Applications (i.e., 0.49 to 0.63 standard deviations) and Operations and Computation subtests (i.e., 0.55 to 0.73 standard deviations).

Seventh grade low SES, non-English proficient, and female students all dramatically outperformed their comparison group counterparts on GMADE Total score (i.e., 0.57 to 0.66 standard deviations) and the Process and Applications subtest (i.e., 1.06 to 1.39 standard deviations). Further, lower-achieving and Hispanic 7th grade SuccessMaker students statistically outperformed their comparison group peers on the Process and Applications subtest (i.e., 0.58 and 1.19 standard deviations).

RQ: Do 3rd, 5th, and 7th grade students using the SuccessMaker Math program demonstrate more positive attitudes toward mathematics and mathematics instruction as their comparison group counterparts?

The 3rd and 7th grade SuccessMaker students both had statistically significantly higher math academic attitudes than the comparison group (i.e., 3rd 0.99 standard deviations, 7th 0.62 standard deviations). The very large statistically significant effects seen at 3rd grade were also seen for Hispanic, low SES, non-English proficient, female, and lower-achieving students (i.e., 0.29 to 1.13 standard deviations). Several 7th grade at- risk populations (i.e., female, low SES, non-English proficient) also had statistically higher math attitudes than the comparison group (i.e., 0.61 to 0.69 standard deviations).

I. INTRODUCTION

As elementary and middle schools strive to meet the adequate yearly progress goals set for them in reading and mathematics achievement, many are attempting to maximize their efforts by turning to instructional technology like the SuccessMaker©[1] program. Gatti Evaluation partnered with Pearson to evaluate the effectiveness of the SuccessMaker program. Information gathered during this study will inform future revisions of the program and provide evidence of program efficacy.

Pearson partnered with Gatti Evaluation to study the efficacy of the SuccessMaker Math program in achieving positive educational attitudes and achievement outcomes.

This report provides methods and results from the first phase of the efficacy research conducted during the 2009-10 school year on the SuccessMaker Math program; including the study methodology, nuanced program usage information, teacher and administrator attitudes, as well as student attitudinal and achievement gains. This efficacy study evaluated the Math program in sixty-three diverse 3rd, 5th and 7th grade classrooms from ten schools in seven different states (i.e., AZ, AR, CA, IN, KS, NY, PA).

Instructional Technology Literature

SuccessMaker is an adaptive computer based learning environment that offers an instructional management system, placement and formative assessment, individualized elementary and middle grade reading and mathematics curriculum resources, and a student progress reporting system.

SuccessMaker is an adaptive computer-based learning environment that offers an instructional management system, placement and formative assessment, individualized elementary and middle grade reading and mathematics curriculum resources, and a reporting system to inform administrators and teachers as to student progress. It is widely believed that making formative assessment an integral part of instructional practice is one of the best ways to improve student learning.[2] The National Council of Teachers of Mathematics also emphasizes that technology can enhance mathematics learning and supports effective mathematics teaching and skills practice. Mathematics education and instruction may be aided by technology in various ways, with the technology assuming the role of enhancing, amplifying, and organizing curricula.[3] It is also well documented that both the scope of ways and effectiveness of technology in aiding instruction is increasing with each passing decade.[4] What remains unclear are the best ways to

[1] http://www.pearsondigital.com/
[2] National Council of Teaching of Mathematics (2000). Principles and Standards for School Mathematics. Reston, Va.: National Council of Teachers of Mathematics.
[3] Heid, M. K. (1997). The technological revolution and the reform of school mathematics. *American Journal of Education, 106*(*1*), p5-61.
[4] Jenks, M. S., & Springer, J. M. (2001). A view of the research on the efficacy of CAI. *Electronic Journal for the Integration of Technology in Education, 1*(*2*).

utilize technology to find significant improvement in student achievement over non-technology methods that make use of the same pedagogy.

Theoretically, well-designed mathematics interventions can increase student achievement, specifically in the acquisition and practice of basic skills, especially when integrated with classroom instruction.[5] Although an intervention may be skillfully applied to create an educational environment that significantly increases achievement, poorly designed and implemented interventions will provide little or no benefit, and may even be detrimental. Poorly designed and implemented curricula can confuse and frustrate students and teachers, proving to be a waste of money and valuable learning time. For these reasons, state adoption committees and the federal government (i.e., No Child Left Behind Act[6]) require publishers to conduct rigorous efficacy research to support their educational materials.

Study Goals And Research Questions

The primary goal of this study is to conduct rigorous research to support the assertion that the SuccessMaker Math program effectively increases students' mathematics achievement and attitudes. This study is testing the SuccessMaker program during the first year of implementation as it is typically the most challenging year for any new program to impact student achievement. The SuccessMaker program was tested against comparison classrooms that did not utilize a computerized intervention program, which were randomly selected within each school .

The second goal of the study was to collect information on teacher and student attitudes toward specific features and aspects of the SuccessMaker program. These research questions are classified into two categories; how do teachers and students respond to the program, and how is the program being used?

The research questions for this study are outlined in the following four parts:

RQ1: Do 3rd, 5th, and 7th grade students making regular use of the SuccessMaker Math program demonstrate higher mathematics achievement as compared to students that did not utilize SuccessMaker Math?

RQ2: Do 3rd, 5th, and 7th grade students using the SuccessMaker Math program demonstrate more positive attitudes toward mathematics and mathematics instruction as their comparison group counterparts?

RQ3: How did teachers and students react to the SuccessMaker Math program?

RQ4: How was the SuccessMaker Math program utilized?

[5] Parr, J. M., & Fung, I. (2000). A review of the literature on computer-assisted learning, particularly integrated learning systems, and outcomes with respect to literacy and numeracy: Final Report. Report to New Zealand Ministry of Education.
[6] http://www.ed.gov/nclb/landing.jhtml

II. METHODOLOGY

The SuccessMaker Math program was evaluated in sixty-three diverse 3rd, 5th and 7th grade classrooms from eight urban and suburban school districts in seven different states (i.e., AZ, AR, CA, IN, KS, NY, PA) during the 2009-10 school year. The program was evaluated via a two-group, classroom level randomized, baseline to post observation assessment research design. Teachers or sections within each school were randomly assigned to one of two study groups (i.e., comparison v. SuccessMaker Math). Students in classrooms randomly assigned to implement SuccessMaker Math made regular use of the program for one hour a week in two or three sessions while students in comparison classrooms generally received supplemental instruction from non-computerized mathematics programs currently in use at their school.

Gatti Evaluation provided research schools all data collection materials, maintained constant communication with the study sites, and followed clear data collection procedures throughout the study to ensure that both study and program implementation ran smoothly and effectively. The following sections provide information on study procedures, including; student and teacher level data collection, site recruitment and selection, the nature of math instruction at the study sites, program training and implementation, detail on educational settings at each study site, demographic information for study participants, and the statistical methodologies used to analyze outcomes.

Student Outcome Measures

A challenging assessment battery was group administered to students to measure achievement and academic attitude growth during the school year.

An assessment battery was group administered to students, proctored by their teachers, at the start of program use (i.e., baseline testing) and again in the last month of the school year (i.e., end-of-year testing). The assessment battery consisted of the *Group Mathematics Assessment and Diagnostic Evaluation* (GMADE) and a mathematics academic attitude survey. The assessment battery was intended to challenge the students; attempting to adequately assess incoming mathematics knowledge for a wide range of abilities while providing room for growth as knowledge was acquired during the school year.

The GMADE is a standardized, nationally norm-referenced mathematics achievement test published by Pearson Assessments. The GMADE was constructed with all fifty states' standards in mind, covering a wide range of content topics and skills. The GMADE includes 9 levels that span grades K-12, each with two parallel forms (i.e., level 3 for 3rd grade, level 5 for 5th grade, level M for 7th grade). Form A was administered at baseline and form B was administered at the end of the school year. The GMADE is not a timed test, but generally takes between 60 and 90 minutes to administer. Sites returned completed student tests to the site coordinators, who then shipped the tests to the research team for hand-scoring.

Both GMADE overall and subtest scores were reported. The subtest scores allowed the research team to evaluate the effectiveness of the curricula on three important dimensions. The subtests are *Concepts and Communication* (28 questions), *Operations and Computation* (24 questions), and *Process and Applications* (28 questions level 3, 30 questions levels 5 and M). These subtests address students' knowledge of mathematics representations and language, use of basic computational algorithms and operations, and the ability to solve problems presented in written form, respectively.

The math academic attitude survey was developed by the Gatti Evaluation principal investigator. Students responded to self-report questions (i.e., 13 questions at 3rd grade, 16 questions at 5th and 7th grade) regarding general math attitude, confidence, motivation, and self-perceived aptitude. Student responses were coded as 1 for a positive response, 0 for a neutral response, and -1 for a negative response. This scoring method anchors a completely neutral student at an overall score of zero with positive total scores indicating an overall positive attitude. Lastly, students in SuccessMaker classrooms were surveyed as to their opinions on several aspects of the program.

The estimated intraclass reliability for GMADE scores tended to be less reliable as grade level increased. However, all subtest scores were deemed reliable enough for the purposes of study analyses. The estimated intraclass reliability for the 3rd, 5th, and 7th grade mathematics attitude scores was 0.75, 0.77, and 0.78 respectively.

3rd Grade Scale	Reliability[1]
GMADE Total	0.96
Concepts and Communication	0.87
Operations and Computation	0.91
Process and Applications	0.92
Math Academic Attitude Survey	0.75

1. Sample estimated coefficient alpha intraclass reliability.

5th Grade Scale	Reliability[1]
GMADE Total	0.94
Concepts and Communication	0.83
Operations and Computation	0.86
Process and Applications	0.88
Math Academic Attitude Survey	0.77

1. Sample estimated coefficient alpha intraclass reliability.

7th Grade Scale	Reliability[1]
GMADE Total	0.91
Concepts and Communication	0.77

Operations and Computation	0.85
Process and Applications	0.77
Math Academic Attitude Survey	0.76

1. Sample estimated coefficient alpha intraclass reliability.

Teacher Measures

The research team also collected data through teacher logs and classroom observations, as well as teacher interviews and focus groups. The teacher and classroom data increased the validity of the research findings related to achievement outcomes by verifying results through multiple data collection methods, adding context for results through the perspectives of various participants, and by collecting data at various time points during the study.

The research team collected achievement, attitudinal, as well as, observational and self-report data making the study both quantitative and qualitative in nature.

In addition to the assessment battery, qualitative data collection methods were also employed. The sources of qualitative data included; program reports, teacher surveys, daily lesson logs, classroom observations, as well as, teacher notes from electronic correspondences. Teachers were routinely asked for their opinions throughout the school year. Weekly mathematics lesson notes were collected for both SuccessMaker and comparison classes (i.e., 10-15 minutes completion time per week). Cumulative usage reports and program implementation logs were regularly collected from SuccessMaker users. All study classrooms were observed twice during the school year teaching routine mathematics lessons and SuccessMaker teachers were further interviewed as to their opinions regarding the program. All this data was compiled and content analyzed to determine teacher attitudes and performance, as well as to illuminate the various ways teachers and students interact with the program.

Weekly Teacher Logs

All study teachers were required to complete weekly logs in which they describe their mathematics lessons. Information from the weekly logs was important for two reasons. The first is to guarantee SuccessMaker teachers fully and regularly utilized all key components of SuccessMaker Math to provide adequate opportunity for the program to positively influence student achievement. The second reason was to document the instructional model for all study teachers, including; classroom environment, teaching style, pacing and mathematics content and methods.

Teachers were asked not to spend more than 15 minutes per week completing the logs. It is clear several teachers spent more time, however, as many of the logs were returned with detailed comments. Teachers often shared candid weekly experiences with the Gatti Evaluation project manager and were typically happy to provide documentation describing weekly instruction and learning experiences related to the program. SuccessMaker and comparison group teachers summarized daily classroom mathematics instruction time, topics, and methods. Daily summaries also included the amount of time spent on these activities. In addition, SuccessMaker

teachers summarized program usage and details of how information from the program was integrated into classroom instruction.

Teacher Observations

Classroom observations took place between mid-November and mid-March and again between April and early June. Classroom observations were conducted by the research team. All study classrooms from each site were observed at least once during routine mathematics lessons. Portions of the observation forms include a description of the classroom environment, summary of the lesson taught, teacher interviews, student comments, observed teaching strengths and weaknesses, pacing, and supplemental instruction information.

Students were also observed using the SuccessMaker Math program in both the classroom and/or computer lab. These observations gave the research team an opportunity to witness the ability and willingness of teachers to properly use the program in the laboratory and/or classroom, verify teacher reported information, identify adherence to the program usage schedule, as well as observe general classroom environment and teaching styles.

It should be noted that two classroom observations provide just a snapshot of the classroom environment and instructional competence. Some teachers were required to change their normal class time due to scheduling conflicts, which occasionally resulted in the observer having less than optimal time to spend in the classroom. The observations are, however, worthwhile because they are the only opportunity the research team has to directly observe the study teachers in action and verify teacher reported information.

Teacher Surveys

All participating teachers were administered two surveys about their teaching background: a baseline survey, and an end-of-year survey. The purpose of the baseline teacher survey was to collect information on teaching experience, math curricula, and prior research study experience. Teachers were asked to indicate their highest level of education and the number of years teaching total, as well as years they had spent at their district, school, and grade level.

The end-of-year teacher survey was focused more on gathering details about school context, teaching philosophy, and math curriculum implementation. Teachers were asked about their curriculum materials, technology usage, and teaching strategies. Teachers were also asked to describe ways in which their school and students are unique. All of this information allowed researchers to gain additional insight into the overall experience at each research site.

SuccessMaker Teacher Focus Group

A focus group style interview process was chosen by the research team to collect teacher attitudes towards the SuccessMaker program. The face-to-face nature of a focus group, though more labor intensive, can be superior to simple questionnaires in collecting detailed attitudinal information from participants. When properly conducted, the focus group discussion gravitates to those topics most important to the participants, and can provide more nuanced information. Collecting attitudinal data in person allows for a better understanding of participant tone and gravity of responses, and provides opportunity to delve deeper into topics.

The focus group results describe what teachers and students liked about the SuccessMaker program, how the program could be improved, and how teachers are using specific features of the system.

Focus group sessions were conducted at each school during site observations between April and early June. Representatives from the research team facilitated each session. The sessions lasted approximately 60 minutes. Twenty-nine of the 32 SuccessMaker teachers participated in the focus group sessions. One teacher who could not participate in the focus group session sent in responses to the focus group questions electronically. The focus group sessions provided a forum for teachers and administrators to answer specific questions, as well as express their professional and personal opinions regarding the SuccessMaker Math program. Each session held the teachers' comfort level as a high priority. The teachers were encouraged to speak without hesitation or inhibition and to be as honest and candid as possible. Though the facilitator followed a structured interview format, the teachers were allowed to direct the discussion and provide their reactions to- and comment on- any and all aspects of the program.

Teachers were asked about their general opinions of the SuccessMaker Math program, as well as their reactions to specific features. In order to uncover how teachers were integrating report information from the program with their classroom instruction and goals, questions were asked pertaining to the reporting system and how teachers were utilizing that system. Teachers were also asked to describe student reactions to the program and how the program impacted their students' learning experience. Efforts were made to minimize response bias by avoiding leading questions and asking for the program's strengths and weaknesses alike.

Extensive notes were taken at each session allowing the research team to compile a large master file of participant responses. Following an exhaustive review of the teacher responses, a two-dimensional coding system was developed to organize those responses. Responses were categorized by *Topic Area* and *Attitude*. The topic areas describe the aspect of the program a response is directed towards. Topic area codes have a two-digit numeric format with the first digit on the left indicating general topic category (ex., teacher opinion, student response to program, program content, specific features) and the second digit indicating a specific topic within a general category. The topic codes are further categorized by grade level, study site, and paired with either a + or - to indicate the general attitude toward an aspect of the program or the tone of the response.

Site Recruitment and Selection

Gatti Evaluation and Pearson Digital Learning account executives identified potential research partners that met certain characteristics important to the study, such as no previous exposure to any version of SuccessMaker, at least 2 teachers per study grade level, and geographic diversity. Potential research schools were contacted by e-mail and given details about the study. Probable sites were further vetted through their Pearson Digital Learning account executive, than invited to participate in the study. As schools responded to the invitation, they were further screened with a detailed questionnaire and an infrastructure checklist. The intent of the questionnaire was to ensure participants understood all the requirements and benefits associated with participation. It was required that schools did not currently use the SuccessMaker program, all participating

teachers abide by the random assignment, and all randomly selected SuccessMaker classroom students use the program for a minimum of one hour per week. The purpose of the infrastructure checklist was to ensure that the SuccessMaker program could be installed and successfully run at each site.

When sites were deemed eligible for participation and demonstrated strong interest, the Principal Investigator completed the research application process with each site. Final acceptance to the study required a district level administrator (ex., curriculum director, superintendent) and a school level administrator (ex., principal) to sign a memorandum of understanding outlining the responsibilities of each stakeholder. No available students of any socio-economic level, English proficiency level, or ethnic background, who opted to participate in the study, were excluded from the study. Passive informed consent of both students and parents/guardians was required by the research team and secured by the schools.

Table 1 **Gatti Evaluation SuccessMaker Math Study Site State Assessment Information**

					School Results		State Wide Results	
School Year	Grade	State	District	School	Meets Math Standards	Meets Reading Standards	Meets Math Standards	Meets Reading Standards
2008-09	3	AZ	1	1	37%	46%	72%	72%
2008-09	5	AZ	1	1	50%	41%	72%	73%
2008-09	7	AZ	1	1	62%	65%	73%	73%
2008-09	3	AZ	2	2	53%	58%	72%	72%
2008-09	5	AZ	2	2	52%	63%	72%	73%
2008-09	7	AZ	2	2	66%	62%	73%	73%
2008-09	7	AZ	2	3	77%	56%	73%	73%
2008-09	3	AR	3	4	94%	91%	80%	66%
2008-09	5	AR	3	4	94%	90%	70%	68%
2008-09	3	CA	4	5	50%	31%	64%	44%
2008-09	5	CA	4	5	26%	39%	57%	54%
2008-09	3	IN	5	6	56%	67%	69%	74%
2008-09	5	IN	5	6	69%	74%	77%	74%
2008-09	3	KS	6	7	81%	72%	86%	84%
2008-09	5	KS	6	7	84%	75%	87%	84%
2008-09	7	KS	6	8	60%	65%	78%	86%
2008-09	3	NY	7	9	98%	85%	93%	76%
2008-09	5	NY	7	9	98%	100%	88%	82%
NA	3	PA	8	10	NA	NA	NA	NA
NA	5	PA	8	10	NA	NA	NA	NA

School Year designates latest school year state assessment information was available. The PA school was new in the 2009-10 school year.

The study schools come from urban or suburban public school districts. A single school represented each of Arkansas, California, Indiana, New York, and Pennsylvania. Two school districts came from Arizona. One school from each of these districts served kindergarten through 8th grade students and the second school from the second Arizona district was a middle school. Lastly, both an elementary and middle school represented the Kansas school district.

Ethnic and socio-economic diversity among the student population were two criteria the evaluation team considered when recruiting study sites. A third criterion was that students exhibit a wide range of ability with respect to mathematics and reading achievement. Table 1 shows, according to recent state achievement testing data, the percent of each school's students meeting state math standards range between 35% below to 24% above statewide results and students meeting state reading standards range between 32% below to 25% above statewide results. The evaluation team sought out diversity in the study sample to ensure the program would be used by learners of all abilities and backgrounds, thus reflecting the reality that is today's elementary classrooms.

Math Instruction

Teachers were expected to implement their current adopted core mathematics curricula as required by their district. Four widely-used classroom mathematics programs were utilized by the sites at 3rd and 5th grade, and three different programs were utilized at 7th grade. The study groups reported somewhat differing levels of adherence to their adopted math programs. Supplemental math instruction seen across sites included commonplace methods such as website exploration, math facts, daily math problems, and test preparation.

Adopted Math Program Adherence		
3rd Grade	SuccessMaker	comparison
strict	12.8%	18.8%
mostly	59.2%	61.9%
some	20.6%	19.3%
none	7.4%	0.0%
Percents are statistically significantly different		

Adopted Math Program Adherence		
5th Grade	SuccessMaker	comparison
strict	10.3%	25.0%
mostly	89.7%	75.0%
some	0.0%	0.0%
none	0.0%	0.0%
Percents are statistically significantly different		

Adopted Math Program Adherence		
7th Grade	SuccessMaker	comparison
strict	35.0%	26.3%
mostly	65.0%	73.7%
some	0.0%	0.0%
none	0.0%	0.0%
Percents are *NOT* statistically significantly different		

SuccessMaker and comparison groups where similar in teacher experience, both in years teaching and years teaching current grade. The study sample did have two years less teaching experience (i.e., 11.6 years) than the national average (i.e., 13.9 years). More of the 3rd grade comparison sample had a higher portion of students taught by a teacher with a Master's degree. Conversely, the 5th and 7th grade SuccessMaker samples had a higher portion of students taught by a teacher with a Master's degree. Average minutes of classroom math instruction were nearly equivalent for 3rd grade students across the treatment and comparison conditions, however, 5th and 7th grade SuccessMaker classrooms averaged fewer minutes of classroom math instruction as students at some sites used the program during their usual math blocks.

3rd Grade	SuccessMaker	comparison
years teaching	12.38	11.14
years at current grade	6.51	7.44
master's degree	54%	76%
years using adopted program	4.15	3.93
minutes math instruction	63.73	63.15
Difference in percent of teachers with master's degree was statistically significant.		

5th Grade	SuccessMaker	comparison
years teaching	11.61	8.87
years at current grade	4.85	5.45
master's degree	54%	30%
years using adopted program	6.09	3.33
minutes math instruction	69.07	78.72
Difference in years teaching was statistically significant.		
Difference in percent of teachers with master's degree was statistically significant.		
Difference in years using adopted math program was statistically significant.		
Difference in minutes of math instruction was statistically significant.		

7th Grade	SuccessMaker	comparison
years teaching	12.60	13.55
years at current grade	8.45	9.85
master's degree	100%	74%
years using adopted program	4.49	4.83
minutes math instruction	59.35	63.48

Difference in percent of teachers with master's degree was statistically significant.
Difference in minutes of math instruction was statistically significant.

SuccessMaker Implementation

Teachers received multiple training sessions by Pearson curriculum specialists. This well-received training allowed teachers to fully implement the SuccessMaker Math program and fostered positive teacher and student attitudes.

SuccessMaker Teacher Training

To initiate the study, Gatti Evaluation conducted study orientations for all teachers at the start of the school year. The study orientation formally introduced the teachers to the research team, explained in detail the requirements and benefits of participation in the study, as well as, addressed any immediate questions or concerns about the research. All teachers were required to read and sign informed consent forms.

The publisher ensured that sites had full access to the program and that access was continual throughout the duration of the study. Pearson also provided free product training and funding to cover the cost of substitute teachers during training. All teachers with SuccessMaker classrooms were required to attend training sessions facilitated by a curriculum specialist. Initial training took place on-site over the course of one full school day. This training introduced administrators, teachers, and technicians to the key components of the SuccessMaker Math program, including; student login, learning environments, classroom management and reporting systems, as well as how to best implement these in practice. Follow-up training was further provided to each site to support consistent usage of the program and to fully acquaint teachers with all aspects of the reporting system. As needed, additional training sessions were also offered to provide a more detailed understanding of the program, identify and correct district or school level technical issues, address student's special needs, and to support consistent implementation of the program.

Initial product training sessions lasted a full school day and typically began with a group presentation. Then teachers moved to computers where they were given the opportunity to use the program as students would. Teachers had the responsibility of training their students to use the program. The follow-up training sessions typically lasted three hours and began with a group presentation, then teachers moved to computers where they were shown how best to monitor their class and individual student progress. The trainings were well-received. The research team strongly believes that ongoing professional development can significantly affect the potential for

a program such as SuccessMaker to foster positive teacher and student attitudes, meet students' needs, and ultimately increase student achievement.

The date of initial training varied, dependent on when a site was added to the study (i.e., see Table 2). Five schools began the study in the first month (i.e., AZ district 1 school, AZ district 2 k-12 school, KS elementary and middle schools, PA school), three began in the third month (AR school, CA school, IN school), one in the fourth (NY school) and the last in the fifth month (AZ district 2 middle school) of the school year.

Table 2			SuccessMaker Math RCT Training Dates			
State	District	School	School Start Date	Initial Training Date	Follow-up Training Date	Additional Trainings
AZ	1	1	08/03/09	07/29/09	11/04/09	02/12/10
AZ	2	1	08/03/09	08/12/09	11/10/09	N/A
AZ	2	2	08/03/09	11/18/09	03/05/10	N/A
AR	1	1	08/19/09	10/28/09	01/21/10	02/18/10
CA	1	1	09/10/09	12/11/09	03/17/10	N/A
IN	1	1	08/24/09	11/13/09	01/28/10	N/A
KS	1	1	08/12/09	08/10/09	09/21/09	12/11/09
KS	1	2	08/12/09	08/10/09	09/21/09	12/11/09
NY	1	1	09/08/09	12/08/09	02/02/10	03/16/10
PA	1	1	08/03/09	08/26/09	10/12/09	03/31/10

SuccessMaker Program Usage

The majority of study teachers implemented SuccessMaker Math in a computer laboratory environment, typically implementing the program 2-3 days per week for an average of 24 minutes. Ten teachers implemented SuccessMaker in the lab more than three times a week. Three teachers utilized a joint-usage model, implementing SuccessMaker in the classroom for 30% to 40% of the total usage, and the remainder in the computer lab. One 3rd grade teacher chose not to utilize the computer laboratory after a couple months of implementation, and implemented SuccessMaker the remainder of the year in the classroom with laptop stations (accounting for 75% of total usage minutes in the classroom). SuccessMaker students in 3rd and 5th grade generally used the program in addition to their regular math block, while 7th grade students used SuccessMaker during their daily math block.

Students in the 3rd grade SuccessMaker classrooms used the math program a median[7] of 19 hours, attempted a median 43 exercises every thirty minutes with a median success rate of 69%. Usage at 5th grade was nearly the same, at approximately 18 hours, attempting 44 exercises every thirty minutes with a success rate of 68%. Students in the 7th grade SuccessMaker classrooms used the math program approximately 17 hours, attempting 38 exercises every thirty minutes with a success rate of 63%.

[7] This value is the median usage rounded to the nearest hour.

Settings

This section summarizes the educational model and environment for each study site as well as a demographic breakdown. This information is crucial for determining how applicable results from this study may be to the consumers of this report.

Arizona District One

The first participating Arizona school resides in a rural fringe area, has a high student turnover rate and frequent changes in staffing positions. Students are expected to follow a strict dress code. According to teachers, many students come from underprivileged backgrounds and do not generally receive a high degree of parental support. Teachers also describe a variety of learning abilities in the classrooms, as well as motivational and behavioral diversity.

In the 2008-09 school year, the district served a community of over 10,000. The median household income is approximately $50,000 indicating a middle-class community. It is a mid-size school serving over 500 students in grades kindergarten through seven. The primary ethnic group, Hispanic, makes up a total of 67% of the school population. Caucasian, African-American and Asian students make up the remaining 33% of the student population. This school falls into the high range for participation in the nation's free or reduced-price lunch program with 78% of students eligible to receive free or reduced-price lunch. Approximately 22% of the students are designated as not English proficient.

This school did not meet AYP in the 2008-09 school year. The percentage of 3rd grade students testing at standard in mathematics in the 2008-09 school year was 37%, 35% lower than the statewide results. The percentage of 5th grade students testing at standard in mathematics in the 2008-09 school year was 50%, 22% lower than the statewide results and the percentage of 7th grade students testing at standard was 62%, 11% lower than the statewide results. Likewise, the percentage of 3rd grade students testing at standard in reading was 46%, 26% lower than the statewide results. The percent of 5th grade students testing at standard in reading was 63%, 10% lower than the statewide results and the percentage of 7th grade students testing at standard was 65%, 8% lower than the statewide results. The student/teacher ratio is approximately 26 to 1.

One 3rd grade classroom was randomly assigned to use the SuccessMaker Math program and another was assigned to the comparison condition. Early in the school year, the teacher assigned to the comparison condition left the school. The comparison students were disbursed into the SuccessMaker classroom and a new 3rd grade classroom. Though they moved to new classrooms, these students maintained their random assignments and did not use the program. Of the three 5th grade teachers, two were randomly assigned to use SuccessMaker and one was assigned to the comparison condition. The comparison teacher used the program with students and thus they were dropped from the study. The 7th grade math teacher had four sections; two were assigned to use the program. The students in one of the two 7th grade sections assigned to use SuccessMaker did not complete the full assessment battery and were dropped from the study. As a result, a total of only three teachers from this school ultimately participated in the study, two 3rd grade teachers and a 7th grade math teacher with three participating sections.

The district adopted a widely published elementary basal mathematics curriculum with a late copyright date. One of the 3rd grade teachers adhered strictly to the district adopted curriculum and the other reported using some supplementation, as did the 7th grade teacher. The teachers incorporated district learning standards and AIMS (Arizona's Instrument to Measure Standards) test preparatory work into their daily mathematics lessons, as well as some speed drills. Teachers had used this same basal program for 2 years, however, one of the 3rd grade teachers had just begun her second year. Teachers in this school have daily math blocks of one hour. The study teachers prefer using a combination of skills- and activity-based teaching styles for math instruction and have no additional assistance in the classroom. They conduct their math lessons using whole group instruction 50% or more of the time. Also, the 3rd grade teachers like to explore educational websites on their interactive white boards.

The school has a large computer lab that is housed in the library. Stations are arranged in long rows, facing the same direction. This computer lab is where students used the SuccessMaker program. SuccessMaker students in 3rd and 5th grades used the program in addition to their regular math block, while 7th grade students used SuccessMaker during their daily math block. Those teachers randomly assigned to use the SuccessMaker Math program were trained the week prior to the start of the school year. These teachers also received additional trainings in November and February. Students completed baseline testing the last week of September and were tested again the week of April 30th. Students' last week using the program was the last week of April.

One 3rd grade teacher used the SuccessMaker program in 25 minute sessions three times per week, while the other used the program for 35 minute sessions, also three times per week. The median 3rd grade student used the math program approximately 18 hours, attempting 43 exercises every thirty minutes with a success rate of 66%. In 7th grade, SuccessMaker usage varied throughout the year. While the minimum usage time was not met during the first half of the year, the 7th grade teacher tried to get in at least 75 minutes per week in two separate sessions during the spring term. The median 7th grade student used the math program approximately 12 hours, attempting 34 exercises every thirty minutes with a success rate of 61%.

Arizona District Two

Two schools in the second Arizona district participated in the study. Both schools reside in a suburban area. In 2008-09 the district served a community of over 70,000. The median household income is approximately $65,000 indicating a high-middle class community. Despite this income statistic, many students at these Title 1 schools come from low-income areas, with a high population of Hispanic students and English language learners. Teachers report having a wide range of learning abilities in their classes, and that getting students interested in classroom material is a big challenge. Both schools enforce a strict dress code for their students. The district adopted two widely published elementary basal mathematics curricula with a late copyright date, one for elementary grades and another for middle grades.

The first school in Arizona is a large size school serving over 1,100 students in grades kindergarten through eight. The school has one primary ethnic group, Hispanic, making up a total of 83% of the school population. This school falls into the high range for participation in the nation's free or reduced-price lunch program with 87% of students eligible to receive free or reduced-price lunch. Approximately 41% of the students are designated as not English proficient. The student/teacher ratio is approximately 19 to 1.

This school did not meet AYP in the 2008-09 school year. The percentage of 3rd grade students testing at standard in mathematics in the 2008-09 school year was 53%, 19% lower than the statewide results. The percentage of 5th grade students testing at standard in mathematics in the 2008-09 school year was 52%, 20% lower than the statewide results. The percentage of 7th grade students testing at standard in mathematics in the 2008-09 school year was 66%, 7% lower than the statewide results. The percentage of 3rd grade students testing at standard in reading was 58%, coming in 14% lower than the statewide results. The percent of 5th grade students testing at standard in reading was 41%, 32% lower than the statewide results. The percentage of 7th grade students testing at standard in reading in the 2008-09 school year was 62%, 11% lower than the statewide results.

A total of nine teachers from the first school participated in the study, four 3rd grade, four 5th grade and one 7th grade with four math sections. Two 3rd and 5th grade classrooms were randomly assigned to use SuccessMaker Math. Likewise, two 7th grade sections were randomly assigned to use SuccessMaker Math. Daily math blocks last one hour, however, teachers reported average daily math instruction lasting from thirty to over one hundred minutes. Most teachers used the district adopted curriculum with some additional supplementation. One teacher reported using heavy supplementation. Six of the teachers reported having little to no training on this curriculum, two teachers reported having some training on the curriculum, and one teacher reported having 5-8 hours of professional development on the district adopted program. None of the teachers were new to the district for the 2009-2010 school year, although one teacher was new to her grade level.

Four of the nine teachers prefer using a combination of skills- and activity-based teaching styles for math instruction, four others expressed a preference for a skills-based, and one, activity-based teaching. They conduct their math lessons using whole group an average of 61% of the time and place heavy emphasis on test preparation for the AIMS. Teachers use several teaching strategies for math instruction. Four teachers reported using leveled instruction, six use cooperative learning strategies, four use center activities, and four use speed drills. In 3rd grade, teachers also employ an outreach program where the students complete a consumable parent/student booklet every month and the school is awarded $1.00 for each student who has completed every lesson. Six teachers reported using educational websites and computer games and two use their interactive white boards. The interactive white boards were later additions to those classrooms, the remainder had digital projectors. Two had student teachers during the year, but no additional classroom assistance was reported during math instruction.

Initially, this school's technological infrastructure was weak and teachers experienced significant problems logging on to the program and the program freezing. It was necessary to borrow teacher computers from this classroom for incorporation into the lab. By the end of the school year the lab was running flawlessly. Computers lined the walls of the room, with an island of stations in the center.

SuccessMaker students used the program three to four times per week in 15-30 minute sessions. Students in 3rd and 5th grade generally used the program in addition to their regular math block, while 7th grade students used SuccessMaker during their daily math block.

The teachers at the first elementary school who were randomly assigned to use SuccessMaker were trained on August 12th, a week-and-a-half into the school year. Theses teacher also received a follow-up training November 10th. Students completed baseline testing on September 26th and completed end-of-year testing the third week of May. Students' last week using the program was the week of May 10th. The median 3rd grade student used the math program approximately 26 hours, attempting 37 exercises every thirty minutes with a success rate of 69%. The median student in the 5th grade used of the math program approximately 31 hours, attempting 38 exercises every thirty minutes with a success rate of 67%.

The second school from this district is a mid-size school serving more than 500 students in grades kindergarten through eight. The school has one primary ethnic group, Hispanic, making up a total of 89% of the school population. This school falls into the high range for participation in the nation's free or reduced-price lunch program with 87% of students eligible to receive free or reduced-price lunch. Approximately 50% of the students are designated as not English proficient. The student/teacher ratio is approximately 16 to 1. The school did meet AYP in the 2008-09 school year.

Only 7th grade students participated at this school. The 7th grade teacher reported language as being one of the biggest challenges in the classroom for her students, she is fluent in Spanish. The percentage of 7th grade students testing at standard in mathematics in the 2008-09 school year was 77%, 4% higher than the statewide results. The percentage of 7th grade students testing at standard in reading in the 2008-09 school year was 56%, 17% lower than the statewide results.

This is the math teacher's second year implementing the district adopted curriculum. She reports that she mostly adheres to the curriculum but with some additional supplementation. The teacher draws on a number of resources to teach students math including a variety of 6th and 7th grade level workbooks and online programs and collaborates with other teachers in the district through meetings held every month. She has been teaching at this school and grade level for 6 years.

The daily math blocks last one-and-a-half hours. The teacher uses whole group instruction about 80% of the time. She sometimes includes cooperative learning and leveled instruction. This teacher also prefers to use a combination of skills-based and discovery-based method. This teacher uses a lot of technology in the classroom, including educational websites two to three times per week, interactive videos once per week, and rounds these out with some educational computer games.

Of the three math sections, two were randomly assigned to use the SuccessMaker math program. The teacher at the second school was trained on November 18th, two and a half months after school began. This teacher also received additional trainings in March. Students completed baseline testing on December 16th and completed end-of-year testing the second week of May. Students' last week using the program was the week of May 10th. The computer lab has two rows of computers directly across from each other, separated by an aisle, with over 30 stations. The set up allows the teacher to walk up and down the aisle to monitor students. The median student used the math program approximately 14 hours, attempting 33 exercises every thirty minutes with a success rate of 66%.

Arkansas District

The participating Arkansas elementary school is a Blue Ribbon School with very high degree of parental support and involvement. Teachers have indicated this also puts a lot of pressure on them to succeed. Students are high achieving and come from higher socioeconomic backgrounds. However, because so many of the students are high achieving, teachers can find it challenging to reach those who are below grade level. The school building is new with high quality facilities.

The school resides in a small city. In 2008-09 the school district served a community of 30,000. The median household income is over $40,000 indicating a middle class community. This elementary school is large, serving almost 700 students in grades kindergarten through five. The school has one primary ethnic group, Caucasian, making up a total of 91% of the school population. This school falls into the medium-low range for participation in the nation's free or reduced-price lunch program with 11% of students eligible to receive free or reduced-price lunch.

This school met AYP in the 2008-09 school year. The percentage of 3rd grade students testing at standard in mathematics in the 2008-09 school year was 94%, 14% higher than the statewide results. The percentage of 5th grade students testing at standard in mathematics in the 2008-09 school year was 94%, 24% higher than the statewide results. The percentage of 3rd grade students testing at standard in reading was 91%, coming in 25% higher than the statewide results. The percent of 5th grade students testing at standard in reading was 90%, 22% higher than the statewide results. The student/teacher ratio is approximately 17 to 1.

A total of 10 teachers participated in the study from the Arkansas school. None of these teachers were new to the school or district or receive additional support in their classrooms. Three 3rd and three 5th grade classrooms were randomly assigned to use the math program. Those teachers randomly assigned to use the SuccessMaker Math program were trained two months after school began on October 28th. These teachers also received additional trainings in January and February. Students completed baseline testing the second week in November and completed end-of-year testing the last week in May. Students' last week using the program was the week of May 24th.

The district adopted a widely published elementary basal mathematics curriculum with a late copyright date. Four teachers adhere strictly to this curriculum, the rest report primarily used the district adopted program, with some supplementation. Teachers have followed this program for a range of 2-13 years. While most teachers report attending two training modules on the district curriculum, or a couple of days worth of training, two teachers report receiving significantly more training.

Third grade teachers have a one-hour-and-fifteen-minute daily math block. Fifth grade teachers have daily math blocks of 55 minutes, and two had an additional 25 minutes of math in the afternoon. Most teachers prefer using a combination of skills-based and discover-based teaching methods for math, and one teacher prefers a skills-based philosophy. Teachers conduct math lessons using whole group approximately 71% of the time. Two teachers report frequently using centers, and two use centers sometimes for teaching math. Most teachers conduct some degree of math test prep with their classroom. Only one teacher reports using leveled math instruction,

and uses this method infrequently. Most teachers incorporate educational websites and computer games into their math instructions and some also use interactive whiteboards.

The Arkansas school has a nice computer lab with more than 30 terminals. The SuccessMaker teachers took their students to use the program in the computer lab two days a week for thirty minutes. SuccessMaker classrooms used the math portion of the program as part of their normal mathematics instruction. The median 3rd grade student used the math program approximately 18 hours, attempting 41 exercises every thirty minutes with a success rate of 70%. The median student in the 5th grade used the math program approximately 18 hours, attempting 49 exercises every thirty minutes with a success rate of 70%.

California District

The participating California elementary school resides in a suburb of a large city. In 2008-09 the school district served a community of more than 100,000. The median household income is approximately $60,000 indicating an upper-middle class community. The school is located in a mostly Hispanic, low socio-economic area and has a high number of students that are English language learners. Students are required to wear uniforms at this Title I school. Teachers are challenged by the fact that many of their students are below-grade level and receive limited support at home. Additionally, the district has recently undergone severe budget cuts and was forced to lay off many teachers.

The elementary school in California is a medium size school serving almost 600 students in grades kindergarten through five. The school has one primary ethnic group, Hispanic, making up a total of 97% of the school population. This school falls into the high range for participation in the nation's free or reduced-price lunch program with 85% of students eligible to receive free or reduced-price lunch. Approximately 55% of the students are designated as not English proficient.

The elementary school did not meet AYP in the 2008-09 school year. The percentage of 3rd grade students testing at standard in mathematics in the 2008-09 school year was 50%, 14% lower than the statewide results. The percentage of 5th grade students testing at standard in mathematics in the 2008-09 school year was 26%, 31% lower than the statewide results. The percentage of 3rd grade students testing at standard in English language arts was 31%, coming in 13% lower than the statewide results. The percent of 5th grade students testing at standard in English language arts was 39%, 15% lower than the statewide results. The student/teacher ratio is approximately 21 to 1.

A total of seven teachers participated in the SuccessMaker study from the California school, five at 3rd grade and two at 5th. None of these teachers were new to the school or district, but three were new to their grade level. Two 3rd grade classrooms and one 5th grade classroom were randomly assigned to use the SuccessMaker math program. Those teachers randomly assigned to use the SuccessMaker Math program were trained three months after school began on December 11th. These teachers also received an additional training in March. Students completed baseline testing the week of December 15th and tested again the week of June 9th. Students' last week using the program was the first week of June.

The district adopted a widely published elementary basal mathematics curriculum with a late copyright date. Most teachers at the California school heavily supplement the district adopted

program with other materials, and work hard to emphasize state standards in their instruction. Most teachers at this school have followed this curriculum for one year, though a few have used it for longer. None of the teachers have received professional development on this curriculum.

Daily math blocks range from 35 minutes to one hour and 25 minutes. Students are instructed in whole group for an average of 60% of the time (i.e., 20% to 75%). Three teachers choose a skills-based teaching philosophy when it comes to mathematics, the rest a combination of skills-based and discovery-based approaches. Many of the teachers use speed games as a daily warm-up for math instruction. Teachers place a heavy emphasis on assessing the progress of their class before moving on to new concepts. All but one of the teachers use some form of technology in their math instruction. The most popular form of technology was educational websites and computer games. There is also frequent use of interactive whiteboards by two teachers.

The school's computer lab is made up of about 35 new Mac computers, and is attached to the school library. Computer stations are in rows, facing the front of the room, with an aisle running down the middle. The set up allows a teacher to be at the back of the room and have a view of every student's computer monitor.

The SuccessMaker teachers took their students to use the program in the computer lab three days a week for twenty minutes. SuccessMaker is generally used in addition to the core block of mathematics instruction. The median 3rd grade student used the math program approximately 17 hours, attempting 49 exercises every thirty minutes with a success rate of 66%. The median students in 5th grade used of the math program approximately 23 hours, attempting 47 exercises every thirty minutes with a success rate of 67%.

Indiana District

The participating Indiana school resides in the fringe of a large city. In 2008-09 the school district served a community of 12,000. The median household income is approximately $43,000 indicating a middle class community. The majority of the students from this Title 1 school are from lower socioeconomic backgrounds. The surrounding area has few opportunities for jobs and economic growth and the school district was recently forced to lay off 40 teachers due to budget shortfalls. Teachers say many of their students face a lot of uncertainty at home, and yet make big efforts to do well in school. Math scores have been low in the past, so the teachers were excited to see what impact SuccessMaker would have on their state math assessments.

The elementary school in Indiana is a mid-size school serving approximately 420 students in grades pre-kindergarten through five. The school has one primary ethnic group, Caucasian, making up a total of 91% of the school population. This school falls into the medium-high range for participation in the nation's free or reduced-price lunch program with 59% of students eligible to receive free or reduced-price lunch. The elementary school did meet AYP in the 2008, but due to the change to spring testing in 2009; AYP was not calculated for 2009.

The percentage of 3rd grade students testing at standard in mathematics in the 2008-09 school year was 56%, 13% lower than the statewide results. The percentage of 5th grade students testing at standard in mathematics in the 2008-09 school year was 69%, 8% lower than the statewide results. The percentage of 3rd grade students testing at standard in English Language Arts was 67%, coming in 7% lower than the statewide results. The percent of 5th grade students testing at

standard in English Language Arts was 74%, which is the same percentage as the statewide results. The student/teacher ratio is approximately 17 to 1.

A total of six teachers participated in the SuccessMaker study with two 3rd and two 5th grade classrooms randomly assigned to use the program. None of these teachers were new to the school or district. The district adopted a widely published elementary basal mathematics curriculum with a late copyright date. Five out of the six teachers report receiving training on this curriculum, while one teacher has received no training. Teachers have followed this curriculum for an average of 6 years. All teachers primarily use the district adopted program with some supplementation. This supplementation includes a paper-based math facts program, which is used by all teachers at the school. Teachers use a variety of other materials (i.e., additional worksheets, teacher-created activities) to supplement the rest of their instruction.

Daily math blocks range from 45 minutes to one hour and 10 minutes. All of the teachers have additional support in their classroom during their math block. Five out of six teachers have help in the form of a teacher's aid, and two of these teachers also have a student teacher. The sixth teacher receives support from a paraprofessional. Most teachers adhere to a math teaching philosophy that combines skills-based and discovery-based methods, though one 3rd grade teacher prefers purely skills-based teaching methods. Teachers conduct math lessons using whole group instruction about 76% of the time.

Only one teacher frequently used leveled instruction for math lessons, while one other teacher used this strategy occasionally. All teachers used cooperative learning to some degree for math instruction, though infrequently for most. All but one teacher reported using centers. Teachers also incorporate some technology use into the classroom during math instruction. All teachers frequently use interactive white boards and occasionally use educational computer games. All but two use instructional websites weekly.

Those teachers randomly assigned to use the SuccessMaker Math program were trained approximately three months after school began on November 13th. These teachers also received an additional training in January. Students completed baseline testing the third week in November and completed end-of-year testing the week of May 10th. Students' last week using the program was the week of May 21st.

The SuccessMaker teachers took their students to use the program in the computer lab two days a week for 30 minutes with the exception of one 5th grade teacher that took their students to the lab four times a week for fifteen minute sessions. Teachers used the program in addition to their block of mathematics instruction. The median 3rd grade student used the math program approximately 18 hours, attempting 48 exercises every thirty minutes with a success rate of 67%. The median 5th grade student used the math program approximately 17 hours, attempting 50 exercises every thirty minutes with a success rate of 66%.

Kansas District

The participating Kansas schools reside in a large city, which in 2009 had a population of more than 100,000. The median household income is approximately $40,000 indicating a middle class community. Two schools, one elementary school and one middle school participated in the SuccessMaker study from this Kansas district.

The elementary school in Kansas is a mid- to large size school serving approximately 400 students in grades kindergarten through five. This school has English language learning and dual-language classrooms, as well as a hearing-impaired program. This elementary school also follows an inclusion model. Most of the population is bused in. The students demonstrate a wide diversity in achievement. Caucasian students make up a total of 62% of the school population. Hispanic students make up the next largest portion of the population at 22%, with African-Americans next at 11%, and a small American-Indian group of 2%. This school falls into the medium range for participation in the nation's free or reduced-price lunch program with 48% of students eligible to receive free or reduced-price lunch. The student/teacher ratio is approximately 13 to 1.

The middle school is a mid-size school serving approximately 460 students in grades six through eight. Many of the students come from families that live in poverty and some are undocumented citizens. Caucasian and Hispanic students equally make up 80% of the school population. African-American students make up about 16% of the school population. American-Indian students make up the remaining 4% the student population. This school falls into the high range for participation in the nation's free or reduced-price lunch program with 87% of students eligible to receive free or reduced-price lunch. The student/teacher ratio is approximately 11 to 1.

The elementary school did meet AYP in the 2008-09 school year. The percentage of 3rd grade students testing at standard in mathematics in the 2008-09 school year was 81%, 5% lower than the statewide results. The percentage of 5th grade students testing at standard in mathematics in the 2008-09 school year was 84%, 3% lower than the statewide results. The percentage of 3rd grade students testing at standard in reading was 72%, coming in 12% lower than the statewide results. The percent of 5th grade students testing at standard in reading was 75%, 9% lower than the statewide results. The middle school did not meet AYP in the 2008-09 school year. The percentage of 7th grade students testing at standard in mathematics in the 2008-09 school year was 60%, 18% lower than the statewide results. The percent of 7th grade students testing at standard in reading was 65%, 21% lower than the statewide results.

Six teachers from the elementary school participated in the study, three 3rd grade teachers and three 5th grade teachers. Two 3rd and 5th grade classrooms were randomly assigned to use the SuccessMaker math program, and one 3rd and 5th grade classroom was assigned to the comparison group. At the middle school, two 7th grade teachers participated in the study. One 7th grade teacher's three classrooms were assigned to use SuccessMaker math and the other 7th grade teacher's three classrooms were assigned to the comparison group. None of these teachers were new to the school or district.

The district adopted a widely published elementary basal mathematics curriculum with an early copyright date for the elementary school. Five of the six teachers primarily use the district adopted curriculum with some supplementation, and one teacher strictly adheres to the district curriculum. Teachers have used this program for an average of four years. Teachers have received training from the district and many have received support from their peers to fill in any training gaps. Daily math blocks range in time from one hour to one-and-a-half hours. All teachers have some form of additional support in the classroom during math instruction. Four of the teachers have a paraprofessional in their classroom, and the other two have student teachers.

Additionally, one teacher has a hearing impaired interpreter in her room, and another has a special education teacher in her room.

All six elementary teachers prefer to use a combination of skills-based and discovery-based teaching methods. Teachers teach math whole group for about 60% of the time, and use small groups about 40% of the time. A number of classroom strategies were present during math instruction. Two of the six teachers frequently use leveled instruction, and half frequently use centers. All teachers used center activities and speed drills to some extent. Only one teacher reported using seatwork, and infrequently at that. As far as technology use for teaching math, educational websites and computer games were employed to varying degrees. One teacher reported frequently using their interactive white board.

The district adopted program for the middle school was a widely published basal math curriculum with an early copyright date. Both 7th grade teachers strictly adhere to the district curriculum and have been using the curriculum for an average of 6 years. The teachers have received training on the district adopted curriculum. The math daily blocks last an hour-and-a-half and both teachers have assistance from paraprofessionals in the classroom. Both teachers prefer to use a combination of skills-based and discovery-based teaching methods and teach whole group about 75% of the time. One of the teachers sometimes used centers for math instruction, and only occasionally used leveled instruction or cooperative learning; the other teacher did not use these teaching strategies. One teacher infrequently used educational websites and computer games while the other teacher used educational technology two to three times per week.

Those teachers randomly assigned to use the SuccessMaker Math program from both schools were trained two days before school began. These teachers also received additional trainings in September and December. Students in the elementary school were baseline tested the third week of September and tested again the second week of May. Students in the middle school were baseline tested the second week of September and were post tested the second week of May. Students in the elementary school stopped using the program the first week of May and students in the middle school stopped using the program the second week of May. These schools place a heavy emphasis on state testing and there is a lot of pressure for students to do well. This year, for the first time, all students were required to take the state assessments online, which limited the amount of time the 7th grade students had on the program during the second half of the school year.

The elementary school has a dedicated up-to-date computer lab with over forty stations. The elementary SuccessMaker teachers took their students to use the program in the computer lab three days a week for twenty minutes. Additionally, one 3rd grade teacher used the program on classroom stations about 30% of their total usage minutes. The median 3rd grade student used the math program approximately 27 hours, attempting 48 exercises every thirty minutes with a success rate of 73%, while the median 5th grade student used the math program approximately 33 hours, attempting 44 exercises every thirty minutes with a success rate of 74%.

The middle school had an older computer lab with at least 30 computer stations. The computers had to be updated in order to meet the technology requirements necessary to run the SuccessMaker program. The SuccessMaker teacher at the middle school took her students to use the program in the computer lab two days a week for 30 minutes as part of to their normal block

of mathematics instruction. The median student used the math program approximately 21 hours, attempting 43 exercises every thirty minutes with a success rate of 63%.

New York District

The participating New York elementary school resides in a suburb of NYC. In 2008-09 the school district served a community of 16,000. The median household income is approximately $74,000 indicating an upper-middle class community. Facilities are new at this school, as the building is only two years old. This school draws from a low-income community, but sets high standards for students, who are required to wear uniforms and demonstrate appropriate school behavior. The school has been recognized for best practices and as a "Closing the Gap" school. Teachers are very proud and supportive of their students and describe them as making "learning their priority" and "surpassing goals despite some of the difficulties they face."

The school is mid-sized serving approximately 400 students in grades kindergarten through five. The school has one primary ethnic group, African-American, making up a total of 57% of the school population. Hispanic students make up 39% of the school population. Multi-racial students make up the remaining 4% of the student population. This school falls into the high range for participation in the nation's free or reduced-price lunch program with 75% of students eligible to receive free or reduced-price lunch. Approximately 22% of the students are designated as not English proficient. The student/teacher ratio is approximately 19 to 1.

The elementary school did meet AYP in the 2008-09 school year. The percentage of 3rd grade students testing at standard in mathematics in the 2008-09 school year was 98%, 5% higher than the statewide results. The percentage of 5th grade students testing at standard in mathematics in the 2008-09 school year was 98%, 10% higher than the statewide results. The percentage of 3rd grade students testing at standard in English Language Arts was 85%, coming in 9% higher than the statewide results. The percent of 5th grade students testing at standard in English Language Arts was 100%, 18% higher than the statewide results.

There were four teachers that participated in the study from the New York elementary school. Out of two teachers in 3rd grade, one was randomly assigned to use the SuccessMaker math program, and one was assigned to the comparison group. In 5th grade, there were also two teachers and one was randomly assigned to the SuccessMaker group, while the other was assigned to the comparison group. None of these teachers were new to the school or district.

The district adopted a widely published elementary basal mathematics curriculum with a late copyright date. Teachers receive training on the curriculum about twice a year and have followed this curriculum for an average of 3 years. The degree of curriculum implementation varies by teacher. One teacher reported strict adherence to the district adopted curriculum, two teachers reported using some supplementation, and the fourth teacher reported using heavy supplementation.

Teachers at the New York school have daily math blocks ranging from 50 minutes to one hour. None of the teachers receive additional support in their classrooms during their math block. All teachers adhere to a math teaching philosophy that combines skills-based and discovery-based methods. Teachers conduct math lessons using whole group instruction about 60% of the time, and small group instruction about 40% of the time. Teachers reported using a number of strategies for math instruction including cooperative learning, speed drills, centers, and leveled

instruction. As for technology, all teachers used educational websites and computer games. One teacher frequently uses their interactive white board for math instruction.

The New York school has a good quality computer lab with over 40 Mac stations. The SuccessMaker teachers took their students to use the program in the computer lab three days a week for twenty minutes. The SuccessMaker classes used the math portion of the program in addition to their block mathematics instruction. The median student in the 3rd grade used the math program approximately 18 hours, attempting 42 exercises every thirty minutes with a success rate of 73%. The median student in the 5th grade used the math program approximately 16 hours, attempting 55 exercises every thirty minutes with a success rate of 67%.

Those teachers randomly assigned to use the SuccessMaker Math program were trained a few months after school began on December 8th. These teachers also received an additional training in February and March. Students completed baseline testing the second week of February and tested again the week of June 14th. Students' last week using the program was the week of June 22nd.

Pennsylvania District

The participating Pennsylvania school resides in a suburban area. In 2008-09 the school district served a community of 8,000. The median household income is approximately $40,000 indicating a middle class community. This is a brand new school that emphasizes technology. It is housed in a renovated building, which was once the local high school. The school day and year are extended, uniforms are required, and students are admitted based on a lottery system. The structure of the school requires high parent involvement, which in turn motivates the students to learn. Teachers describe having diverse classrooms in terms of learning abilities.

Three teachers participated in the study from the Pennsylvania school: two 3rd grade teachers and one 5th grade teacher. Out of the two 3rd grade teachers, one was randomly assigned to use the SuccessMaker math program, and the other was assigned to the comparison group. The participating 5th grade teacher had two classroom sections of math. One classroom was randomly chosen to use SuccessMaker math, and the other was assigned to the comparison group.

The district adopted a widely published elementary basal mathematics curriculum. None of the teachers strictly adhered to this curriculum. The 3rd grade teachers supplemented the basal program with an activity based program developed by a local retired teacher. Teachers have daily math blocks of one-and-a-half hours. The 3rd grade comparison and 5th grade teacher receive additional support in their classrooms during their math block. All teachers share a combined skills-based and discovery-based math teaching philosophy, all conduct math lessons using whole group and small group instruction in equal parts, and speed drills. Technology was very prevalent in math instruction. All teachers used various educational websites and computer games. Each teacher also used their interactive white board regularly.

Those teachers randomly assigned to use the SuccessMaker Math program were trained a few weeks after school began on August 26th. These teachers also received an additional training in October and March. Students completed baseline testing the first week of September and tested again the week of June 8th. Students' last week using the program was the first week of June. SuccessMaker classrooms used the math portion of the program in addition to their block mathematics instruction.

After initially using the program in the computer lab, the 3rd grade SuccessMaker students settled on using the program in the classroom three days a week for twenty minutes. Classroom use accounted for 75% of the total usage time. The 5th grade SuccessMaker students used the program in the computer lab four days a week for fifteen minutes. The median 3rd grader used the math program approximately 24 hours, attempting 42 exercises every thirty minutes with a success rate of 72%. The median 5th grade student used the math program much less, approximately 9 hours, attempting 43 exercises every thirty minutes with a success rate of 71%.

Table 3		Gatti Evaluation SuccessMaker Math RCT Sample Demographic Information							
Group	Grade	[1]Student Count	[2]Percent One Grade Equivalent Below	Percent Not English Proficient	Percent Reduced Lunch	Percent Caucasian	Percent Hispanic/ Native American	Percent African American/ Caribbean	Other Ethnicity or No Information
Arizona District 1									
SM	3	30 (73%)	47%	57%	97%	17%	63%	13%	7%
Comparison		15 (68%)	53%	33%	100%	13%	80%	7%	0%
SM	7	22 (69%)	55%	23%	91%	9%	77%	14%	0%
Comparison		42 (72%)	57%	24%	93%	14%	79%	0%	7%
Arizona District 2									
SM	3	44 (69%)	45%	48%	86%	5%	93%	2%	0%
Comparison		43 (68%)	74%	53%	95%	5%	84%	9%	2%
SM	5	38 (76%)	45%	50%	87%	5%	87%	5%	3%
Comparison		42 (82%)	83%	38%	86%	10%	86%	2%	2%
SM	7	67 (74%)	46%	33%	84%	4%	91%	4%	1%
Comparison		59 (81%)	53%	25%	95%	7%	86%	5%	2%
Arkansas District									
SM	3	64 (97%)	8%	2%	9%	89%	2%	5%	4%
Comparison		43 (96%)	7%	2%	21%	95%	2%	0%	3%
SM	5	49 (98%)	12%	4%	12%	88%	4%	2%	6%
Comparison		43 (91%)	9%	7%	12%	93%	0%	0%	7%
California District									
SM	3	38 (90%)	68%	0%	79%	0%	100%	0%	0%
Comparison		57 (95%)	35%	0%	78%	0%	100%	0%	0%
SM	5	24(100%)	46%	0%	63%	4%	96%	0%	0%
Comparison		24 (96%)	46%	0%	63%	8%	92%	0%	0%
Indiana District									
SM	3	29 (76%)	31%	0%	69%	93%	0%	0%	7%
Comparison		16 (89%)	50%	6%	69%	88%	6%	0%	6%

SM	5	36 (90%)	47%	8%	83%	83%	3%	3%	11%
Comparison		16 (84%)	25%	0%	69%	100%	0%	0%	0%
Kansas District									
SM	3	41 (95%)	44%	0%	56%	54%	24%	15%	7%
Comparison		19 (95%)	63%	0%	79%	47%	37%	11%	5%
SM	5	43 (98%)	28%	0%	49%	63%	26%	7%	4%
Comparison		20 (91%)	35%	0%	60%	60%	10%	25%	5%
SM	7	48 (81%)	74%	0%	83%	45%	32%	23%	0%
Comparison		36 (80%)	50%	0%	78%	42%	33%	19%	6%
New York District									
SM	3	15 (94%)	27%	0%	87%	0%	60%	40%	0%
Comparison		11 (79%)	73%	0%	100%	0%	64%	36%	0%
SM	5	13 (68%)	62%	0%	85%	0%	54%	46%	0%
Comparison		20 (95%)	90%	0%	90%	0%	45%	55%	0%
Pennsylvania District									
SM	3	21 (91%)	38%	0%	62%	43%	0%	52%	5%
Comparison		19 (90%)	32%	0%	58%	42%	0%	58%	0%
SM	5	21 (91%)	38%	0%	81%	38%	0%	62%	0%
Comparison		19 (90%)	0%	0%	58%	42%	0%	58%	0%

1. Percents within parentheses next to student counts indicate the percent of students tested at baseline that were also tested at the end of the school year.
2. Study sample was broken out by baseline GMADE national norm cutoff score for 1.0 grade equivalent below grade and month at the time of testing.

Participants

The final diverse sample consisted of 1,186 3rd, 5th, and 7th grade students from eight school districts in seven states located in different regions of the US.

The research team recruited sixty-three diverse 3rd, 5th and 7th grade classrooms from eight urban and suburban school districts in seven different states (i.e., AZ, AR, CA, IN, KS, NY, PA). The final study sample consisted of 505 3rd grade (i.e., SuccessMaker = 282, comparison = 223), 408 5th grade (i.e., SuccessMaker = 224, comparison = 184) and 273 7th grade (i.e., SuccessMaker = 136, comparison = 137) students. It can be seen from Table 3 that the AZ and NY sites had considerable attrition. These three districts have a highly transient population and thus had comparatively high attrition. Eighty-five percent of the 3rd grade students tested at baseline remained in the final study sample (i.e., SuccessMaker = 85%, comparison = 85%). Likewise, 80% of the 5th grade (i.e., SuccessMaker = 74%, comparison = 89%) and 71% of the 7th grade (i.e., SuccessMaker = 65%, comparison = 78%) students tested at baseline remained in the final study sample.

It can be also be seen from Table 3 the study sites show considerable variation in math achievement and ethnicity, as well as percent of students eligible for reduced priced lunch.

Although, overall low-achieving at baseline (i.e., 3rd = 40%, 5th = 39%, 7th = 55% one grade equivalent below), the study groups do not statistically vary on baseline achievement at the three grade levels. Also, the study groups at the three grades did not vary in percent of English proficient students (i.e., 3rd = 86%, 5th = 90%, 7th = 81%). The percent of the students eligible to receive free or reduced priced lunch was high (i.e., 3rd = 68%, 5th = 63%, 7th = 87%) and statistically different at 3rd grade (i.e., SuccessMaker = 63%, comparison = 74%). The sample also tended to be heavily Hispanic (i.e., 3rd Hispanic = 47%, Caucasian = 39%, African-American = 11%; 5th Hispanic = 36%, Caucasian = 47%, African-American = 13%; 7th Hispanic = 69%, Caucasian = 19%, African-American = 10%).

Data Analysis Procedures

Statistical analyses were performed on students' end-of-year GMADE Total score and subtests, as well as, math academic attitude survey raw scores for each grade level. Results were also broken out and analyzed for separate levels of four key demographic variables (i.e., English proficiency, ethnicity, gender, meal status[8]). In addition, results were calculated for those students performing one grade equivalent below their current grade and month at the time of testing. Further, the performance for the comparison group was compared to four blocks of program usage (i.e., block 1 = 1 to 9 hours, block 2 = 10 to 19 hours, block 3 = 20 to 29 hours, block 4 = 30 or more hours).

Statistical analyses were performed on students' end-of-year GMADE Total and subtests, as well as, academic attitude survey scores for the three grade levels. Results were also broken out and analyzed for key subpopulations of students.

Rigorous research design dictates that all characteristics of the study participants and their environmental influences that may impact the results must be equated across study groups. This is advised even when classrooms of students are randomly assigned to study groups. Random assignment can only probabilistically equate study groups prior to the start of the study. The statistical equating of confounding factors and maintaining a controlled and consistent environment for the study participants ensures that differences found in the study groups on outcomes of interest may more confidently be attributed to the study conditions assigned to these groups.

Comparisons were made between study groups (i.e., comparison vs. SuccessMaker) using model adjusted group mean differences. Model adjusted group mean differences were calculated holding all covariates constant in an attempt to statistically equate the study groups on those constructs and remove their influence from the study group effect. Covariates included baseline scores, student demographic,[9] and 2009-2010 school year classroom environment indicators.[10]

[8] The CA site could not provide meal program status for individual students. The CA site did, however, provide the percent of students receiving free or reduced priced lunch in each classroom. Participation in the meal program for each student was estimated by choosing the most likely participants as determined via the EM algorithm using all available known student and classroom level information.

[9] gender, meal program status, ethnicity, English proficiency

[10] teacher education and experience, classroom assistance, teacher substitution, regular math instruction in minutes, classroom demographics, class size, baseline classroom achievement and variation, testing time span, program usage time span, current curricular choices, basal curricula adherence, years using basal curricula, self report of frequency of use of specific teaching strategies (i.e., leveled instruction, cooperative learning/peer tutoring, center rotations, speed drills/math facts, test preparation, progress monitoring)

When results are broken out by a demographic variable or a grouping indicator, such as the below one grade level designation, the group mean difference is no longer adjusted by that variable along with the remaining model covariates, rather, these differences are separated by the levels of that variable.

A random intercepts model was employed to estimate and test model adjusted group mean differences. While students were the unit of analysis, the nine school districts were the independent units. The hierarchical nature of the data (i.e., students nested within classrooms, classrooms nested within schools, schools nested within districts) has the effect of reducing the amount of independent information available in the sample, therefore decreasing the precision of estimates and the power of hypothesis tests to find these estimates statistically significant.[11] A naïve covariance structure[12] within a robust empirical standard error formulation was used to calculate confidence intervals for estimated effects. The result of this procedure is group mean differences are unbiased and statistical hypothesis tests are consistent[13,14] despite the nested nature of data.

All statistical significance tests are two-tailed, with a Type I error rate of 0.05. Statistically significant estimates mean the probability of sampling scores that result in a value that much greater than zero, when it is in fact null, is p = 0.05 or 1 in 20 samples. Statistical significance implies that the samples are likely drawn from two separate populations or that the group averages are unlikely to be the same in the population. Standardized effect size estimates (i.e., effect size = estimated adjusted group difference / comparison sample standard deviation) are computed for statistically significant model adjusted group mean differences using the sample standard deviation for the comparison group's end-of-year scores.[15] The statistical models were able to find moderate to large effect sizes statistically significant. The average minimal detectable effect sizes for 3rd, 5th, and 7th grade were 0.39, 0.29, and 0.43 respectively. Effect sizes as large as these are most likely of practical significance. The careful review of efficacy studies for educational materials[16] indicate that the average adjusted group mean difference for studies with large samples (i.e., more than 250 students) is only 0.13 standard deviations.

[11] Donnar, A. & Klar, N. (2000) *Design and analysis of cluster randomization trials in health research*. Arnold Publishers, London.
[12] Initially a compound symmetric structure was assumed for the error variances but the extra parameter was not statistically significant for any of the statistical models.
[13] Liang, N. M. & and Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika, 73*, pp. 13-22.
[14] SAS's Mixed procedure was used to analyze the data, see SAS Institute Inc. (2008) Online documentation 9.2. A linear model was defined with all fixed effects, full degrees of freedom (i.e., N-2), using the sandwich estimator for all standard errors with districts set as the subject or independent level of nesting and a naïve, independent working covariance structure.
[15] Hedges, L. V. & Olkin, I. (1985). *Statistics methods for meta-analysis*. Academic Press, NY.
[16] Slavin, R. & Smith, D. (2009). The relationship between sample sizes and effect sizes in systematic reviews in education. *Educational Evaluation and Policy Analysis, 31*(*4*) pp. 500-506.

III. RESULTS

Report section III summarizes the results of data analyses, including statistical and qualitative results, and group comparisons at baseline. The first subsection demonstrates the closeness of the samples on the quantitative outcome measures at baseline. The second subsection addresses research question one, comparing achievement for the SuccessMaker group to that of the comparison group. Section two further addresses achievement for increasing levels of SuccessMaker usage. The third subsection then breaks out the SuccessMaker v. comparison group achievement results by subpopulations.

The fourth and fifth subsections address both research questions two and three. That is, do SuccessMaker students demonstrate more positive attitudes toward mathematics and mathematics instruction, and, how did teachers and students react to the program? Section five summarizes comments collected from SuccessMaker teachers during focus groups interviews and end-of-year student SuccessMaker opinion surveys.

Baseline Group Equivalence

Tables 4-7 present both the simple sample[17] and model adjusted[18] baseline group mean differences for each measure of achievement and attitude for 3rd, 5th and 7th grade classrooms. These tables also show statistical significance test results and effect size measures for the baseline group mean differences. No achievement or attitude outcomes were statistically significantly different between the study groups at baseline, and no effects were of practical significance.

Table 4	Third Grade Baseline GMADE Score Study Group Comparisons						
Measure	Sample Size SM/CP	Sample Difference	Sample p-value	Sample Effect Size	Adjusted Difference	Adjusted p-value	Adjusted Effect Size
GMADE Overall	505	1.59	0.5047	0.11	0.54	0.8188	0.04
GMADE Subtest 1	505	0.57	0.4004	0.12	0.12	0.8615	0.03
GMADE Subtest 2	505	0.41	0.6891	0.07	0.12	0.9028	0.02
GMADE Subtest 3	505	0.63	0.4804	0.10	0.29	0.7447	0.04
Adjusted baseline group mean differences are estimated holding student demographic variables constant across groups. Sample group mean differences are estimated allowing student demographics to vary as they were sampled and randomly assigned.							

[17] Sample group mean differences are estimated allowing student demographics to vary as they were sampled and randomly assigned.
[18] Adjusted baseline group mean differences are estimated holding student demographic variables constant across groups.

Table 5	Fifth Grade Baseline GMADE Score Study Group Comparisons						
Measure	Sample Size SM/CP	Sample Difference	Sample p-value	Sample Effect Size	Adjusted Difference	Adjusted p-value	Adjusted Effect Size
GMADE Overall	408	2.77	0.2280	0.19	1.96	0.3792	0.13
GMADE Subtest 1	408	1.08	0.1853	0.20	0.74	0.3274	0.13
GMADE Subtest 2	408	1.00	0.2197	0.20	0.71	0.3631	0.14
GMADE Subtest 3	408	0.70	0.3833	0.12	0.51	0.5323	0.09
Adjusted baseline group mean differences are estimated holding student demographic variables constant across groups. Sample group mean differences are estimated allowing student demographics to vary as they were sampled and randomly assigned.							

Table 6	Seventh Grade Baseline GMADE Score Study Group Comparisons						
Measure	Sample Size SM/CP	Sample Difference	Sample p-value	Sample Effect Size	Adjusted Difference	Adjusted p-value	Adjusted Effect Size
GMADE Overall	273	-0.44	0.6767	-0.04	-0.27	0.7987	-0.02
GMADE Subtest 1	273	-0.57	0.4059	-0.12	-0.56	0.4166	-0.12
GMADE Subtest 2	273	0.02	0.9676	0.00	0.05	0.9122	0.01
GMADE Subtest 3	273	0.12	0.6703	0.03	0.24	0.4294	0.06
Adjusted baseline group mean differences are estimated holding student demographic variables constant across groups. Sample group mean differences are estimated allowing student demographics to vary as they were sampled and randomly assigned.							

Table 7	Baseline Math Academic Attitude Survey Score Comparisons						
Grade	Sample Size SM/CP	Sample Difference	Sample p-value	Sample Effect Size	Adjusted Difference	Adjusted p-value	Adjusted Effect Size
Grade 3 Survey	497	-0.01	0.9678	0.00	-0.07	0.8365	-0.02
Grade 5 Survey	406	0.78	0.0958	0.15	0.68	0.1645	0.13
Grade 7 Survey	269	-0.16	0.7053	-0.03	0.03	0.9467	0.01
Adjusted baseline group mean differences are estimated holding student demographic variables constant across groups. Sample group mean differences are estimated allowing student demographics to vary as they were sampled and randomly assigned.							

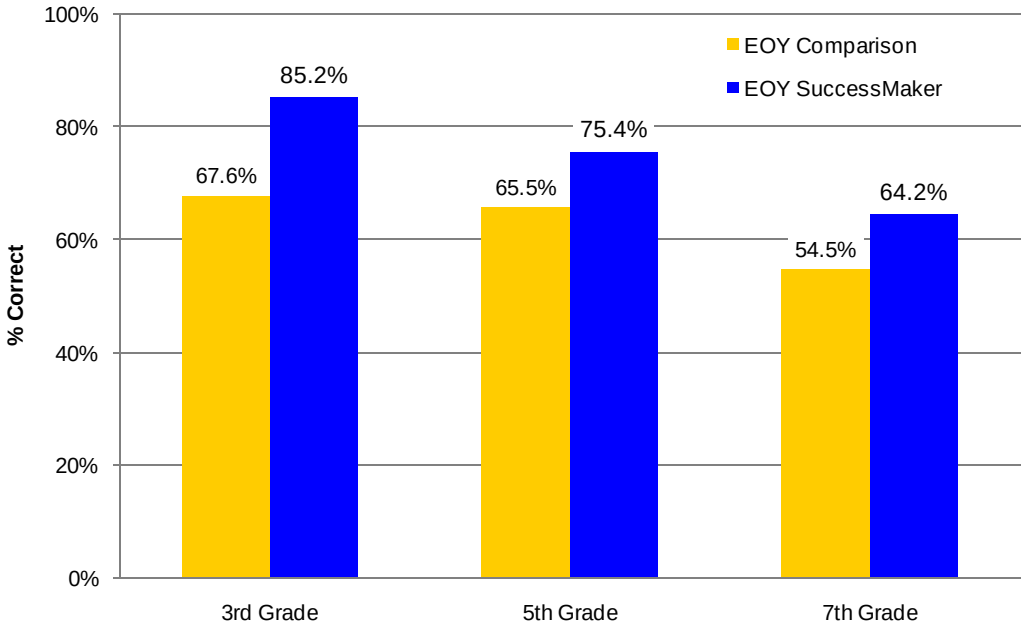
Group Comparisons of Achievement Gains

This section will address research question one:

RQ1: Do 3rd, 5th, and 7th grade students making regular use of the SuccessMaker Math program demonstrate higher mathematics achievement as compared to students that did not utilize SuccessMaker Math?

Figures 1 through 4 present the SuccessMaker and comparison model adjusted group mean differences on the GMADE total and subtest scores.

Figure 1: GMADE Total



After adjusting for student & classroom characteristics, 3rd, 5th & 7th grade SuccessMaker Math users out scored their comparison group counterparts by 17.5% (SE=2.19%), 10.0% (SE=2.72%) and 9.8% (SE=2.23%) respectively.

Figure 2: GMADE Concepts and Communication

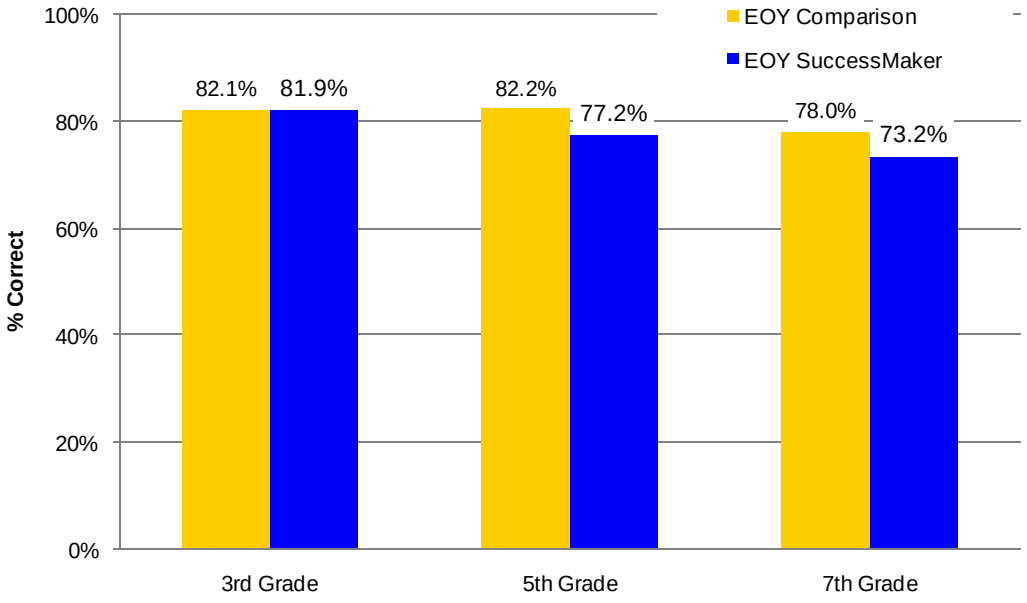
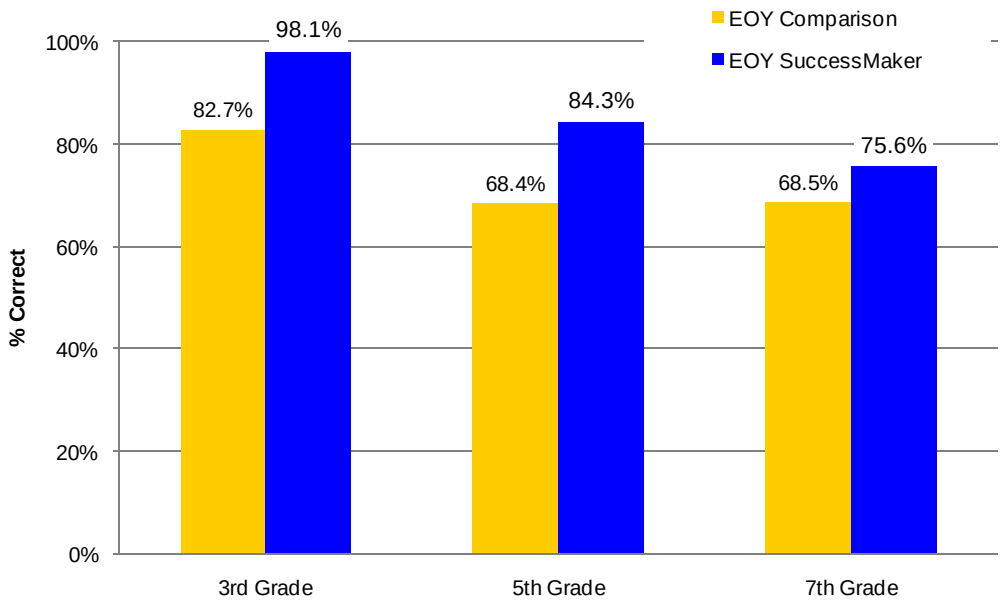
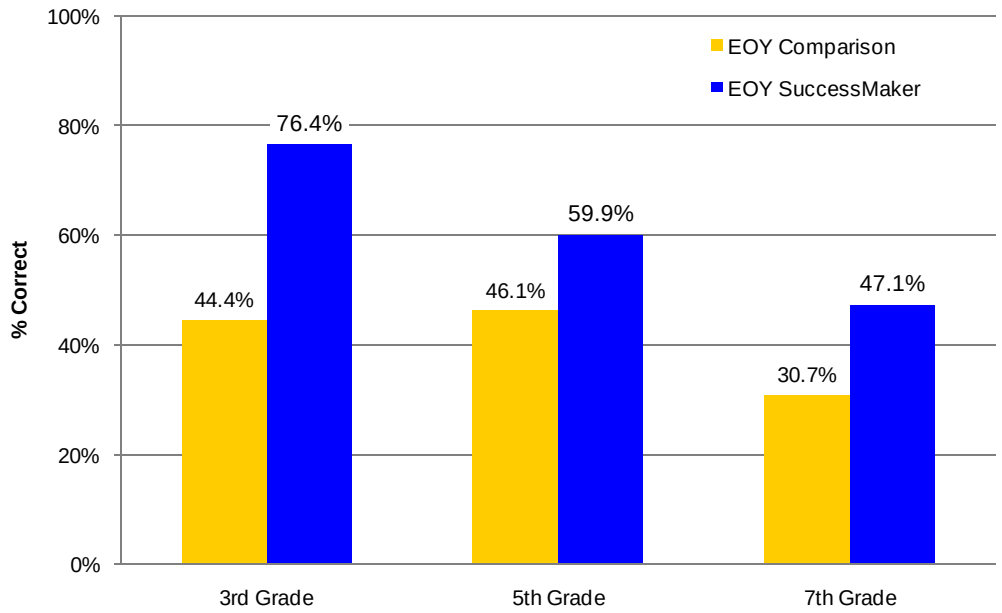


Figure 3: GMADE Operations and Computation



After adjusting for student and classroom characteristics, 3rd, 5th and 7th grade SuccessMaker Math users outscored their comparison group counterparts by 15.4% (SE=4.95%), 15.9% (SE=3.65%) and 7.0% (SE=5.46%) respectively.

Figure 4: GMADE Process and Applications



After adjusting for student and classroom characteristics, 3rd, 5th and 7th grade SuccessMaker Math users outscored their comparison group counterparts by 32.0% (SE=3.71%), 13.8% (SE=2.58%) and 16.4% (SE=3.62%) respectively.

SuccessMaker students in 3rd, 5th, and 7th grade statistically significantly outperformed their comparison group counterparts on the GMADE Total score. The magnitude of the difference in performance observed at all three grades was remarkable, 1.00, 0.53, and 0.61 standard deviations for 3rd, 5th, and 7th grade respectively. These effects were consistently large across usage levels. The comparative effect sizes observed for the blocks of program usage (i.e., block 1 = 1 to 9 hours, block 2 = 10 to 19 hours, block 3 = 20 to 29 hours, block 4 = 30 or more hours) can be found in Appendix 1.

3rd Grade Scale	GMADE Effect Size[1,2]
GMADE Total	1.00
Concepts and Communication	***
Operations and Computation	0.75
Process and Applications	1.32

*** Indicates group means are not statistically significantly different

1. effect size = estimated adjusted group difference / comparison sample standard deviation

2. The average effect size for studies with large samples (i.e., more than 250 students) has been recently estimated at 0.13 standard deviations.

5th Grade Scale	GMADE Effect Size[1,2]
GMADE Total	0.53
Concepts and Communication	-0.29
Operations and Computation	0.75
Process and Applications	0.59

1. effect size = estimated adjusted group difference / comparison sample standard deviation

2. The average effect size for studies with large samples (i.e., more than 250 students) has been recently estimated at 0.13 standard deviations.

7th Grade Scale	GMADE Effect Size[1,2]
GMADE Total	0.61
Concepts and Communication	***
Operations and Computation	***
Process and Applications	1.01

*** Indicates group means are not statistically significantly different

1. effect size = estimated adjusted group difference / comparison sample standard deviation

2. The average effect size for studies with large samples (i.e., more than 250 students) has been recently estimated at 0.13 standard deviations.

After adjusting for student & classroom characteristics, 3rd, 5th & 7th grade SuccessMaker Math users statistically outperformed their comparison group peers on the Process and Applications subtest by 32.0% (SE=3.71%), 13.8% (SE=2.58%) and 16.4% (SE=3.62%) correct respectively. The magnitude of the difference in performance observed at all three grades was very large, 1.32, 0.59, and 1.01 standard deviations for 3rd, 5th, and 7th grade respectively. These effects were also consistently large across usage levels.

The 3rd, 5th, and 7th grade SuccessMaker Math students statistically significantly outperformed the comparison group students on the GMADE Process and Applications subtest by a staggering 1.32, 0.59, and 1.01 standard deviations respectively.

Similarly, 3rd and 5th grade SuccessMaker Math users statistically outperformed their comparison group counterparts on the Operations and Computation subtest by 15.4% (SE=4.95%) and 15.9% (SE=3.65%) correct respectively. The magnitude of the differences in performance observed at both grades were equivalently very large, 0.75 standard deviations. And yet again, these effects were consistently large across usage levels. The 7th grade SuccessMaker students performed statistically the same as the comparison group on this subtest.

Finally, the SuccessMaker students in 3rd and 7th grade performed similarly to their comparison peers on the Concepts and Communication subtest. The 5th grade comparison group performed statistically significantly greater than 5th grade SuccessMaker students on this subtest.

Group Comparisons by Subpopulations

When the data was broken out for student subpopulations, 3rd grade Hispanic, low SES, non-English proficient, female, and lower-achieving SuccessMaker students all statistically significantly outperformed their comparison group peers on GMADE Total score (i.e., 0.50 to 1.31 standard deviations), as well as the Process and Applications (i.e., 0.91 to 1.65 standard deviations) and the Operations and Computation subtests (i.e., 0.49 to 1.19 standard deviations). The 3rd graders performed statistically similar on the Concepts and Communication subtest.

3rd Grade Subpopulation	GMADE Effect Size[1,2]
Lower achieving	0.50
Male	0.98
Female	1.06
Reduced priced lunch	1.01
Full priced lunch	0.82
Not English proficient	1.31
English proficient	0.88
African American	***
Hispanic	0.95
Caucasian	0.64

*** Indicates group means are not statistically significantly different.

1. effect size = estimated adjusted group difference / comparison sample standard deviation

2. The average effect size for studies with large samples (i.e., more than 250 students)

has been recently estimated at 0.13 standard deviations.

3rd Grade Subpopulation	Concepts and Communication Effect Size[1,2]
Lower achieving	***
Male	***
Female	***
Reduced priced lunch	***
Full priced lunch	***
Not English proficient	***
English proficient	***
African American	***
Hispanic	***
Caucasian	***

*** Indicates group means are not statistically significantly different.

1. effect size = estimated adjusted group difference / comparison sample standard deviation

2. The average effect size for studies with large samples (i.e., more than 250 students) has been recently estimated at 0.13 standard deviations.

3rd Grade Subpopulation	Operations and Computation Effect Size[1,2]
Lower achieving	0.49
Male	0.72
Female	0.79
Reduced priced lunch	0.76
Full priced lunch	***
Not English proficient	1.19
English proficient	0.60
African American	***
Hispanic	0.72
Caucasian	***

*** Indicates group means are not statistically significantly different.

1. effect size = estimated adjusted group difference / comparison sample standard deviation

2. The average effect size for studies with large samples (i.e., more than 250 students) has been recently estimated at 0.13 standard deviations.

3rd Grade Subpopulation	Process and Applications Effect Size[1,2]

Lower achieving	0.91
Male	1.26
Female	1.35
Reduced priced lunch	1.34
Full priced lunch	1.25
Not English proficient	1.65
English proficient	1.29
African American	1.52
Hispanic	1.41
Caucasian	1.18

1. effect size = estimated adjusted group difference / comparison sample standard deviation

2. The average effect size for studies with large samples (i.e., more than 250 students) has been recently estimated at 0.13 standard deviations.

Low SES, non-English proficient and female 5th grade SuccessMaker students statistically significantly outperformed their comparison group peers on GMADE Total score (i.e., 0.48 to 0.53 standard deviations), as well as, both the Process and Applications (i.e., 0.49 to 0.63 standard deviations) and Operations and Computation subtests (i.e., 0.55 to 0.73 standard deviations). In addition, 5th grade African-American students using SuccessMaker statistically outperformed their peers not using SuccessMaker on the Process and Applications subtest.
Conversely, 5th grade African-American comparison group students statistically outscored the SuccessMaker group on the Concepts and Communication subtest.

5th Grade Subpopulation	GMADE Effect Size[1,2]
Lower achieving	***
Male	0.60
Female	0.49
Reduced priced lunch	0.53
Full priced lunch	0.50
Not English proficient	0.48
English proficient	0.55
African American	***
Hispanic	***
Caucasian	0.58

*** Indicates group means are not statistically significantly different.

1. effect size = estimated adjusted group difference / comparison sample standard deviation

2. The average effect size for studies with large samples (i.e., more than 250 students) has been recently estimated at 0.13 standard deviations.

5th Grade Subpopulation	Concepts and Communication Effect Size[1,2]
Lower achieving	***
Male	***
Female	***
Reduced priced lunch	***
Full priced lunch	-0.40
Not English proficient	***
English proficient	-0.25
African American	-0.48
Hispanic	***
Caucasian	***

*** Indicates group means are not statistically significantly different.

1. effect size = estimated adjusted group difference / comparison sample standard deviation

2. The average effect size for studies with large samples (i.e., more than 250 students) has been recently estimated at 0.13 standard deviations.

5th Grade Subpopulation	Operations and Computation Effect Size[1,2]
Lower achieving	***
Male	0.81
Female	0.70
Reduced priced lunch	0.73
Full priced lunch	0.73
Not English proficient	0.55
English proficient	0.77
African American	***
Hispanic	***
Caucasian	0.88

*** Indicates group means are not statistically significantly different.

1. effect size = estimated adjusted group difference / comparison sample standard deviation

2. The average effect size for studies with large samples (i.e., more than 250 students) has been recently estimated at 0.13 standard deviations.

5th Grade Subpopulation	Process and Applications Effect Size[1,2]
Lower achieving	***

Male	0.64
Female	0.52
Reduced priced lunch	0.63
Full priced lunch	0.57
Not English proficient	0.49
English proficient	0.64
African American	0.61
Hispanic	***
Caucasian	0.68

*** Indicates group means are not statistically significantly different.

1. effect size = estimated adjusted group difference / comparison sample standard deviation

2. The average effect size for studies with large samples (i.e., more than 250 students) has been recently estimated at 0.13 standard deviations.

Seventh grade low SES, non-English proficient, and female students all dramatically outperformed their comparison group counterparts on GMADE Total score (i.e., 0.57 to 0.66 standard deviations) and the Process and Applications subtest (i.e., 1.06 to 1.39 standard deviations). Further, lower-achieving and Hispanic 7th grade SuccessMaker students statistically outperformed their comparison group peers on the Process and Applications subtest (i.e., 0.58 and 1.19 standard deviations). The study groups scored statistically the same for all 7th grade populations on the Concepts and Communication and the Operations and Computation subtests.

7th Grade Subpopulation	GMADE Effect Size[1,2]
Lower achieving	***
Male	0.61
Female	0.66
Reduced priced lunch	0.57
Full priced lunch	0.78
Not English proficient	0.60
English proficient	0.57
African American	***
Hispanic	***
Caucasian	***

*** Indicates group means are not statistically significantly different.

1. effect size = estimated adjusted group difference / comparison sample standard deviation

2. The average effect size for studies with large samples (i.e., more than 250 students) has been recently estimated at 0.13 standard deviations.

7th Grade Subpopulation	Concepts and Communication Effect Size[1,2]
Lower achieving	***
Male	***
Female	***
Reduced priced lunch	***
Full priced lunch	***
Not English proficient	***
English proficient	***
African American	***
Hispanic	***
Caucasian	***

*** Indicates group means are not statistically significantly different.

1. effect size = estimated adjusted group difference / comparison sample standard deviation

2. The average effect size for studies with large samples (i.e., more than 250 students) has been recently estimated at 0.13 standard deviations.

7th Grade Subpopulation	Operations and Computation Effect Size[1,2]
Lower achieving	***
Male	***
Female	***
Reduced priced lunch	***
Full priced lunch	***
Not English proficient	***
English proficient	***
African American	***
Hispanic	***
Caucasian	***

*** Indicates group means are not statistically significantly different.

1. effect size = estimated adjusted group difference / comparison sample standard deviation

2. The average effect size for studies with large samples (i.e., more than 250 students) has been recently estimated at 0.13 standard deviations.

7th Grade Subpopulation	Process and Applications Effect Size[1,2]
Lower achieving	0.58

Male	0.85
Female	1.14
Reduced priced lunch	1.06
Full priced lunch	0.80
Not English proficient	1.39
English proficient	0.99
African American	***
Hispanic	1.19
Caucasian	***

*** Indicates group means are not statistically significantly different.

1. effect size = estimated adjusted group difference / comparison sample standard deviation

2. The average effect size for studies with large samples (i.e., more than 250 students) has been recently estimated at 0.13 standard deviations.

Student Academic Attitudes

SuccessMaker Math students at 3rd and 7th grade demonstrated statistically higher attitudes than their comparison group counterparts. These very large effects were also seen for several at-risk populations.

This section will attempt to answer research question two:

RQ2: Do 3rd, 5th, and 7th grade students using the SuccessMaker Math program demonstrate more positive attitudes toward mathematics and mathematics instruction as their comparison group counterparts?

Figure 5 presents the average model adjusted math attitude survey score mean differences. The 3rd and 7th grade SuccessMaker students both had statistically significantly higher math academic attitudes than the comparison group (i.e., 3rd = 0.99 standard deviations, 7th = 0.62 standard deviations). The 5th grade SuccessMaker students had similar attitudes to their peers not using SuccessMaker.

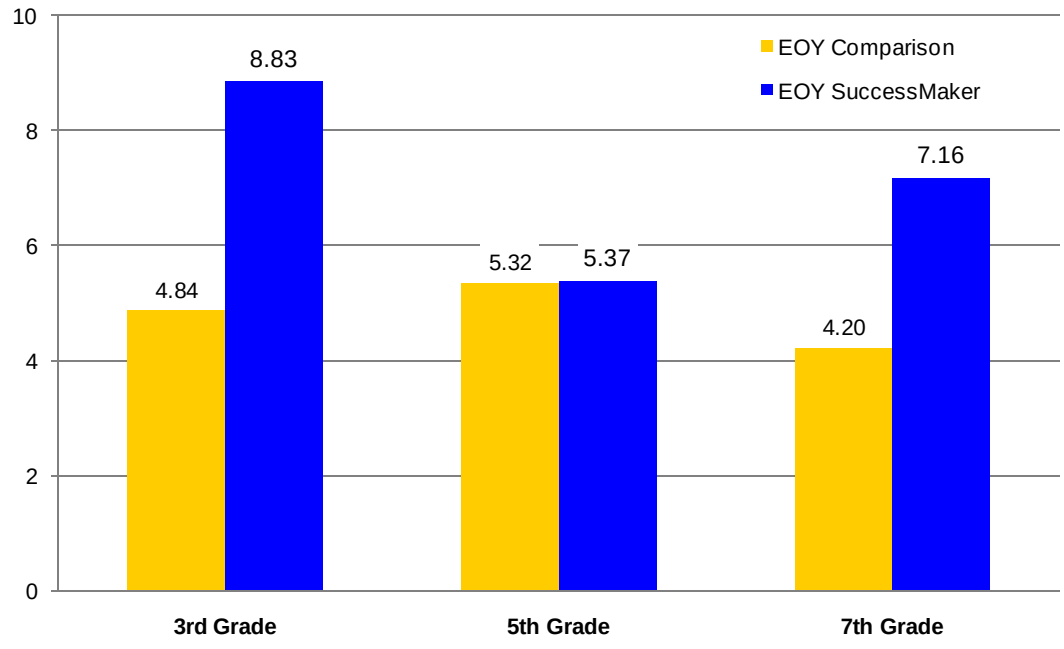
Student Math Attitude Scale	Effect Size[1,2]
3rd Grade	0.99
5th Grade	***
7th Grade	0.62

*** Indicates group means are not statistically significantly different

1. effect size = estimated adjusted group difference / comparison sample standard deviation

2. The average effect size for studies with large samples (i.e., more than 250 students) has been recently estimated at 0.13 standard deviations.

Figure 5: Math Academic Attitude Survey



After adjusting for student and classroom characteristics, 3rd, 5th & 7th grade SuccessMaker Math users outscored their comparison group counterparts by 3.99 (SE=1.52), 0.06 (SE=0.99) and 2.95 (SE=1.43) respectively.

The very large effects seen at 3rd grade were consistent for students in at-risk populations or Hispanic, lower SES, not English proficient, female, and lower achieving students (i.e., 0.29 to 1.13 standard deviations).

3rd Grade Subpopulation	Student Math Attitude Effect Size[1,2]
Lower achieving	0.29
Male	0.91
Female	0.96
Reduced priced lunch	1.03
Full priced lunch	***
Not English proficient	1.13
English proficient	0.95
African American	***
Hispanic	0.98
Caucasian	***

*** Indicates group means are not statistically significantly different

1. Cohen's *d* effect size = estimated group difference / comparison sample standard deviation

2. The average effect size for studies with large samples (i.e., more than 250 students) has been recently estimated at 0.13 standard deviations.

The 5th grade SuccessMaker and comparison group students, across all populations, had similar attitudes.

5th Grade Subpopulation	Student Math Attitude Effect Size[1,2]
Lower achieving	***
Male	***
Female	***
Reduced priced lunch	***
Full priced lunch	***
Not English proficient	***
English proficient	***
African American	***
Hispanic	***
Caucasian	***

*** Indicates group means are not statistically significantly different

1. Cohen's *d* effect size = estimated group difference / comparison sample standard deviation

2. The average effect size for studies with large samples (i.e., more than 250 students) has been recently estimated at 0.13 standard deviations.

Several 7th grade at-risk populations (i.e., female, lower SES, not English proficient) had statistically higher math attitudes than the comparison group (i.e., 0.61 to 0.69 standard deviations).

7th Grade Subpopulation	Student Math Attitude Effect Size[1,2]
Lower achieving	***
Male	***
Female	0.63
Reduced priced lunch	0.69
Full priced lunch	***
Not English proficient	0.61
English proficient	***
African American	***

Hispanic	***
Caucasian	***

*** Indicates group means are not statistically significantly different

1. Cohen's *d* effect size = estimated group difference / comparison sample standard deviation

2. The average effect size for studies with large samples (i.e., more than 250 students) has been recently estimated at 0.13 standard deviations.

The comparative effect sizes observed for the blocks of program usage can be found in Appendix 1.

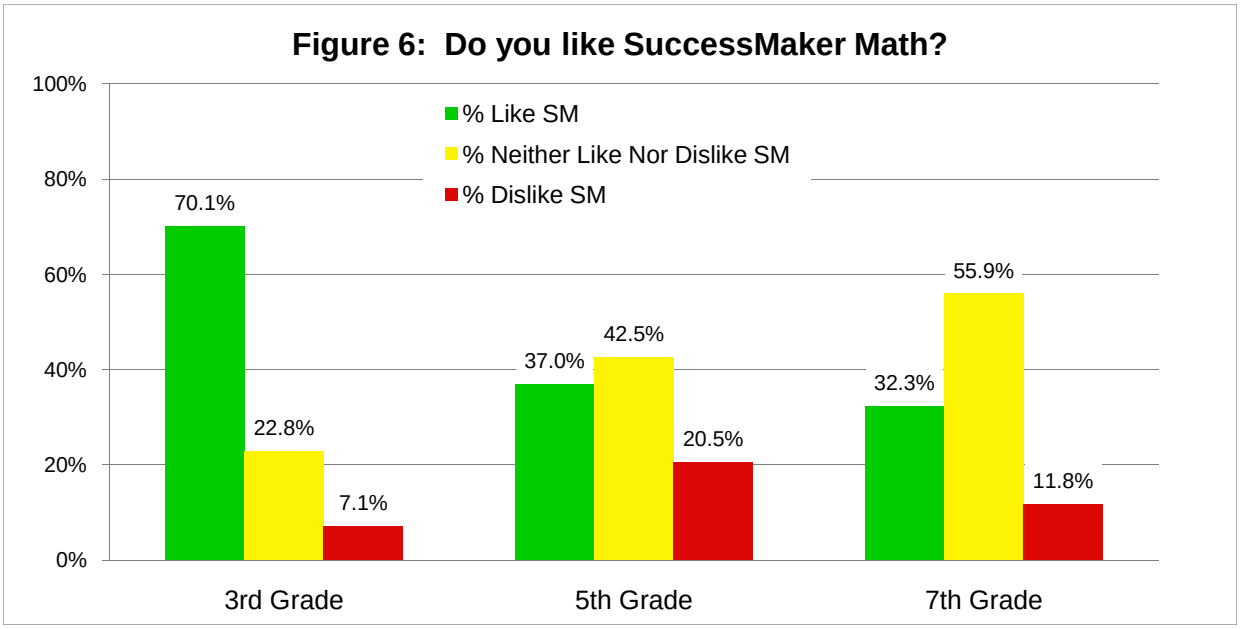
Teacher and Student SuccessMaker Opinions

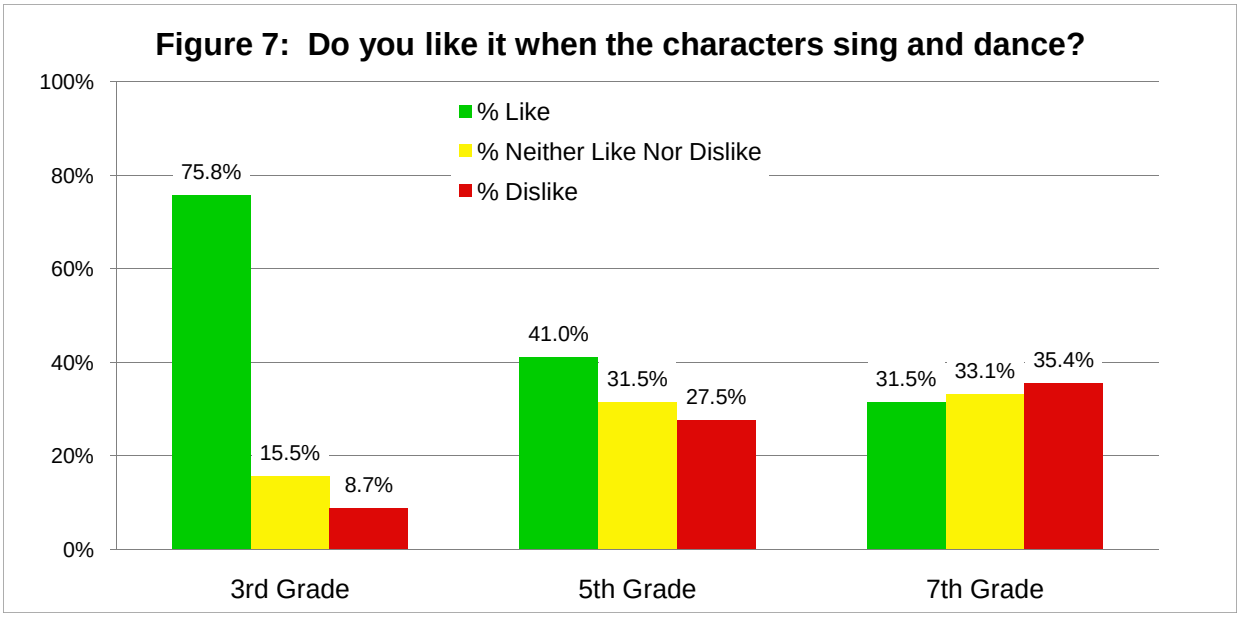
This section addresses research question three:

RQ3: How did teachers and students react to the SuccessMaker Math program?

The first sub-section summarizes the student math academic attitude survey results. The second and third sub-sections summarize the end-of-year student SuccessMaker opinion surveys and comments collected from SuccessMaker teachers during focus groups interviews, respectively.

When students were surveyed, 93% of 3rd grade, 79% of 5th grade, and 88% of 7th grade students indicated they liked using the SuccessMaker program.





Student SuccessMaker Math Attitudes

SuccessMaker students were surveyed at the end of the school year as to their opinions on several aspects of the program (i.e., 3rd = 268, 5th = 200, and 7th = 127 responses). Figures 4 and 5 show students' reactions to the math program. The overwhelming majority of 3rd grade students (i.e., 70%) indicated they liked using the program, and only 21% of 5th grade and 12% of 7th grade students indicated they disliked using the program. Similarly, 3rd grade students responded most positively to the characters and animation, and found the learning activities engaging with 90% reporting they liked the characters and 76% reported liking the animation. Not as many older students found the characters and animation engaging. Of 5th graders, 18% disliked the learning activities and 28% disliked the animation. Increasingly at 7th grade, 25% of students reported disliking the learning activities and 35% indicated they disliked the characters and animation.

Teacher SuccessMaker Attitudes

Opinions about the SuccessMaker program were systematically collected from teachers during focus group sessions. Focus groups were conducted at each school during site visits between April and early June. These sessions provided a forum for teachers and administrators to answer specific questions as well as express their professional and personal opinions regarding the program. The teachers were encouraged to speak without hesitation or inhibition, and to be as candid as possible. The focus group sessions provided extensive insight into teacher and student experiences with, and attitudes about, the SuccessMaker Math program. This information was supplemented with opinions gained from students when students were observed using the program.

The focus group results describe what teachers and students liked about the SuccessMaker program, how the program could be improved, and how teachers are using specific features of the system.

The sessions provided the research team with the following insights into teacher and student experiences with the program. Teachers and students quickly became comfortable with the SuccessMaker program, and felt the program was a good educational investment. The teacher response to the program was overwhelmingly positive, with 80% of the 646 recorded comments coded as positive in nature.

Teacher response to SuccessMaker was overwhelmingly positive, with 80% of all responses coded as positive in nature.

Teachers felt that their current print supplements or past computer-based interventions could not compete with SuccessMaker when it comes to interactivity, differentiated content, immediate feedback, and student engagement.

5th grade teacher: "I love how it differentiated for me. It gave them the test. It found out what their weaknesses were without me going in there. It did all the work for me."

7th grade teacher: SuccessMaker puts them where they need to be and builds them up. With [previously used computer program] they wouldn't go to certain areas if they didn't know them.

Teachers like the interactive nature of the educational activities that comprise the program. Teachers also like that the instruction is differentiated for the individual student. The marriage of the interactive learning objects to the differentiated content keeps students engaged and challenged in their own independent learning environment.

3rd grade teacher: "It's good because you feel like everyone got what they needed. Felt like SuccessMaker was your co teacher."

7th grade teacher: "I think it's a very essential tool for students that are at different levels."

Teachers stressed the importance of having a program that is aligned to the content of the current curriculum as well as state standards and assessments. An overwhelming majority of the teachers felt that the program was aligned with both state and district educational objectives, as well as to curriculum content. Several teachers used the program specifically to prepare for benchmark and state testing.

3rd grade teacher: "[My students] hit all skills possible in the beginning. My kids are ready to go."

5th grade teacher: "I noticed fractions. In 5th grade we spend a lot of time on that. I was excited when that came up for some of our students."

Teachers indicated that students are learning concepts from the program that are different from what has been traditionally taught or before it is even introduced in the classroom. This provides a new and exciting dimension to learning as it creates an environment of confidence and discussion for the students when a concept they have experienced on the program is identified in class.

5th grade teacher: "They see it [new material] for the first time in SuccessMaker instead of seeing it in class under pressure. It takes some of the pressure out. They are not as intimidated."

5th grade teacher: "I found my kids were already motivated, they would recognize when we got to a new concept in class, "well I've already had that on the computer". I had one girl who said, "I saw that on SuccessMaker two months ago," gave them more confidence in the classroom."

Further, teachers felt the program reinforces skills already discussed in class.

3rd grade teacher: "Some of my kids are very hard to motivate, but with SuccessMaker they will do it. So if I can link what we are doing in the classroom with what they did in SuccessMaker, they are automatically more interested."

First 3rd grade teacher: "Really reinforces. Second 3rd grade teacher: "Vocabulary too, they will say we heard that in SuccessMaker."

Teachers felt the initial placement and adaptive motion through the content worked well.

3rd grade teacher: "The IP on math, I thought was great."

5th grade teacher: "I didn't see any frustration, it seemed like they were progressing at their own pace. It was great."

The program's reporting feature was also well-received by the teachers. Though all teachers were trained on the reporting feature by the time of the site visits, many teachers were still relatively new to the reporting feature for a variety of reasons, including; starting the program later in school year, time constraints, and lack of interest.

3rd grade teacher: Then I notice wow, most of my kids have mastered that skill and we don't have to review that. It was pretty easy once we figured out what we were doing

3rd grade teacher: "I did a little bit with it. Didn't do near as much as I wanted to. Think I did three separate lessons. I liked it because I could base my lessons off of it. I like it because there are a lot of questions, but you could kind of navigate through those questions. Like little modules you could check off."

5th grade teacher: "I wish I had used more of the reports. I did not utilize that enough."

Most teachers tended to walk around the room when students were using SuccessMaker in the lab, looking over students' shoulders, monitoring their progress and answering their questions. In doing this, teachers gained a lot of insight into their students' development as well as the ability to deliver personal instruction.

3rd grade teacher: "I had one student; she would just sit there and look at me. I don't understand this. I found out she did not know how to count by fives. I didn't know that."

Individual preference and teacher expectations dictated how teachers utilized the reports and what they liked most about the reporting system. The research team did not find that teachers

often used the reports to inform classroom instruction. Teachers tended to use the reporting system to identify students for remediation and discover off-task behavior, as well as to monitor and report student progress.

3rd grade teacher: "At the last parent-teacher conference, I ran off the areas of difficult report for each parent. They liked it."

3rd grade teacher: "I have used those [reports] for leveling students, to split them into groups."

5th grade teacher: "I look at how many questions they have answered. Sometimes they have been on there for 20 minutes and answered 2 questions. I do look at that. It tells me who is on it and [who is] just sitting there."

5th grade teacher: "When I would see the students struggling the next day I could go back to their last session and see what their score was. I could say, oh this was not the score you told me yesterday. This is what you need to work on, [for example] if it was integers or something."

Teachers used program reports most often to identify students for remediation as well as to monitor student progress. Teachers also used the reports to convey student progress information to curriculum specialists and parents.

Teachers firmly believe that their students like using the program. When formally interviewed, teachers were overwhelmingly positive about their students' interactions with the program. Of the 179 recorded comments, 79% were positive in nature. Teachers felt that the program ultimately makes math more attractive to their students than it has been in the past.

3rd grade teacher: "My kids enjoyed it. There was not a day or a moment where they would say, "Oh why do we have to be here?" They look forward to going."

3rd grade teacher: "My kids were really excited to show me their scores at the end of the day. Just that competition with themselves to do better."

3rd grade teacher: "And the speed games. I hear a lot of good feedback about the speed games."

5th grade teacher: "My kids really like it; they really look forward to it."

7th grade teacher: "The 7th graders, they'd rather do math on the computer than in the classroom"

Teachers firmly believe that their students like using SuccessMaker Math and feel that the program makes the learning process more fun for students.

Although most teachers felt that the characters and animation were appropriate, a few found the characters too immature and the animation distracting. Whereas third 3rd teachers overwhelmingly found the animation and graphics a welcome component to the program, negative response to the graphics and animation were most prevalent with the 5th and 7th grade teachers.

3rd grade teacher: "The animation hooked them in."

5th grade teacher: "They think it's silly. One girl complained about the dog licking the screen. They just want to move on."

A majority of teachers felt that the program challenged both their special needs and higher achieving student populations. Teacher also felt the SuccessMaker math program was more engaging and challenging than previous printed and computer-based supplements, helpful for ELL students and struggling readers, and an overall good educational investment.

3rd grade teacher: "I saw the kids picking up a lot more English."

3rd grade teacher: "I really like it for enrichment for my high kids."

5th grade teacher: "I do think it was really beneficial for those kids that need that enrichment. The kids that just don't get it, even my low kids had great gains."

7th grade teacher: "I have an ELL and he does better on SuccessMaker than he does in the classroom."

A majority of teachers felt the initial placement and adaptive motion of students through the program was effective and the learning activities were well-differentiated and aligned to their current curricula and state educational objectives. Although most teachers made minimal use of the reporting system, the teachers overwhelmingly responded positively to the reporting system and believe it met their needs. Teachers reported rare minor technical issues (ex., logging in, activities loading), most likely a result of their district and school infrastructure. Teachers also felt the SuccessMaker math program was more engaging and challenging than previous printed and computer-based supplements, helpful for ELL students and struggling readers, and an overall good educational investment.

IV. DISCUSSION

Teachers and students quickly became comfortable with the SuccessMaker program, and felt the program was a good educational investment. When interviewed, the teacher response to the program was overwhelmingly positive. Teachers appreciated the reporting system, felt the initial placement and adaptive motion of students through the program were effective, the learning activities were well-differentiated and aligned to their current curricula and state educational objectives, the program challenged both their lower and higher achieving student populations, and that the audio and graphics allowed ELL and lower reading achieving populations to learn.

Teachers firmly believe that their students like using SuccessMaker Math and feel that the program makes the learning process more fun for students. Students themselves reported positive attitudes towards the program as well as more positive academic attitudes than non-users.

Teachers also firmly believe that their students like using the program and feel that the program makes the learning process more fun. Students appreciate the capacity of the program to allow them to laugh and interact with their own virtual learning environment. When surveyed, only a small minority of students indicated they disliked the program. Further evidence that the program resonated positively with students can be seen in the math attitude survey results where SuccessMaker students had higher scores than did their comparison group counterparts. The 3rd and 7th grade differences were both statistically significant, very large (i.e., 3rd 0.99 standard deviations, 7th 0.62 standard deviations) and also seen for several at risk populations.

Teachers came up with creative solutions to get all students on the program each week, overcoming packed classroom lesson plans and filled computer lab schedules. Most teachers went to the lab 2 or 3 times a week for an average of 24 minutes. Ten teachers went to the lab more than three times a week. Only four teachers had their students use the program in the classroom for 30% or more of their total usage. Total program usage was a median of 19, 18, and 17 hours, for 3rd, 5th, and 7th grade respectively.

The final study sample was diverse and very large at 1,186 students. Three districts have a highly transient population and thus had comparatively high attrition. Though diverse, the sample was specifically heavily Hispanic, low SES, and overall low achieving, including the type of at-risk students that would benefit from a well-conceived and implemented mathematics intervention.

The data indicates clearly that diverse populations of students receiving SuccessMaker Math can be successful in significantly increasing achievement.

The achievement data indicates clearly that diverse populations of students receiving SuccessMaker Math can be successful when receiving as little as ten hours on the program. After holding confounding factors constant for both groups (i.e., baseline scores, student demographic information, and classroom environment indicators) and estimating end-of-year raw score group mean differences SuccessMaker students in all three grades statistically

significantly outperformed their comparison group counterparts on the GRADE Total score and *Process and Applications* subtest. Likewise, SuccessMaker students in 3rd and 5th grade statistically significantly outperformed their comparison group counterparts on the *Operations and Computation* while 7th grade students performed similarly to their comparison peers on this subtest. SuccessMaker students in 3rd and 7th grade performed similarly to their comparison peers on the Concepts and Communications subtest. The 5th grade comparison group performed statistically significantly greater than 5th grade SuccessMaker students on this subtest.

In summary, the SuccessMaker Math program was found to significantly positively impact student achievement scores in various domains of math achievement. Large comparative effects were also seen for at-risk populations. Furthermore, student attitudes were positively impacted by the SuccessMaker Math program.

A.1 Comparative Study Group Results by Program Usage

Appendix 1 lists the comparative study group results (i.e., comparison group vs. SuccessMaker group) broken down by program usage time. Comparisons on assessment outcomes (i.e., GMADE, GMADE subtests, mathematics attitude survey) were made between study groups using model adjusted end-of-year raw score group mean differences. Adjusted group mean differences are calculated holding the effects of confounding factors constant for both groups, that is, baseline scores, student demographic information, and classroom environment indicators are set to the sample mean. Usage time is rounded down to the nearest hour. Effect sizes reported here are calculated using standard deviation for the comparison group (i.e., Effect Size = estimated adjusted group difference / comparison sample standard deviation). Also reported is the number of students at each grade and usage level as well as the average usage time for these students.

3rd Grade Usage[1]	Ave. Hours[2]	GMADE Effect Size[3]
less than 10 hours	9 (3)	***
10 to 19 hours	17 (147)	1.23
20 to 29 hours	25 (106)	1.18
30 or more hours	32 (26)	1.21

1. usage time rounded down to nearest hour
2. Ave. Hours = average of students' usage in hours, parentheses indicate sample size
3. Effect Size = estimated adjusted group difference / comparison sample standard deviation

5th Grade Usage[1]	Ave. Hours[2]	GMADE Effect Size[3]
less than 10 hours	7 (11)	0.74
10 to 19 hours	15 (94)	0.70
20 to 29 hours	23 (54)	0.64
30 or more hours	35 (65)	0.55

1. usage time rounded down to nearest hour
2. Ave. Hours = average of students' usage in hours, parentheses indicate sample size
3. Effect Size = estimated adjusted group difference / comparison sample standard deviation

7th Grade Usage[1]	Ave. Hours[2]	GMADE Effect Size[3]
less than 10 hours	9 (4)	0.70
10 to 19 hours	16 (72)	0.75
20 to 29 hours	24 (51)	0.93(2)
30 or more hours	31 (9)	1.14(1,2,3)

1. usage time rounded down to nearest hour
2. Ave. Hours = average of students' usage in hours, parentheses indicate sample size
3. Effect Size = estimated adjusted group difference / comparison sample standard deviation

3rd Grade Usage[1]	Ave. Hours[2]	Concepts and Communication Effect Size[3]
less than 10 hours	9 (3)	***
10 to 19 hours	17 (147)	***
20 to 29 hours	25 (106)	***
30 or more hours	32 (26)	***

1. usage time rounded down to nearest hour
2. Ave. Hours = average of students' usage in hours, parentheses indicate sample size
3. Effect Size = estimated adjusted group difference / comparison sample standard deviation

5th Grade Usage[1]	Ave. Hours[2]	Concepts and Communication Effect Size[3]
less than 10 hours	7 (11)	***
10 to 19 hours	15 (94)	***
20 to 29 hours	23 (54)	-0.57
30 or more hours	35 (65)	***

1. usage time rounded down to nearest hour
2. Ave. Hours = average of students' usage in hours, parentheses indicate sample size
3. Effect Size = estimated adjusted group difference / comparison sample standard deviation

7th Grade Usage[1]	Ave. Hours[2]	Concepts and Communication Effect Size[3]
less than 10 hours	9 (4)	***
10 to 19 hours	16 (72)	***
20 to 29 hours	24 (51)	***
30 or more hours	31 (9)	***

1. usage time rounded down to nearest hour
2. Ave. Hours = average of students' usage in hours, parentheses indicate sample size
3. Effect Size = estimated adjusted group difference / comparison sample standard deviation

3rd Grade Usage[1]	Ave. Hours[2]	Operations and Computation Effect Size[3]
less than 10 hours	9 (3)	***
10 to 19 hours	17 (147)	0.93(1)
20 to 29 hours	25 (106)	0.80(1)
30 or more hours	32 (26)	1.02(1)

1. usage time rounded down to nearest hour
2. Ave. Hours = average of students' usage in hours, parentheses indicate sample size
3. Effect Size = estimated adjusted group difference / comparison sample standard deviation

5th Grade Usage[1]	Ave. Hours[2]	Operations and Computation Effect Size[3]
less than 10 hours	7 (11)	0.81
10 to 19 hours	15 (94)	0.82
20 to 29 hours	23 (54)	1.09(4)
30 or more hours	35 (65)	0.79

1. usage time rounded down to nearest hour
2. Ave. Hours = average of students' usage in hours, parentheses indicate sample size
3. Effect Size = estimated adjusted group difference / comparison sample standard deviation

7th Grade Usage[1]	Ave. Hours[2]	Operations and Computation Effect Size[3]
less than 10 hours	9 (4)	***
10 to 19 hours	16 (72)	***
20 to 29 hours	24 (51)	***
30 or more hours	31 (9)	***

1. usage time rounded down to nearest hour
2. Ave. Hours = average of students' usage in hours, parentheses indicate sample size
3. Effect Size = estimated adjusted group difference / comparison sample standard deviation

3rd Grade Usage[1]	Ave. Hours[2]	Process and Applications Effect Size[3]
less than 10 hours	9 (3)	1.28
10 to 19 hours	17 (147)	1.46
20 to 29 hours	25 (106)	1.46
30 or more hours	32 (26)	1.50

1. usage time rounded down to nearest hour
2. Ave. Hours = average of students' usage in hours, parentheses indicate sample size
3. Effect Size = estimated adjusted group difference / comparison sample standard deviation

5th Grade Usage[1]	Ave. Hours[2]	Process and Applications Effect Size[3]
less than 10 hours	7 (11)	0.95(3,4)
10 to 19 hours	15 (94)	0.77(3,4)
20 to 29 hours	23 (54)	0.47
30 or more hours	35 (65)	0.42

1. usage time rounded down to nearest hour
2. Ave. Hours = average of students' usage in hours, parentheses indicate sample size
3. Effect Size = estimated adjusted group difference / comparison sample standard deviation

7th Grade Usage[1]	Ave. Hours[2]	Process and Applications Effect Size[3]
less than 10 hours	9 (4)	1.65(2)
10 to 19 hours	16 (72)	1.16
20 to 29 hours	24 (51)	1.45(2)
30 or more hours	31 (9)	1.81(2,3)

1. usage time rounded down to nearest hour
2. Ave. Hours = average of students' usage in hours, parentheses indicate sample size
3. Effect Size = estimated adjusted group difference / comparison sample standard deviation

3rd Grade Usage[1]	Ave. Hours[2]	Student Math Attitude Effect Size[3]
less than 10 hours	9 (3)	***
10 to 19 hours	17 (143)	1.14
20 to 29 hours	25 (104)	1.27
30 or more hours	32 (25)	***

1. usage time rounded down to nearest hour
2. Ave. Hours = average of students' usage in hours, parentheses indicate sample size
3. Effect Size = estimated adjusted group difference / comparison sample standard deviation

5th Grade Usage[1]	Ave. Hours[2]	Student Math Attitude Effect Size[3]
less than 10 hours	7 (10)	***
10 to 19 hours	15 (93)	***
20 to 29 hours	24 (51)	***
30 or more hours	35 (64)	***

1. usage time rounded down to nearest hour
2. Ave. Hours = average of students' usage in hours, parentheses indicate sample size
3. Effect Size = estimated adjusted group difference / comparison sample standard deviation

7th Grade Usage[1]	Ave. Hours[2]	Student Math Attitude Effect Size[3]
less than 10 hours	9 (3)	0.74
10 to 19 hours	15 (48)	***
20 to 29 hours	24 (46)	1.07
30 or more hours	31 (9)	***

1. usage time rounded down to nearest hour
2. Ave. Hours = average of students' usage in hours, parentheses indicate sample size
3. Effect Size = estimated adjusted group difference / comparison sample standard deviation

Instructional Support Alignment

	ROLES and Responsibilities	**Monitoring Structure**
Principal Lead	Directly work with Principal to monitor operational, managerial, and academic progress	• Weekly with Executive Directors of School Development • Monthly with Executive staff • Bi Monthly school visits with principal • Quarterly data review with Executive Team
Professional Development and Leadership Specialist	20% of time teaching most at risk students 80% of time coaching, training, and mentoring faculty and teacher leaders	• Weekly with Directors of Curriculum and Staff Development • Monthly with Principal Leads • Bi monthly school visits with teacher leaders, coaches, and administration • Quarterly data review with Executive Team
Teacher Leader	40% of time teaching, 60% of time coaching, modeling, and mentoring faculty	• Bi monthly with Professional Development and Leadership Specialists • Daily meets with assigned faculty

The School District of Lee County
Job Description

JOB TITLE: **Principal Lead**

FLSA STATUS:	Exempt	**PAY GRADE:**	Supplemental
SALARY SCHEDULE:	Administrator	**JOB CODE:**	NA
BARGAINING UNIT:	Non-bargaining	**DAYS PER YEAR:**	Supplemental
WORKER'S COMP CATEGORY:	9101 - All Other		

MAJOR FUNCTION:

To lead principals toward fulfillment of their potential in support of the Superintendent's priorities to include coaching on management, operations, instruction, and student achievement.

MINIMUM QUALIFICATIONS:

- Master's degree or higher.
- Valid Florida School Principal certificate or the equivalent covering Educational Leadership or Administration and Supervision.
- A minimum of four (4) years of Principal experience; must have been a principal in the School District of Lee County for a minimum of one (1) year.
- A minimum of four (4) years of Effective or Highly Effective administrative evaluation ratings.

Such alternatives to the above qualifications as the Board may find acceptable.

KNOWLEDGE, SKILLS, AND ABILITIES:

- Experience in developing and providing professional development to adult learners.
- Established skill in continuous improvement processes such as Quality, Best Practices, and Interest-Based Problem Solving.
- Demonstrated evidence of strong organizational, leadership, and managerial skills.
- Ability to complete Instructional Coaching Training course (if applicable).
- Established skill in oral and written communication.
- Demonstrated ability to lead diverse groups of people.
- Experience with industry-standard computer applications.

REPORTS TO: Designated Administrator

ESSENTIAL JOB FUNCTIONS:

- Work collaboratively with principals, assistant principals, faculty, and staff in assigned schools to build a capacity to increase student achievement, managerial, and coaching skills with a primary focus on new principals.
- Interpret data to guide teaching, learning and managerial decision making.
- Develop, coordinate, and provide job embedded professional development opportunities for principals, assistant principals, and faculty.

- Support colleagues by providing information, mentoring, modeling, and problem solving strategies that align with the School Improvement Plan.
- Design, select, modify, and evaluate research based instructional strategies that reflect core standards, curriculum goals, and the interests, motivation, and needs of adult/student learners.
- Develop, manage, coordinate, and deliver leadership development and technology development identified as necessary for improving leadership and technology skills for the district.
- Evaluate diverse learning activities related to improving leadership and technology in terms of their impact on quality of implementation and quality of outcomes for students.
- Provide the executive management with regular updates on professional development and principal requests as needed.
- Assist in identifying and developing future administrators.
- Create opportunities for school leaders across school sites to collaborate and learn from one another.
- Provide a schedule of training and development opportunities.
- Provide broadminded, specific, and constructive feedback and advice.
- Plan and direct a system of feedback and assessment of the effectiveness of training and development programs.
- Attend and deliver Principal Lead trainings as required.
- Respond to internal and external customers in a timely, accurate, courteous, and empathetic manner representing the School District of Lee County in a positive light.
- Participate in school advisory, business, and community groups and activities.
- Serve with other educational leaders on work groups, committees, and project action teams that directly support schools.
- Responsible for self-development and keeping up to date on the current trends and best practices regarding educational leadership and technology training at district, state, and national levels.

OTHER JOB FUNCTIONS:

- Generate creative solutions to District challenges.
- Lead and monitor division/departmental in progress toward attainment of strategic goals and objectives.
- Interpret and apply local, state, and/or federal legislation, requirements, and standards to district programs and services.
- Responsible for self-development and keeping up to date on the current trends and best practices regarding educational leadership and technology training at district, state, and national levels.
- Serve with other educational leaders on work groups, committees, and project action teams that directly support schools.

The School District of Lee County
Job Description

EXERTION TYPE:

- Light work. Position requires exerting up to 20 pounds of force occasionally, and/or up to 10 pounds of force frequently, and/or a negligible amount of force constantly to move objects.

OTHER PHYSICAL REQUIREMENTS:
The following selected physical activities are required to perform the essential functions of this position. *Reasonable accommodations may be made to enable individuals with disabilities to perform the essential functions.*

The physical requirements of this position. (Please check all boxes that apply)		
Physical Requirement	**Description**	**Percent of Time**
☐ **Balancing**	Maintaining body equilibrium to prevent falling and walking, standing or crouching on narrow, slippery, or erratically moving surfaces. This factor is important if the amount of balancing exceeds that needed for ordinary locomotion and maintenance of body equilibrium.	0%
☐ **Climbing**	Ascending or descending ladders, stairs, scaffolding, ramps, poles and the like, using feet and legs and/or hands and arms. Body agility is emphasized. This factor is important if the amount and kind of climbing required exceeds that required for ordinary locomotion.	0%
☐ **Crawling**	Moving about on hands and knees or hands and feet.	0%
☒ **Crouching**	Bending the body downward and forward by bending leg and spine.	10%
☒ **Feeling**	Perceiving attributes of objects, such as size, shape, temperature or texture by touching with skin, particularly that of fingertips.	30%
☒ **Finger Dexterity**	Picking, pinching, typing or otherwise working, primarily with fingers rather than with the whole hand as in handling.	70%
☒ **Grasping**	Applying pressure to an object with the fingers and palm.	30%
☒ **Hearing**	Perceiving the nature of sounds at normal speaking levels with or without correction. Ability to receive detailed information through oral communication, and to make the discriminations in sound.	100%
☒ **Kneeling**	Bending legs at knee to come to a rest on knee or knees.	10%
☒ **Lifting**	Raising objects from a lower to a higher position or moving objects horizontally from position-to-position. This factor is important if it occurs to a considerable degree and requires substantial use of upper extremities and back muscles.	10%
☐ **Pulling**	Using upper extremities to exert force in order to draw, haul, or tug objects in a sustained motion.	0%
☐ **Pushing**	Using upper extremities to press against something with steady force in order to thrust forward, downward, or outward.	0%

The School District of Lee County
Job Description

Physical Requirement	Description	Percent of Time
☒ **Reaching**	Extending hand(s) and arm(s) in any direction.	30%
☐ **Repetitive Motion**	Substantial movements (motions) of the wrists, hands, and/or fingers.	0%
☒ **Seeing**	The ability to perceive the nature of objects by the eye.	100%
☒ **Sitting**	Particularly for sustained periods of time.	70%
☒ **Standing**	Particularly for sustained periods of time.	10%
☒ **Stooping**	Bending body downward and forward by bending spine at the waist. This factor is important if it occurs to a considerable degree and requires full motion of the lower extremities and back muscles.	10%
☒ **Talking**	Expressing or exchanging ideas by means of the spoken word. Those activities in which they must convey detailed or important spoken instructions to other workers accurately, loudly, or quickly.	90%
☒ **Walking**	Moving about on foot to accomplish tasks, particularly for long distances or moving from one work site to another.	20%

TERMS OF EMPLOYMENT:

Twelve month year. Salary as established by the Board.

JDE NUMBER: S – 34.01

BOARD ADOPTION: 6/28/2016

REVISIONS:

REVIEWED:

Every job duty in a job description need not always be specifically described, and any omission does not preclude the required performance of all duties that are job related.

TITLE: **Professional Development and Leadership Specialist**

QUALIFICATIONS:

1. Bachelor's degree or higher. Master's degree preferred.
2. Valid Florida Professional teaching certificate.
3. Minimum 4 years of teaching experience; must have been teaching in the School District of Lee County for a minimum of 1 year.
4. Minimum 2 years of experience in a leadership/mentoring role.
5. Minimum 2 years of Effective or Highly Effective evaluation ratings.
6. Experience in developing and providing professional development to adult learners.
7. Demonstrated proficiency in oral and written communication.
8. Demonstrated ability to work with diverse groups.
9. Experience with industry-standard computer applications.
10. Successful completion of School District of Lee County Clinical Educator training.
11. Such alternatives to the above qualifications as the Superintendent and Board may find appropriate and acceptable.

REPORTS TO: Appropriate Administrator

JOB GOAL: To lead teachers toward the fulfillment of their potential in support of student's intellectual, emotional, physical and social growth in a safe and cost effective manner that supports the goals of the District.

ESSENTIAL FUNCTIONS:

1. Instructs students approximately 20% of the time; and fulfills Professional Development and Leadership Specialist duties approximately 80% of the time.
2. Works collaboratively with the principal, assistant principals, faculty, and staff in assigned schools to build a capacity to increase student achievement.
3. Develops, coordinates and provides professional development opportunities for faculty and administrators, including teacher leaders.
4. Supports colleagues by providing information, mentoring, modeling, and problem solving strategies that align with the School Improvement Plan.
5. Designs, selects, modifies, and evaluates research based instructional strategies that reflect core standards, curriculum goals, and the interests, motivation, and needs of adult/student learners.
6. Provides the building principal with regular updates on professional development and teacher needs.
7. Assists in identifying and developing future Teacher Leaders in the building.
8. Possesses strong oral and written communication skills.
9. Provides a schedule of activities including lesson/coaching plans to be shared with administration.

10. Attends and delivers Teacher Leader trainings as required.
11. Performs tasks or services consistent with the job goal of this position.

OTHER RESPONSIBILITIES:

1. Performs related work as required. (Note: The omission of specific statements of duties does not exclude them from the position if the work is similar, related, or a logical assignment to the position.)

PHYSICAL REQUIREMENTS:

Position requires exerting up to 20 pounds of force occasionally and/or up to 10 pounds of force frequently as needed to move objects.

TERMS OF EMPLOYMENT:

Work year and salary as established by the Board and the TALC bargaining unit through the collective bargaining process.

ASSESSMENT:

Performance of this job will be assessed annually in accordance with provisions of the Board's policy on assessment of certificated personnel.

Adopted: 05-07-13

TITLE: **Teacher Leader**

QUALIFICATIONS:

1. Bachelor's degree or higher.
2. Valid Florida Professional teaching certificate.
3. Minimum 4 years of teaching experience; preference given to current building teachers and must have been teaching in the School District of Lee County for a minimum of 1 year.
4. Minimum 2 years of experience as a leader.
5. Minimum 2 years of Effective or Highly Effective evaluation ratings.
6. Demonstrated proficiency in oral and written communication.
7. Demonstrated ability to work with diverse groups.
8. Experience with industry-standard computer applications.
9. Successful completion of School District of Lee County Clinical Educator training.
10. Such alternatives to the above qualifications as the Superintendent and Board may find appropriate and acceptable.

REPORTS TO: Principal and or Designated Administrator

JOB GOAL: To lead teachers toward the fulfillment of their potential in support of student's intellectual, emotional, physical and social growth in a safe and cost effective manner that supports the goals of the District.

ESSENTIAL FUNCTIONS:

1. Instructs students approximately 40% of the day and teaches a part-time schedule to include a planning period; fulfills teacher leader duties approximately 60% of his/her schedule.
2. Works collaboratively with the principal, assistant principals, faculty, and staff to build a capacity to increase student achievement.
3. Provides professional development opportunities for all faculty and staff members.
4. Supports colleagues by providing information, mentoring, modeling, and problem solving strategies that align with the School Improvement Plan.
5. Designs, selects, modifies, and evaluates instructional strategies that reflect curriculum goals, current knowledge, and the interests, motivation, and needs of individual learners.
6. Provides the building principal with regular updates on professional development and teacher needs.
7. Assists in identifying and developing future Teacher Leaders in the building.
8. Possesses strong oral and written communication skills.
9. Provides a schedule of weekly activities including lesson plans to be shared with administration, faculty, and staff.
10. Attends Teacher Leader trainings as scheduled.
11. Performs tasks or services consistent with the job goal of this position.

OTHER RESPONSIBILITIES:

1. Performs related work as required. (Note: The omission of specific statements of duties does not exclude them from the position if the work is similar, related, or a logical assignment to the position.)

PHYSICAL REQUIREMENTS:

Position requires exerting up to 20 pounds of force occasionally and/or up to 10 pounds of force frequently as needed to move objects.

TERMS OF EMPLOYMENT:

Work year and salary as established by the Board and the TALC bargaining unit through the collective bargaining process.

ASSESSMENT:

Performance of this job will be assessed annually in accordance with provisions of the Board's policy on assessment of certificated personnel.

Adopted: 11-20-12