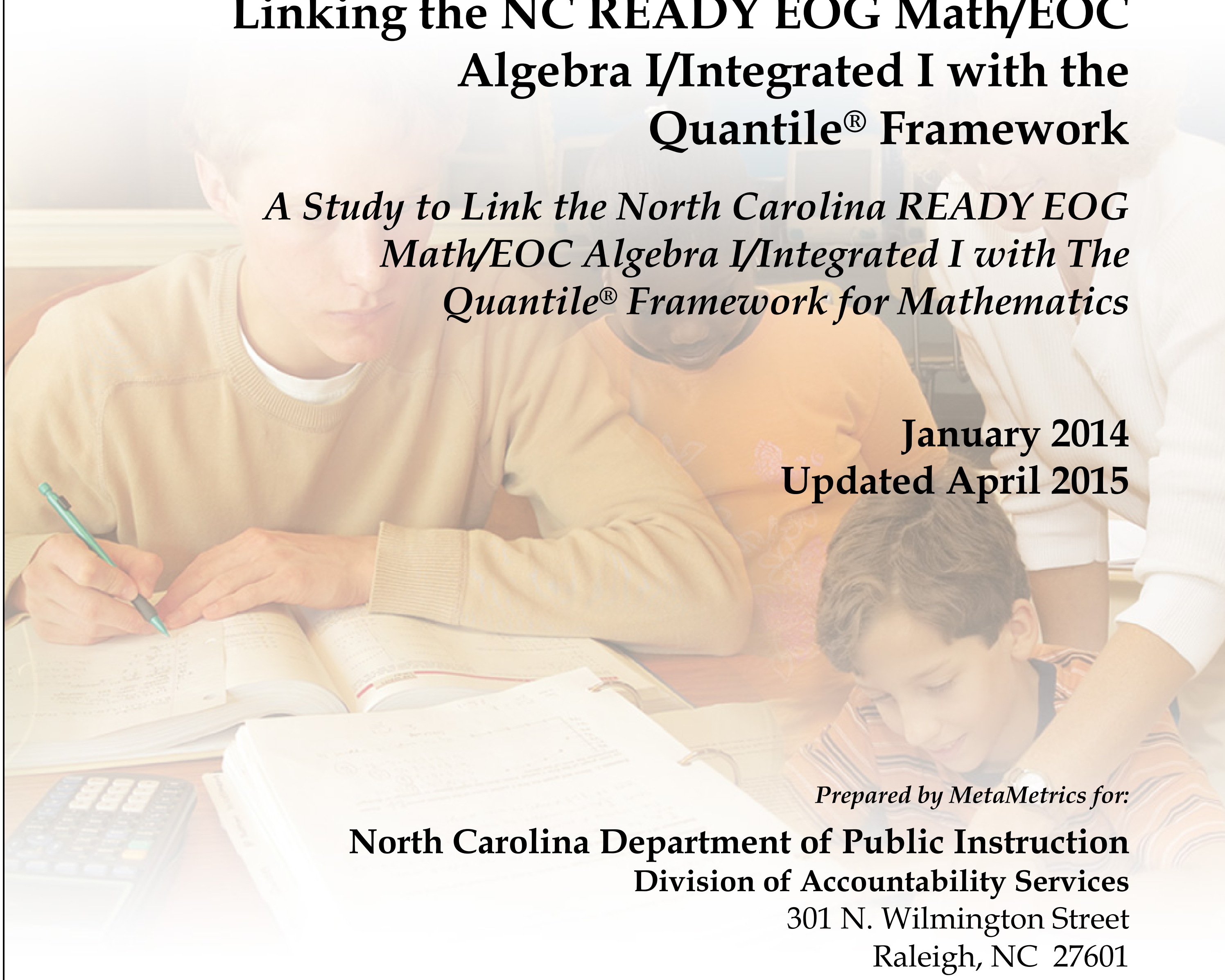
Linking the NC READY EOG Math/EOC Algebra I/Integrated I with the Quantile® Framework

A Study to Link the North Carolina READY EOG Math/EOC Algebra I/Integrated I with The Quantile® Framework for Mathematics



January 2014
Updated April 2015

Prepared by MetaMetrics for:

North Carolina Department of Public Instruction
Division of Accountability Services
301 N. Wilmington Street
Raleigh, NC 27601

MetaMetrics.

1000 Park Forty Plaza Drive, Suite 120
Durham, North Carolina 27713
www.MetaMetricsInc.com
www.Quantiles.com

Preface
Quantile Framework/Scale Enhancements

The Quantile® Framework for Mathematics is a scientific approach to measuring mathematics achievement and concept/application solvability. The Quantile Framework consists of a Quantile measure and the Quantile scale. A Quantile measure represents the difficulty of a mathematical skill, concept, or application. A Quantile measure also describes a student's understanding of the Quantile Skills and Concepts (QSCs) in the areas of geometry, measurement, numbers and operations, algebra, and data analysis and probability.

Quantile measures are expressed as numeric measures followed by a "Q" (e.g., 850Q), and are placed on the Quantile scale. (There is no space between the measure and the "Q.") The Quantile Framework spans the developmental continuum from prekindergarten mathematics through the content typically taught in Algebra II, Geometry, Trigonometry and Pre-Calculus -- from below 0Q (Emerging Mathematician) to above 1600Q. Quantile measures of one thousand or greater are reported without a comma (e.g., 1050Q). All Quantile measures are rounded to the nearest 5Q. If the Quantile measure is xxx2.5 or higher or xxx7.5 or higher, it is rounded up to the next highest 5Q; below those points should be rounded down. For example, if a computed Quantile measure is 772.51, it should be reported as 775Q. If the computed Quantile measure is 777.42, it should be reported as 775Q.

Prior to May 1, 2014, all Quantile measures at or below 0Q were reported as EM (Emerging Mathematician). Starting in spring 2014, Quantile measures below 0Q can be reported with a more specific measure. These EM measures are shown as "EMxxxQ." For example, a Quantile measure of -150 is reported as EM150Q where "EM" stands for "Emerging Mathematician" and replaces the negative sign in the number. The Quantile scale is like a thermometer, with numbers below zero indicating decreasing mathematical demand or achievement as the number moves away from zero. The smaller the number following the EM code, the more advanced the student is or the more demanding the skill or concept. For example, an EM150Q student is more advanced than an EM200Q student. Above 0Q, measures indicate increasing mathematical achievement as the numbers increase. For example, a 200Q QSC is more demanding than a 150Q QSC.

Quantile measures that are reported for an individual student should reflect the purpose for which they will be used. If the purpose is research (e.g., to measure growth at the student, grade, school, district, or state level), then actual measures should be used at all score points, rounded to the nearest integer. If the purpose is instructional, then the Quantile measures should be capped at the upper bound of measurement error (e.g., at the 95th percentile of the national Quantile norms) to ensure developmental appropriateness of the material. MetaMetrics expresses these measures used for instructional purposes as "Reported Quantile Measures" and recommends that they be used on individual score reports. In an instructional environment, all scores below 0Q should be reported as "EMxxxQ" (Emerging Mathematician); no student should receive a negative Quantile measure. As with any test score, uncertainty is present in the form of measurement error. The lowest reported value below 0Q is EM400Q.

Table i. Maximum reported Quantile measures by grade.

Grade	Quantile Cap
K	600Q
1	675Q
2	725Q
3	975Q
4	1075Q
5	1125Q
6	1200Q
7	1325Q
8	1450Q
9	1475Q
10	1500Q
11	1575Q
12	1600Q

Some assessments report a Quantile range for each student, which is 50Q above and 50Q below the student's actual Quantile measure. This range represents the limits within which instruction should be focused to ensure that the student understands the prerequisite skills and concepts associated with a specific QSC. Once a student's Quantile measure and grade are known, mathematical concepts, topics, materials, and resources can be identified within the same Quantile range.

The Quantile Framework has been aligned more closely with the Common Core State Standards for Mathematics. This was done by:

1. Moving from 5 to 6 strands, and
2. Adding approximately 70 QSCs.

Text on the following pages in the Technical Report has been updated to correspond with the language of the enhanced Quantile Framework/scale.

Table of Contents

Introduction

Often it is desirable to convey more information about test performance than can be incorporated into a single primary score scale. When two score scales are linked, the linkage can be used to provide a context for understanding the results of one of the assessments. It is often hard to explain what mathematical skills and concepts a student actually understands based on the results of a mathematics test. Parents typically ask the question, "Based on my child's test results, what math problems can he or she understand and how well?" Once a linkage is established with an assessment that is reported as specific concepts and skills, then the results of the assessment can be explained and interpreted in the context of the specific concepts and skills that a student will likely understand.

Auxiliary score scales can be used to "convey additional normative information, test-content information, and information that is jointly normative and content based" (Petersen, Kolen, and Hoover, 1989, p. 222). One such auxiliary scale is The Quantile® Framework for Mathematics, which was developed to appropriately match students with materials at a level where the student has the background knowledge necessary to be ready for instruction on the new mathematical skills and concepts.

The Quantile Framework for Mathematics takes the guesswork out of mathematics instruction. It serves as a hands-on tool which demonstrates which mathematics skills a learner has likely learned and which ones require further instruction. Teachers can also use the Quantile Framework to determine a student's readiness to learn more advanced skills. Because the Quantile Framework uses a common, developmental scale to measure both student mathematical achievement and task difficulty, educators can also determine how well a student is likely to be able to solve more complex problems (if provided with targeted instruction). The Quantile Framework includes the Quantile® measure and the Quantile® scale. The Quantile Framework targets instruction, forecasts understanding, and helps improve mathematics instruction and achievement by placing the mathematics curriculum, the materials to teach mathematics, and the students themselves on the same scale.

The Quantile Framework for Mathematics can be used to:

- Monitor student mathematics progress.
- Forecast student performance on end-of-year assessments.
- Match students with appropriate materials at their level.
- Determine if a student is ready for a new mathematics skill or concept.
- Link big mathematical concepts with state curriculum objectives.
- Identify student strengths and weaknesses.

- Understand the prerequisite skills needed to learn more advanced concepts in mathematics.
- Adapt instructional methods in the classroom to ensure a greater level of understanding and application.

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. With this valuable information in the hands of educators, instruction can be more accurately tailored to the mathematical achievement of individual students. The structure of the Quantile Framework is organized around two principles—(1) mathematics and mathematical achievement are developmental in nature and (2) mathematics is a content area.

Linking assessment results with the Quantile Framework provides a mechanism for matching each student with materials on a common scale. It serves as an anchor to which resources, concepts, skills, and assessments can be connected allowing parents, teachers, and administrators to speak the same language. By using the Quantile Framework, the same metric is applied to the materials the children use, the tests they take, and the results that are reported. Parents often ask questions like the following:

- How can I help my child become better at mathematics?
- How do I challenge my child to think mathematically?

Questions like these can be challenging for parents and educators. By linking the North Carolina READY EOG Mathematics/EOC Algebra I/Integrated I Math scales with the Quantile Framework, educators and parents will be able to answer these questions and will be better able to use the results from the tests to improve instruction and to develop each student's level of mathematics understanding.

This research study was designed to determine mathematics achievement levels that can be matched with mathematical skills and concepts based on test results on the NC READY EOG Mathematics/EOC Algebra I/Integrated I assessments. The study was conducted by MetaMetrics, Inc. in collaboration with the North Carolina Department of Public Instruction (NCDPI) (Contract No. NC10025818 dated December 17, 2012). The primary purposes of this study were to:

- provide tools (Math@Home, Quantile Teacher Assistant, and Math Skills Database) and information that can be used to answer questions related to standards, student-level accountability, test score interpretation, and test validation;
- create conversion tables for determining Quantile measures from the scores on the NC READY EOG Mathematics/EOC Algebra I/Integrated I assessments; and
- produce a report that describes the linking analysis procedures.

The Quantile Framework for Mathematics

Just as for reading, there are dozens of tests of mathematics ability measuring a common construct and all reporting the results in proprietary, non-exchangeable metrics. The benefits of having a common supplemental metric to describe mathematics ability include the following:

(1) Individual growth trajectories spanning the educational experience can be developed because the Quantile scale is developmental in nature and spans this range.
(2) Various state definitions of grade-level proficiency can be compared by re-expressing scores on a common scale.
(3) Textbook publishers can build links between mathematics curricula and major mathematics tests.
(4) Test publishers can develop classroom/interim assessments that can link to the major mathematics tests and forecast how likely the student is to meet the state performance standards.
(5) The classroom teacher can link his or her day-to-day instructional needs to the year-to-year needs of a state-level accountability system.

The Quantile Framework consists of a common supplemental metric—the Quantile—that is employed to scientifically measure a student's ability to think mathematically and his or her mathematics achievement and to locate the student in a taxonomy of mathematical skills, concepts, and applications. In order to develop the Quantile Framework, several tasks were undertaken: (1) develop a structure of mathematics that spans the developmental continuum from first grade content through Algebra I, Geometry, and Algebra II content, (2) develop a bank of items that have been field tested, (3) develop the Quantile scale (multiplier and anchor point) based on the calibrations of the field-test items, and (4) validate the measurement of mathematics ability as defined by the Quantile Framework.

Structure of the Quantile Framework

In order to develop a framework of mathematical ability, first a structure needs to be established. The structure of the Quantile Framework is organized around two principles—(1) mathematics and mathematical ability are developmental in nature and (2) mathematics is a content area.

Developmental Nature of Mathematics. The developmental nature of mathematics over time describes the increase in sophistication of the problems that can be addressed and the increase in the integration of skills and content to address these problems. The

National Research Council (2001, 2002) described mathematical proficiency as "…five intertwined strands: (1) understanding mathematics; (2) computing fluently; (3) applying concepts to solve problems; (4) reasoning logically; and (5) engaging with mathematics, seeing it as sensible, useful, and doable" (p. 1). Geary and Hamson (2002) observed that much of mathematics can be understood as an interlocking triad of competencies: conceptual competence, procedural competence, and utilization competence. In short, these competencies refer, respectively, to (1) understanding the natural language of mathematics, (2) knowing how to read mathematical expressions and employ algorithms to solve decontextualized problems, and, finally, (3) knowing why the conceptual and procedural knowledge is important and how and when to apply it. The descriptions of these three competencies follow.

A. *Vocabulary of Mathematics*. This aspect concerns the recognition of a concept either verbally or pictorially. Concepts are drawn from the mathematical content (e.g., alternate interior angles, mean, tangent) and the mathematical process (e.g., compare, estimate, etc.) strands of the National Council of Teachers of Mathematics (NCTM) framework, and include contexts (e.g., sales tax, commission) and measurement concepts (e.g., time, weight). The NCTM Standards describe this as the language of mathematics.

B. *Procedures of Mathematics*. This aspect concerns being able to apply mathematical procedures in a controlled environment (decontextualized). Procedural items ask the student to perform operations and can include graphics. For example, (1) simplifying $(3x + 2)(4x - 8)$; or (2) identifying which three angles could form a triangle knowing that the sum of the angles of a triangle equals 180°. Procedures of mathematics can also be described as algorithmic, symbolic computation, and skills.

C. *Applications of Mathematics*. This aspect involves being able to apply a mathematical procedure to solve a problem (contextualized). Application items ask the student to apply operations and concepts and can include graphics. For example, (1) determining how many cars are needed to transport the class to the museum knowing that each car can hold four students; or (2) determining how much soil is needed for a garden plot that is 3 feet wide, 6 feet long, and 8 inches deep. Applications of mathematics can also be described as problem solving, reasoning, projects, and experiences.

MetaMetrics recognizes that in order to adequately address the scope and complexity of mathematics, multiple proficiencies/competencies must be utilized. Just as the "math wars" have brought to the forefront the various aspects of mathematics instruction, we must also address these same issues. On the issue of the "math wars," Richard Riley stated "We are suffering here from an 'either-or' mentality. As any good K-12 teacher will tell you, to get a student enthused about learning, you need a mix of information and styles of providing that information. You need to provide traditional basics, along with more challenging concepts, as well as the ability to problem-solve, and to apply concepts in real world settings" (Starr, 2002). The Quantile Framework is an effort to recognize and define a basis for this "mix of information and styles" in the developmental context of mathematics instruction.

Content of Mathematics. A strand is a major subdivision of mathematical content. The strands describe what students should know and be able to do. The five strands of the Quantile Framework are based on the five Content Standards in the National Council of Teachers of Mathematics framework (NCTM, 2000), which are as follows:

1. *Number and Operations*. The development of number sense. Students with number sense naturally decompose numbers, use particular numbers as referents, solve problems using the relationships among operations and knowledge about the base-ten system, estimate a reasonable result for a problem, and have a disposition to make sense of numbers, problems, and results. Includes computational fluency.

   Instructional programs should enable all students to—
   - Understand numbers, ways of representing numbers, relationships among numbers, and number systems;
   - Understand meanings of operations and how they relate to one another;
   - Compute fluently and make reasonable estimates.

2. *Geometry*. The study of geometric shapes and structures; specifying their characteristics and relationships. A means to interpret and reflect on our physical environment and serve as tools for the study of other topics.

   Instructional programs should enable all students to—
   - Analyze characteristics and properties of two- and three-dimensional geometric shapes and develop mathematical arguments about geometric relationships;
   - Specify locations and describe spatial relationships using coordinate geometry and other mathematical systems;
   - Apply transformations and use symmetry to analyze mathematical situations;
   - Use visualization, spatial reasoning, and geometric modeling to solve problems.

3. *Algebra/Patterns and Functions*. The relationships among quantities, the use of symbols, the modeling of phenomena, and the mathematical study of change. Instructional programs should enable all students to—
   - Understand patterns, relations, and functions;

- Represent and analyze mathematical situations and structures using algebraic symbols;
- Use mathematical models to represent and understand quantitative relationships;
- Analyze change in various contexts.

4. *Data Analysis and Probability*. The collection, analysis, and interpretation of data.

   Instructional programs should enable students to—
   - Formulate questions that can be addressed with data and collect, organize, and display relevant data to answer them;
   - Select and use appropriate statistical methods to analyze data;
   - Develop and evaluate inferences and predications that are based on data;
   - Understand and apply basic concepts of probability.

5. *Measurement*. The assignment of a numerical value to an attribute of an object.

   Instructional programs should enable students to—
   - Understand measurable attributes of objects and the units, systems, and processes of measurement;
   - Apply appropriate techniques, tools, and formulas to determine measurements.

The Quantile Skills and Concepts. Within the Quantile Framework, a "Quantile Skill or Concept" (QSC) describes a specific mathematical skill and is used to annotate the Quantile scale. For example, a QSC under the Numbers and Operations strand is "Model and identify the place value of each digit in a multi-digit numeral to the hundredths place;" and a QSC under the Geometry strand is "Identify and distinguish among similar, congruent, and symmetric figures; name corresponding parts." The content taxonomy of QSCs used with the Quantile Framework was developed during the spring of 2003 for grades 1 through 8, Algebra I, and Geometry. The framework was extended to Algebra II and revised during the summer and fall of 2003. The first step in developing a content taxonomy was to review the curricular frameworks from the following sources:

- National Council of Teachers of Mathematics (NCTM).

- National Assessment of Educational Progress: 2005 Pre-Publication Edition.

- North Carolina Standard Course of Study (Revised in 2003 for grades kindergarten through high school).

- California Mathematics Framework and state assessment blueprints: *Mathematics Framework for California Public Schools: Kindergarten through Grade Twelve* (2000 Revised Edition); *Mathematics Content Standards for California Public Schools: Kindergarten through Grade Twelve* (December 1997); Blueprints document for the Star Program California Standards Tests: Mathematics (California Department of Education, adopted by SBE 10/9/02), and sample items for the California Mathematics Standards Tests (California Department of Education, January 2002).

- Florida Sunshine State Standards: Sunshine State Standards Grade Level Expectations for Mathematics, grade 2 through Mathematics. The Sunshine State Standards "are the centerpiece of a reform effort in Florida to align curriculum, instruction, and assessment." They identify what students should know and be able to do for the 21st century. Publishers are required to correlate instructional materials submitted for state adoption to the standards.

- Illinois: Illinois teachers for Illinois schools developed The Illinois Learning Standards for Mathematics. Their Goals 6 through 10 emphasize the following: numbers and operations, measurement, algebra, geometry, and data analysis and statistics—*Mathematics Performance Descriptors, Grades 1-5 and Grades 6-12* (2002).

- Texas Essential Knowledge and Skills: Texas Essential Knowledge and Skills for Mathematics (TEKS) were adopted by the Texas State Board of Education and became effective on September 1, 1998. The Texas Essential Knowledge and Skills (TEKS), the state-mandated curriculum, was "specifically designed to help students to make progress … by emphasizing the knowledge and skills most critical for student learning" (TEA, 2002b, p. 4).

  The Texas Assessment of Knowledge and Skills (TAKS) was mandated by the 76th Texas Legislature in 1999 and was administered for the first time during the 2002-2003 school year (TEA, 2002a). The TAKS was developed to assess the TEKS and ask questions in more authentic ways. The TAKS test objectives, " 'umbrella statements' generated by TEA staff with input from educators," were used to develop the items (p. 2). These statements serve as headings under which the TAKS are meaningfully grouped. The TAKS measures the statewide curriculum in reading at grades 3-9; in writing at grades 4 and 7; in English Language Arts at grades 10 and 11; in mathematics at grades 3-11; in science at grades 5, 10, and 11; and in social studies at grades 8, 10, and 11. The Spanish TAKS is administered at grades 3 through 6. Satisfactory performance on the TAKS at Grade 11 is prerequisite to a high school diploma.

The review of the content frameworks resulted in the development of a list of QSCs spanning the content typically taught in kindergarten through Algebra I, Geometry and

Algebra II. Each QSC is aligned with one of the five content strands. The QSCs can be viewed and searched at www.Quantiles.com. Each QSC consists of a description of the content, a content identification number (C_ID), the grade at which it typically first appears (Grade), and the strand it is associated with (1 = Numbers and Operations, 2 = Geometry, 3 = Algebra/Patterns & Functions, 4 = Data Analysis & Probability, and 5 = Measurement).

Although states have developed their own individual curriculum standards for years, recently there has been an unprecedented focus on developing common curriculum standards for use throughout the United States of America. Guided and supported by the Council of Chief State School Officers (CCSSO) and the National Governors Association (NGA), departments of education in the states, the United States territories and the District of Columbia have collaborated to identify common standards in English/language arts, mathematics and other content areas. Educators, researchers and educational policy makers were involved extensively in the effort to identify, catalog, review and adopt standards that would lead to students being "college and career ready" by the end of high school. The Common Core State Standards (CCSS) are the culmination of this work. They were released in June 2010 by the CCSSO and the NGA Center for Best Practices. Currently, forty-five states have adopted the CCSS for Mathematics. The standards may be viewed at http://www.corestandards.org/ (NGA Center & CCSSO, 2010a, 2010b). Additional information about the development of the CCSS may be found at the CCSSO website (http://www.ccsso.org/) and the website of the NGA (http://www.nga.org/). The Quantile Framework's QSCs have been aligned with the CCSS for mathematics and, where necessary, QSCs were revised to more closely align (e.g., specifically mentioning number and word problems should be addressed by a QSC) and additional QSCs were added (e.g., margin of error, residuals of a distribution). The alignment may be viewed and searched at www.Quantiles.com.

The Quantile Framework map (Appendix A) presents a picture of the construct of mathematics ability. The map is organized by the five strands and describes the development of mathematics from basic skills to sophisticated problem solving. Exemplar QSCs and problems are used to annotate the Quantile scale and the strands. QSCs are located on the Quantile scale at the point corresponding to the mean of the ensemble of items addressing that QSC from two large, national studies (Quantile Framework field study and *PASeries* Math field study described later in this document). Items are located on the Quantile scale corresponding to their Quantile measure based on the Quantile Framework field study.

Quantile Item Bank Development

The second step in the process of developing The Quantile Framework of Mathematics was to develop and field test a bank of items that could be used in future linking studies. Item bank development for the Quantile Framework went through several stages—content specification, item writing and review, field-testing and analyses, and final evaluation.

Item Specification and Development. Based on the list of QSCs aligned to the five strands, QSCs were identified as typically being taught at a particular grade level. The curricular frameworks from Florida, North Carolina, Texas, and California were synthesized to identify the QSCs instructed and/or assessed at each grade level. If a QSC was included in any state framework it was included in the list of QSCs for which items were to be developed for use with the Quantile Framework field study.

During the summer and fall of 2003, over 1,400 items were developed to assess the QSCs associated with content in grades 1 through Algebra II. The items were written and reviewed by mathematics educators trained to develop multiple-choice items (Haladyna, 1994). The items for the pool were specified by both strand and QSC. At least three items were written for each QSC within each grade.

With the current increased focus on authentic assessment and solving problems in context using real-world applications, mathematics items now tend to require more reading. While the vocabulary specific to mathematical content is used (e.g., congruent), every attempt is made to have the non-content vocabulary below the grade level.

Item Writer Training. Item writers were experienced teachers and item-development specialists who had experience with the everyday mathematical ability of students at various levels. The use of individuals with these types of experiences helped to ensure that the items were valid measures of mathematics. Item writers were provided with training materials concerning the development of multiple-choice items and the Quantile Framework. The item writing materials also contained incorrect and ineffective items that illustrated the criteria used to evaluate items and corrections based on those criteria. The final phase of item writer training was a short practice session with three items.

Item writers were also given additional training related to "sensitivity" issues. Part of the item writing materials addressed these issues and identify areas to avoid when selecting passages and developing items. These materials were developed based on material published on universal design and fair access—equal treatment of the sexes, fair representation of minority groups, and the fair representation of disabled individuals.

Items were reviewed and edited by a group of specialists that represented various perspectives—test developers, editors, and curriculum specialists. These individuals examined each item for sensitivity issues and for the quality of the response options.

During the second stage of the item review process, items were either "approved," "approved with edits," or "deleted."

Linking- and Field-Test Design. Tests were developed for ten levels: Levels 2 through 8 were aligned with the typical content taught in grades 2 through 8, Level 9 was aligned with the typical content taught in Algebra I, Level 10 was aligned with the typical content taught in Geometry, and Level 11 was aligned with the typical content taught in Algebra II. For each level, three forms were developed with each form containing 30 items.

First, each form consisted of 22 unique items that were targeted specifically for the grade level. Across the three grade-level forms, 66 unique items were identified. These items were selected from a pool of items that covered the content of a particular grade level. For grades 2 through 8, 22 items were from Strand 1—Numbers and Operations and 11 items were from each of the other four strands (Strand 2—Geometry, Strand 3—Algebra/Patterns & Functions, Strand 4—Data Analysis & Probability, and Strand 5—Measurement). For Algebra I and Algebra II, the primary focus of the 66 items was Strand 3—Algebra/Patterns & Functions (33 items, 50%) with the remaining items evenly distributed across the other four strands; and for Geometry, the primary focus of the 66 items was Strand 2—Geometry (33 items, 50%) with the remaining items evenly distributed across the other four strands.

Next, for each grade level, 12 of the 66 grade-level items were designated "linking" items. For each grade level set, 4 items were from Strand 1—Numbers and Operations and 2 items were from each of the other four strands (Strand 2—Geometry, Strand 3—Algebra/Patterns & Functions, Strand 4—Data Analysis & Probability, and Strand 5—Measurement). For Algebra I and Algebra II, 6 items (50%) were from Strand 3—Algebra/Patterns & Functions with the remaining six items randomly selected from the other four strands. For Geometry, 6 items (50%) were from Strand 2—Geometry with the remaining six items randomly selected from the other four strands. For Grade 1, only the "linking" set of items was included in the field-test item pool.

The linking set of items for a grade level was used to link (1) the field-test forms within the grade, (2) the field-test forms from the grade below, and (3) the field-test forms from the grade above. The final field tests were comprised of 658 unique items. Two grade 10 forms only had 29 items (one on-grade level item was dropped from each of two forms due to graphics problems).

A common-item test design was employed to vertically link the test levels. In this design, multiple tests are given to non-random groups, and a set of common items is included in the test administration to allow some statistical adjustments for possible sample-selection bias. This design is most advantageous where the number of items to be tested (treatments) is large and the consideration of cost (in terms of time) forces the experiment to be smaller than is desired (Cochran and Cox, 1957). The multiple test forms were developed using a domain-sampling model where items were randomly assigned within QSC to a test form.

Quantile Framework Field Study—Sample. The Quantile Framework field study was conducted in February 2004. Thirty-seven schools from 14 districts across six states (California, Indiana, Massachusetts, North Carolina, Utah, and Wisconsin) agreed to participate in the study. Data were received from 34 of the schools (two elementary and one middle-school did not return data). A total of 9,847 students in grades 2 through 12 were tested. The number of students per school ranged from 74 to 920. The schools were diverse in terms of geographic location, size, and type of community (e.g., suburban; small town, city, or rural communities; and urban). *Table 1* provides information about the sample at each grade level and by gender.

Table 1. Field-study participation by grade and gender.

Grade Level	*N*	Percent Female (*N*)	Percent Male (*N*)
2	1,283	48.1 (562)	51.9 (606)
3	1,354	51.9 (667)	48.1 (617)
4	1,454	47.7 (644)	52.3 (705)
5	1,344	48.9 (622)	51.1 (650)
6	976	47.7 (423)	52.3 (463)
7	1,250	49.8 (618)	50.2 (622)
8	1,015	51.9 (518)	48.1 (481)
9	489	52.0 (252)	48.0 (233)
10	259	48.6 (125)	51.4 (132)
11	206	49.3 (101)	50.7 (104)
12	143	51.7 (74)	48.3 (69)
Missing	74	39.1 (9)	60.9 (14)
Total	9,847	49.6 (4,615)	50.4 (4,696)

Students given Levels 2 through 11 were provided with rulers and students given Levels 3 through 11 were provided with protractors. For students given taking Levels 5 through 8 and 10 and 11, formulas were provided on the back of the test booklet. Administration time was approximately 45 minutes at each level. Students given Level 2 could have the test read aloud and mark in the test booklet if that was typical of instruction.

Table 2. Test-form administration by level.

Test Level	*N*	Missing	Form 1	Form 2	Form 3
2	1,283	4	453	430	397
3	1,354	7	561	387	399
4	1,454	17	616	419	402
5	1,344	3	470	448	423
6	917	13	322	293	289
7	1,309	6	463	429	411
8	1,181	16	387	391	387
9	415	4	141	136	134
10	226	5	73	77	71
11	313	10	102	101	100
Missing	51	31	9	8	3
Total	9,847	116	3,596	3,119	3,016

Table 2 shows the number of students by level and form. The final sample included 9,678 students with complete data. Data were deleted if test level or test form was not indicated or the answer sheet was blank.

Field-Test Analyses. The field-test data were analyzed using both the classical measurement model and the Rasch (one-parameter logistic item response theory) model. Item statistics and descriptive information (item number, field test form and item number, QSC, and answer key) were printed for each item and attached to the item record. The item record contained the statistical, descriptive, and historical information for an item; a copy of the item itself as it was field-tested; any comments by reviewers; and the psychometric notations. Each item had a separate item record.

Field-Test Analyses – Classical Measurement. For each item, the *p*-value (percent correct) and the point-biserial correlation between the item score (correct response) and the total test score were computed. Point-biserial correlations were also computed between each of the incorrect responses and the total score. In addition, frequency distributions of the response choices (including omits) were tabulated (both actual counts and percents). Items with point-biserial correlations less than 0.10 were removed from the item bank. *Table 3* displays the summary item statistics.

Table 3. Summary item statistics from the Quantile Framework field study (February 2004).

Level	Number of Items Tested	Mean P-value (Range)	Mean Correct Response Point-Biserial Correlation (Range)	Mean Incorrect Responses Point-Biserial Correlation (Range)
2	90	0.583 (0.12 – 0.95)	0.322 (-0.15 – 0.56)	-0.209 (-0.43 – 0.12)
3	90	0.532 (0.11 – 0.93)	0.256 (-0.08 – 0.52)	-0.221 (-0.54 – 0.02)
4	90	0.552 (0.12 – 0.92)	0.242 (-0.21 – 0.50)	-0.222 (-0.48 – 0.12)
5	90	0.535 (0.12 – 0.95)	0.279 (-0.05 – 0.50)	-0.225 (-0.45 – 0.05)
6	90	0.515 (0.04 – 0.86)	0.244 (-0.08 – 0.45)	-0.218 (-0.46 – 0.09)
7	90	0.438 (0.10 – 0.77)	0.294 (-0.12 – 0.56)	-0.207 (-0.46 – 0.25)
8	90	0.433 (0.10 - 0.81)	0.257 (-0.15 – 0.50)	-0.201 (-0.45 – 0.13)
9	90	0.396 (0.10 – 0.79)	0.208 (-0.19 – 0.52)	-0.193 (-0.53 – 0.22)
10	88	0.511 (0.01 – 0.97)	0.193 (-0.26 – 0.53)	-0.205 (-0.55 – 0.18)
11	90	0.527 (0.09 – 0.98)	0.255 (-0.09 – 0.51)	-0.223 (-0.52 – 0.07)

Field-Test Analyses – Bias. Differential item functioning (DIF) examines the relationship between the score on an item and group membership while controlling for ability. The Mantel-Haenszel procedure has become "the most widely used methodology [to examine differential item functioning] and is recognized as the testing industry standard" (Roussos, Schnipke, and Pashley, 1999, p. 293). The Mantel-Haenszel procedure examines DIF by examining j 2×2 contingency tables, where j is the number of different levels of ability actually achieved by the examinees (actual total scores received on the test). The focal group is the group of interest and the reference group serves as a basis for comparison for the focal group (Dorans and Holland, 1993; Camilli and Shepherd, 1994).

The Mantel-Haenszel chi-square statistic tests the alternative hypothesis that there is a linear association between the row variable (score on the item) and the column variable (group membership). The χ^2 distribution has 1 degree of freedom and is determined as

$$Q_{MH} = (n-1)r^2 \qquad \text{(Equation 1)}$$

where *r* is the Pearson correlation between the row variable and the column variable (SAS Institute, 1985).

The Mantel-Haenszel (MH) Log Odds Ratio statistic is used to determine the direction of differential item functioning (SAS Institute Inc., 1985). This measure is obtained by combining the odds ratios, α_j, across levels with the formula for weighted averages (Camilli and Shepherd, 1994, p. 110):

$$\alpha_j = \frac{p_{Rj}/q_{Rj}}{p_{Fj}/q_{Fj}} = \frac{\Omega_{Rj}}{\Omega_{Fj}} \qquad \text{(Equation 2)}$$

For this statistic, the null hypothesis of no relationship between score and group membership, or that the odds of getting the item correct are equal for the two groups, is not rejected when the odds ratio equals 1. For odds ratios greater than 1, the interpretation is that an individual at score level *j* of the Reference Group has a greater chance of answering the item correctly than an individual at score level *j* of the Focal Group. Conversely, for odds ratios less than 1, the interpretation is that an individual at score level *j* of the Focal Group has a greater chance of answering the item correctly than an individual at score level *j* of the Reference Group. The Breslow-Day Test is used to test whether the odds ratios from the *j* levels of the score are all equal. When the null hypothesis is true, the statistic is distributed approximately as a χ^2 with j-1 degrees of freedom (Camilli and Shepherd, 1994; SAS Institute, 1985).

For the gender analyses, males (approximately 50.4% of the population) were defined as the reference group and females (approximately 49.6% of the population) were defined as the focal group. The results from the Quantile Framework field study were reviewed for inclusion on later linking studies. The following statistics were reviewed for each item: *p*-value, point-biserial correlation, and DIF estimates. Items that exhibited extreme statistics were removed from the item bank (47 out of 685).

From the studies conducted with the Quantile Framework item bank (Palm Beach County [FL] linking study, Mississippi linking study, DoDEA/TerraNova linking study, and Wyoming linking study), approximately 6.9% of the items in any one study were flagged as exhibiting DIF using the Mantel-Haenszel statistic and the *t*-statistic from Winsteps. For each linking study the following steps were used to review the items: (1) flag items exhibiting DIF, (2) review items to determine if the content of the item is something that all students should know and be able to do, and (3) make decision to retain or delete the item.

Field-Test Analyses—Rasch Item Response Theory. Classical test theory has two basic shortcomings: (1) the use of item indices whose values depend on the particular group of examinees from which they were obtained, and (2) the use of examinee ability estimates that depend on the particular choice of items selected for a test. The basic premises of item response theory (IRT) overcome these shortcomings by predicting the performance of an examinee on a test item based on a set of underlying abilities (Hambleton and Swaminathan, 1985). The relationship between an examinee's item performance and the set of traits underlying item performance can be described by a monotonically increasing function called an item characteristic curve (ICC). This function specifies that as the level of the trait increases, the probability of a correct response to an item increases.

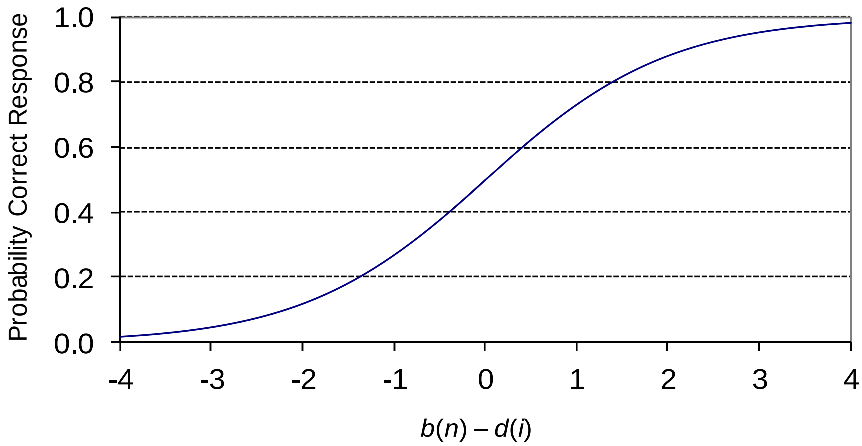
The conversion of observations into measures can be accomplished using the Rasch (1980) model, which states a requirement for the way that item calibrations and observations (count of correct items) interact in a probability model to produce measures. The Rasch IRT model expresses the probability that a person (*n*) answers a certain item (*i*) correctly by the following relationship:

$$P_{ni} = \frac{e^{b_n - d_i}}{1 + e^{b_n - d_i}} \qquad \text{(Equation 3)}$$

where d_i is the difficulty of item *i* (*i* = 1, 2, …, number of items);
b_n is the ability of person *n* (*n* = 1, 2, …, number of persons);
$b_n - d_i$ is the difference between the ability of person *n* and the difficulty of item *i*; and
P_{ni} is the probability that examinee *n* responds correctly to item *i*
(Hambleton and Swaminathan, 1985; Wright and Linacre, 1994).

This measurement model assumes that item difficulty is the only item characteristic that influences the examinee's performance such that all items are equally discriminating in their ability to identify low-achieving persons and high achieving persons (Bond and Fox, 2001; and Hambleton, Swaminathan, and Rogers, 1991). In addition, the lower asymptote is zero, which specifies that examinees of very low ability have zero probability of correctly answering the item. The Rasch model has the following assumptions: (1) unidimensionality—only one ability is assessed by the set of items; and (2) local independence—when abilities influencing test performance are held constant, an examinee's responses to any pair of items are statistically independent (conditional independence, i.e., the only reason an examinee scores similarly on several items is because of his or her ability, not because the items are correlated). The Rasch model is based on fairly restrictive assumptions, but it is appropriate for criterion-referenced assessments. *Figure 1* graphically shows the probability that a person will respond correctly to an item as a function of the difference between a person's ability and an item's difficulty.

Figure 1. The Rasch Model—the probability person *n* responds correctly to item *i*.



An assumption of the Rasch model is that the probability of a response to an item is governed by the difference between the item calibration (d_i) and the person's measure (b_n). From an examination of the graph in *Figure 1*, when the ability of the person matches the difficulty of the item ($b_n - d_i = 0$), then the person has a 50% probability of responding to the item correctly.

The number of correct responses for a person is the probability of a correct response summed over the number of items. When the measure of a person greatly exceeds the calibration (difficulties) of the items ($b_n - d_i > 0$), then the expected probabilities will be high and the sum of these probabilities will yield an expectation of a high "number correct." Conversely, when the item calibrations generally exceed the person measure ($b_n - d_i < 0$), the modeled probabilities of a correct response will be low and the expectation will be a low "number correct."

Thus, Equation 3 can be rewritten in terms of the number of correct responses of a person on a test

$$O_p = \sum_{i=1}^{L} \frac{e^{b_n - d_i}}{1 + e^{b_n - d_i}} \qquad \text{(Equation 4)}$$

where O_p is the number of correct responses of person p and L is the number of items on the test.

When the sum of the correct responses and the item calibrations (d_i) is known, an iterative procedure can be used to find the person measure (b_n) that will make the sum of the modeled probabilities most similar to the number of correct responses. One of the key features of the Rasch IRT model is its ability to place both persons and items on the same scale. It is possible to predict the odds of two individuals being successful on an item based on knowledge of the relationship between the abilities of the two individuals. If one person has an ability measure that is twice as high as that of another person (as measured by b—the ability scale), then he or she has twice the odds of successfully answering the item.

Equation 4 possesses several distinguishing characteristics:

- The key terms from the definition of measurement are placed in a precise relationship to one another.

- The individual responses of a person to each item on an instrument are absent from the equation. The only information that appears is the "count correct" (O_p), thus confirming that the raw score (i.e., number of correct responses) is "sufficient" for estimating the measure.

For any set of items the possible raw scores are known. When it is possible to know the item calibrations (either theoretically or empirically from field studies), the only parameter that must be estimated in Equation 4 is the person measure that corresponds to each observable count correct. Thus, when the calibrations (d_i) are known, a correspondence table linking observation and measure can be constructed without reference to data on other individuals.

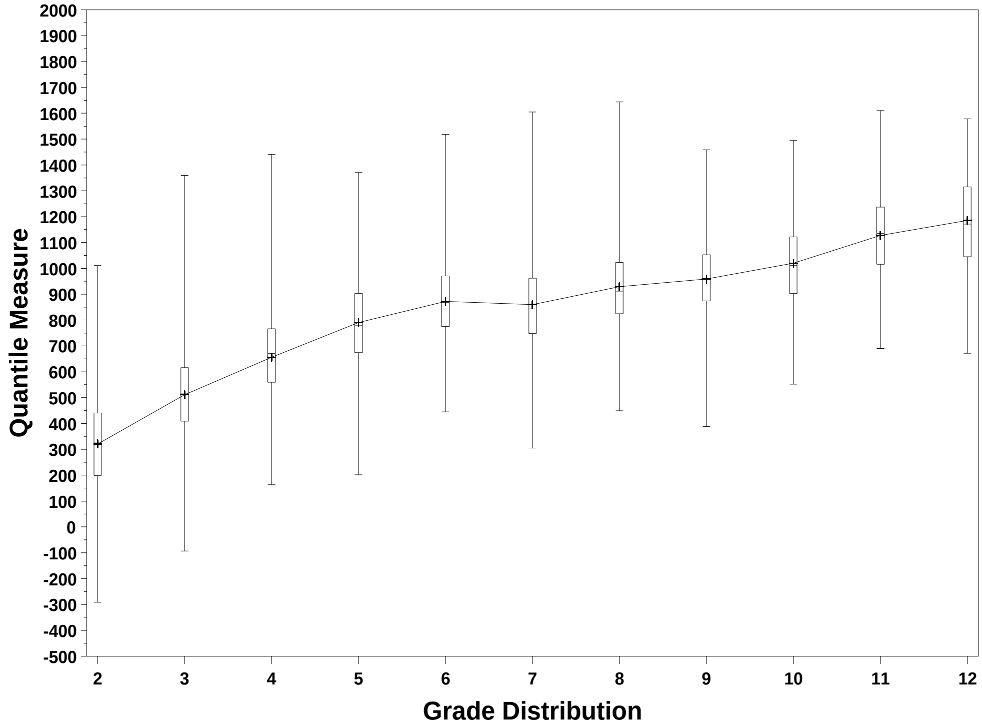
All students and items were submitted to a Winsteps analysis using a logit convergence criterion of 0.0001 and a residual convergence criterion of 0.001. Items that a student skipped were treated as missing, rather than being treated as incorrect. Only students who responded to at least 20 items were included in the analyses (22 students were omitted, 0.22%). *Table 4* shows the mean and median Quantile measures for all students with complete data at each grade level. While there is not a monotonically increasing trend in the mean and median Quantile measures in Grades 6 and 7, the measures are not significantly different. Results from other studies (e.g., *PASeries* Math described beginning on page 25 exhibit a monotonically increasing function).

Table 4. Mean and median Quantile measures for students with complete data (*N* = 9,656).

Grade Level	*N*	Mean Quantile measure (SD)	Median Quantile measure
2	1,275	320.68 (189.11)	323
3	1,339	511.41 (157.69)	516
4	1,427	655.45 (157.50)	667
5	1,337	790.06 (167.71)	771
6	959	871.82 (153.02)	865
7	1,244	860.52 (174.16)	841
8	1,004	929.01 (157.63)	910
9	482	958.69 (152.81)	953
10	251	1019.97 (162.87)	1005
11	200	1127.34 (178.57)	1131
12	138	1185.90 (189.19)	1164

Figure 2 shows the relationship between grade level and Quantile measure. The following box and whisker plots (*Figures* 2, 3, and 4) show the progression of the *y*-axis scores from grade to grade (the *x*-axis). For each grade, the box refers to the inter-quartile range. The line within the box indicates the median and the + indicates the mean. The end of each whisker shows the minimum and maximum values of the *y*-axis which is the Quantile measure. Across all students, the correlation between grade and Quantile measure was 0.76.

Figure 2. Box and whisker plot of the Rasch ability estimates of all students with complete data (N = 9,656).



All students with outfit mean square statistics greater than or equal to 1.8 were removed from further analyses. A total of 480 students (4.97%) were removed from further analyses. The number of students removed ranged from 8.47% (108) in grade 2 to 2.29% (22) in grade 6 with a mean percent decrease of 4.45% per grade.

All remaining students (9,176) and all items were submitted to a Winsteps analysis using a logit convergence criterion of 0.0001 and a residual convergence criterion of 0.001. Items that a student skipped were treated as missing, rather than being treated as incorrect. Only students who responded to at least 20 items were included in the analyses. *Table 5* shows the mean and median Quantile measures for the final set of students at each grade level. *Figure 3* shows the results from the final set of students. The correlation between grade level and Quantile measure was 0.78.

Table 5. Mean and median Quantile measures for the final set of students (*N* = 9,176).

Grade Level	*N*	Median Logit Value	Mean Quantile measure (Median)
2	1,167	-2.800	289.03 (292)
3	1,260	-1.650	502.18 (499)
4	1,352	-0.780	652.60 (656)
5	1,289	0.000	795.25 (796)
6	937	0.430	880.77 (874)
7	1,181	0.370	877.75 (863)
8	955	0.810	951.41 (942)
9	466	1.020	982.62 (980)
10	244	1.400	1044.08 (1048)
11	191	2.070	1160.49 (1169)
12	134	2.295	1219.87 (1210)

Figure 3. Box and whisker plot of the Rasch ability estimates for the final sample of students with outfit statistics less than 1.8 (*N* = 9,176).

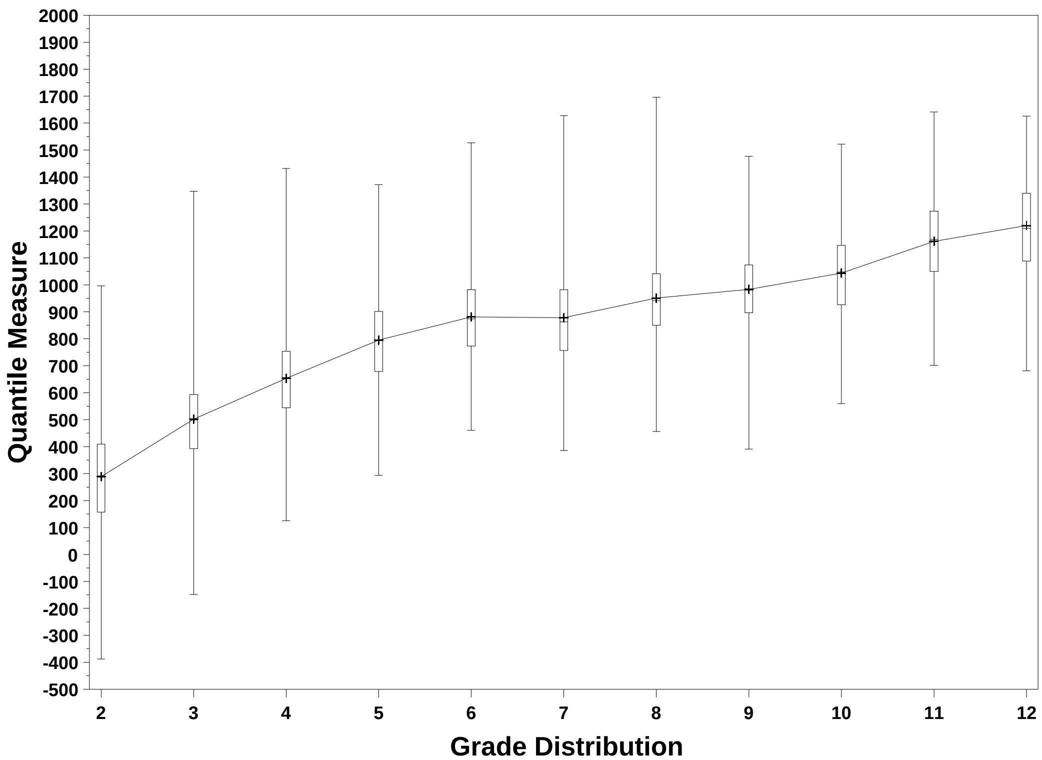
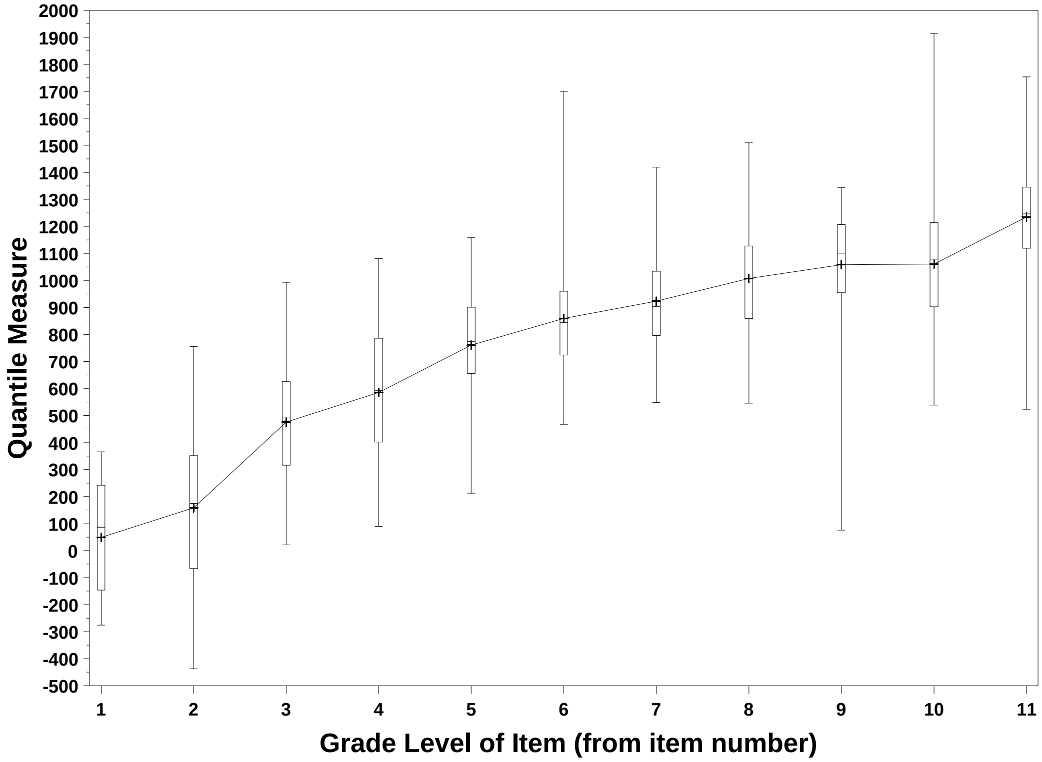


Figure 4 shows the distribution of item difficulties based on the final sample of students. For this analysis, missing data were treated as "skipped" items and not counted as wrong. There is a gradual increase in difficulty when items are sorted by level of test for which the items were written. This distribution appears to be non-linear, which is

consistent with other studies. The correlation between the grade level for which the item was written and the Quantile measure of the item was 0.80.

Figure 4. Box and whisker plot of the Rasch difficulty estimates of the 685 Quantile Framework items for the final sample of students (*N* = 9,176).



Calibration of Items on the Quantile Scale

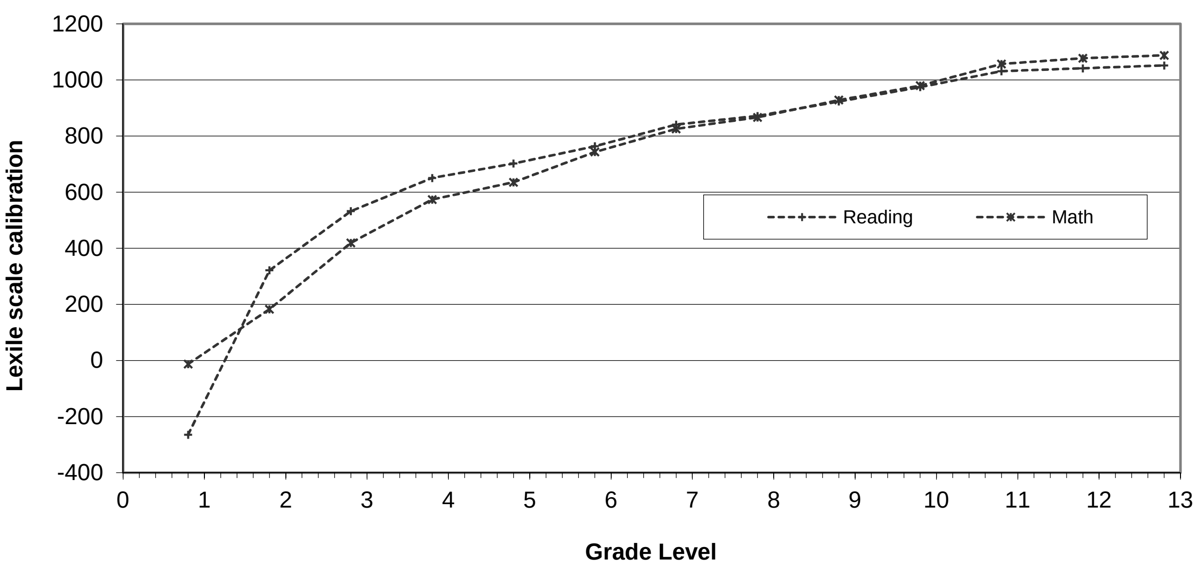
In developing the Quantile scale, two features of the scale were needed: (1) scale multiplier (conversion factor) and (2) anchor point. The Rasch item response theory model (Wright and Stone, 1979) was used to estimate the difficulties of items and the abilities of persons on the logit scale.

The calibrations of the items from the Rasch model are objective in the sense that the relative difficulties of the items will remain the same across different samples of persons (specific objectivity). When two items are administered to the same person it can be determined which item is harder and which one is easier. This ordering should hold when the same two items are administered to a second person. If two different items are administered to the second person, there is no way to know which set of items is harder and which set is easier. The problem is that the location of the scale is not known. General objectivity requires that scores obtained from different test administrations be tied to a common zero—absolute location must be sample independent (Stenner, 1990).

To achieve general objectivity, the theoretical logit difficulties must be transformed to a scale where the ambiguity regarding the location of zero is resolved.

The first step in developing the Quantile scale was to determine the conversion factor (CF) to be used to go from logits to Quantile measure. Based on prior research with reading and the Lexile scale, the decision was made to examine the relationship between reading and mathematics scales used with other assessments. The median scale score for each grade level on a norm-referenced assessment linked with the Lexile scale is plotted in *Figure 5* using the same conversion equation for both reading and mathematics.

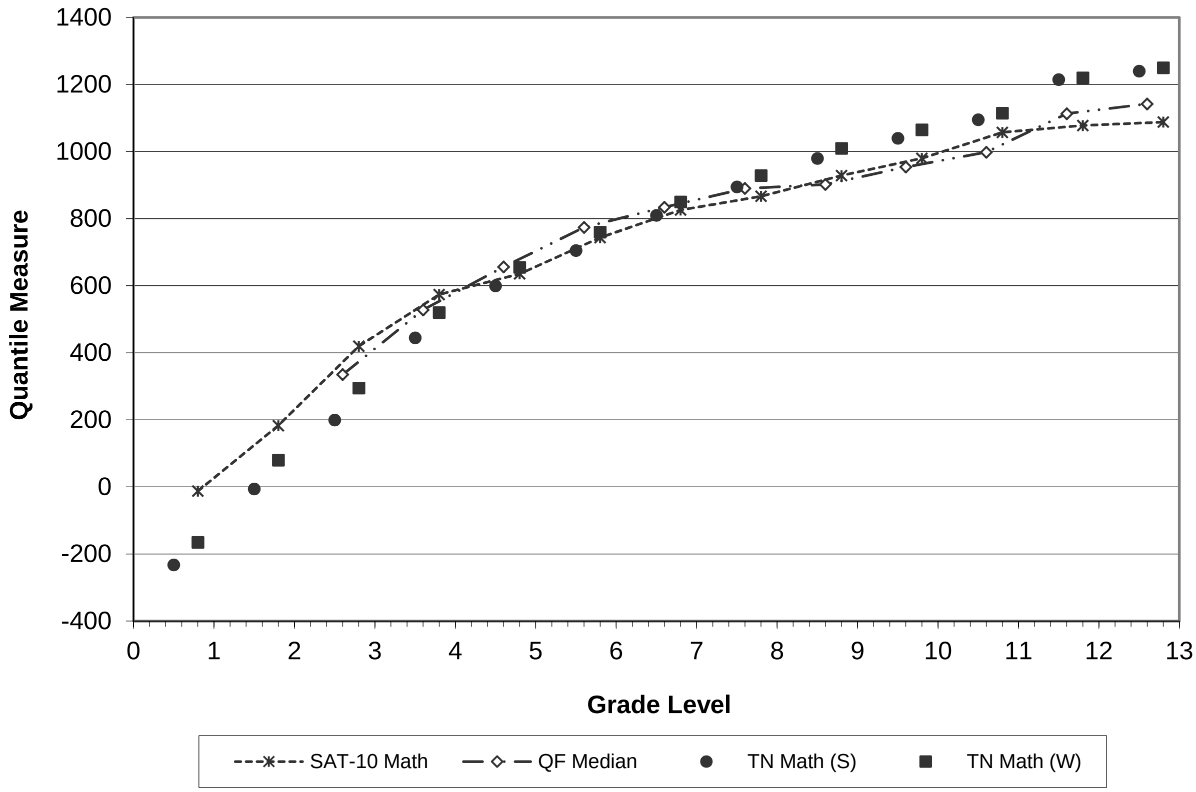
Figure 5. Relationship between reading and mathematics scale scores on a norm-referenced assessment linked to the Lexile scale in reading.



Based on an examination of *Figure 5*, it was concluded that the same conversion factor of 180 that is used with the Lexile scale could be used with the Quantile scale. Both sets of data exhibited a similar pattern across grades.

The second step in developing the Quantile scale with a fixed zero was to identify an anchor point for the scale. Given the number of students at each grade level in the field study, it was concluded that the scale should be anchored at grade 4 or 5 (middle of grade span typically tested by state assessment programs). Median performance at the end of grade 3 on the Lexile scale is 590L. The Quantile Framework field study was conducted in February and this point would correspond to six months (0.6) through the school year. Median performance at the end of grade 4 on the Lexile scale is 700L. To determine the location of the scale, 66Q were added to the median performance at the end of grade 3 to reflect the growth of students in grade 4 prior to the field study (700 – 590 = 110; 110 × 0.6 = 66). The value of 656Q was used for the location of grade 4 median performance. The anchor point was validated with other assessment data and collateral data from the Quantile Framework field study (see *Figure 6*).

Figure 6. Relationship between grade level and mathematics performance on the Quantile Framework field study and other mathematics assessments.



Finally, a linear equation of the form

$$[(\text{Logit} - \text{Anchor Logit}) \times \text{CF}) + 656 = \text{Quantile measure} \qquad \text{(Equation 5)}$$

was developed to convert logit difficulties to Quantile calibrations where the anchor logit is the median for grade 4 in the Quantile Framework field study.

Quantile Skill and Concept (QSC) Quantile Measures

In order to use the Quantile Framework to examine the difficulty of skills and concepts and the complexity of resources, the Quantile measure of each QSC must be estimated. The Quantile measure of a QSC estimates its solvability, or a prediction of how difficult the skill or concept will be for the learner with a Quantile measure of his or her own. The QSCs fall into knowledge clusters along a content continuum.

The Quantile measures and knowledge clusters for QSCs are determined by a group of three to five subject-matter experts (SMEs). Each SME has had classroom experience at multiple developmental levels, has completed graduate-level courses in mathematics education, and understands basic psychometric concepts and assessment issues.

Knowledge Clusters. Knowledge clusters are a family of skills, like building blocks, that depend one upon the other to connect and demonstrate how skills are founded, supported, and extended along the continuum. The knowledge clusters illustrate the interconnectivity of the Quantile Framework and the natural progression of mathematical skills (content progressions) needed to solve increasingly complex problems (Hudnutt, 2012).

Each QSC was classified as having "prerequisite" and "supplemental" QSCs or as being a "foundational" QSC by the SMEs. A *prerequisite QSC* is a QSC that describes a skill or concept that provides the foundation necessary for another QSC. For example, adding single-digit numbers is a prerequisite for adding two-digit numbers. A *supplemental QSC* is a QSC which describes supplementary skills or knowledge that assists and enriches the understanding of another QSC. An *impending QSC* describes the skills and concepts that will be built from a primary QSC and helps the teacher or parent to see a trajectory of knowledge across grades and content strands. The SMEs examined each QSC to determine where the specific QSC comes in the content progression based on classroom experience, instructional resources (e.g., textbooks), and other curricular frameworks (e.g., NCTM Standards). A QSC that is classified as "foundational" means this QSC describes a skill or concept that only requires readiness to learn. Readiness is based upon the learner's cognitive experiences rather than knowledge of specific mathematical concepts. It is the basis upon which other QSCs are built.

Once the knowledge cluster for a QSC was established, the information was used when determining the Quantile measure of a QSC (described below). If necessary, knowledge clusters were reviewed and refined if the Quantile measures of the QSCs in the cluster were not monotonically increasing or there was not an instructional explanation for the pattern.

Quantile measures of QSCs. To determine the Quantile measure of a QSC, actual performance by examinees was used. While expert judgment alone could have been used to scale the QSCs, empirical scaling was more replicable. Items and resulting data from two national field studies were used in the process:

- Quantile Framework field study (685 items, *N* = 9,647, grades 2 through Algebra II) which is described earlier in this section; and
- *PASeries* Mathematics field study (7,080 items, *N* = 27,329, grades 2 through 9/Algebra I) which is described in the *PASeries* Mathematics Technical Manual (MetaMetrics, 2005).

The items initially associated with each QSC were reviewed by SMEs and accepted for inclusion in the set of items, moved to another QSC, or not included in the set. The following criteria were used:

- Psychometric (responded to by at least 50 examinees, administered at the target grade level, point-biserial correlation greater than or equal to 0.16);
- Matched grade level of introduction of concept/skill from national review of curricular frameworks (described on pages 3 and 4); and,
- Appropriate for instruction of concept (first night's homework; from the A and B sections of the lesson problems) based on consensus of the SMEs.

Once the set of items meeting the inclusion criteria was identified, the set of items was reviewed to ensure that the curricular breadth of the QSC was covered. If the group of SMEs considered the set of items to be acceptable, then the Quantile measure of the QSC was calculated. The Quantile measure of a QSC is defined as the mean Quantile measure of items that met the criteria. The standard deviation of the item difficulties was also calculated (mean standard deviation of item difficulties across QSCs was 177.3Q). The final step in the process was to review the Quantile measure of the QSC in relationship to the Quantile measures of the QSCs identified as prerequisite and supplementary to the QSC. If the group of SMEs did not consider the set of items to be acceptable, then the Quantile measure of the QSC was estimated and assigned a Quantile zone. By assigning a Quantile zone instead of a Quantile measure to a QSC, the SMEs were able to provide a valid estimate of the skill or concept's difficulty.

QSC Quantile measures are used in the calibration of resources (e.g., textbooks, instructional materials, supplemental materials, workplace documents, everyday documents) used with the Quantile Framework.

Validation of The Quantile Framework for Mathematics

Validity is the extent to which a test measures what its authors or users claim it measures; specifically, test validity concerns the appropriateness of inferences "that can be made on the basis of observations or test results" (Salvia and Ysseldyke, 1998, p. 166).

The 1999 *Standards for Educational and Psychological Testing* (America Educational Research Association, American Psychological Association, and National Council on Measurement in Education) state that "validity refers to the degree to which evidence and theory support the interpretations of test scores entailed in the uses of tests" (p. 9). For the Quantile Framework, which measures student understanding of mathematical skills and concepts, the most important aspect of validity that should be examined is construct validity. The construct validity of The Quantile Framework for Mathematics can be evaluated by examining how well Quantile measures relate to other measures of mathematics described in the following sections.

Standardization set of items used with PASeries Mathematics. *PASeries* Mathematics is a series of classroom-based, progress monitoring assessments designed for use in the US school market in grades 3 through 8 (MetaMetrics, 2005). Each *PASeries* Mathematics assessment measures a range of mathematics skills appropriate to a specific grade. For each grade, *PASeries* Mathematics includes a screener test (pre-test) and six progress assessments designed to be administered approximately every six weeks. Each assessment contains 30 items; an assessment can be administered in one typical class period. As the school year progresses, each assessment is designed to be at a higher Quantile level, resulting in progressively more challenging tests.

For the standardization set, the items in the Quantile Framework field study that were also in the *PASeries* Mathematics field study were examined. Only items that were presented in exactly the same form in both studies were retained. A total of 213 items were identified that were administered in both studies. One item was dropped because none of the responses were correct, five items were dropped because they were too easy, and five items were dropped because there were presentation differences between the studies. The final number of items in the standardization set was 207. The test level breakdown is presented in *Table 6*.

Table 6. Number of items in the Quantile Framework standardization set by grade level of the item content.

Content Level of Items (by Grade)	Number of Items in Standardization Set
1	6
2	32
3	25
4	29
5	27
6	26
7	27
8	19
9	15
10	1

The relationship between the calibrations of the standardization set of items used in the Quantile Framework field study and on *PASeries* Mathematics (the calibration of the *PASeries* Mathematics items will be described later in this technical manual) was examined. The correlation of the Quantile measures of the 207 items was 0.92. The mean difference was -186Q and the standard deviation of the differences was 153Q. The standardization set of items is validated by consistency of measures between the two studies. Characteristics of the items in the standardization set from the two field studies are presented in *Figures 7* and *8*.

Figure 7. Comparison of the difficulty (Quantile measure) of the standardization set of items across two field studies.

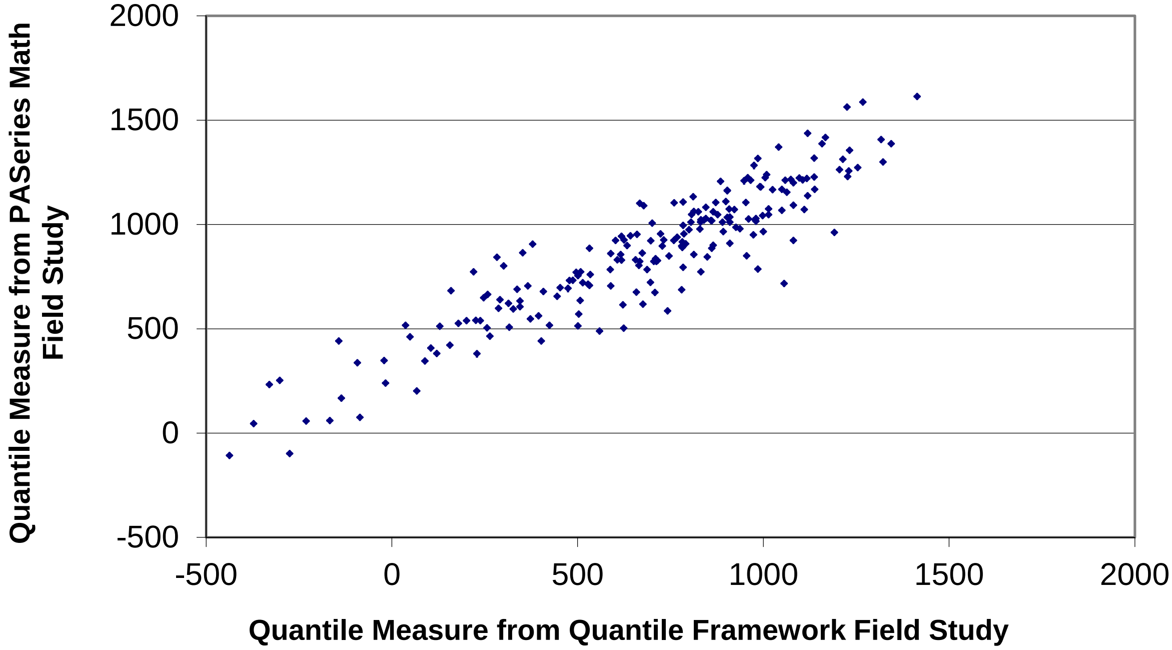
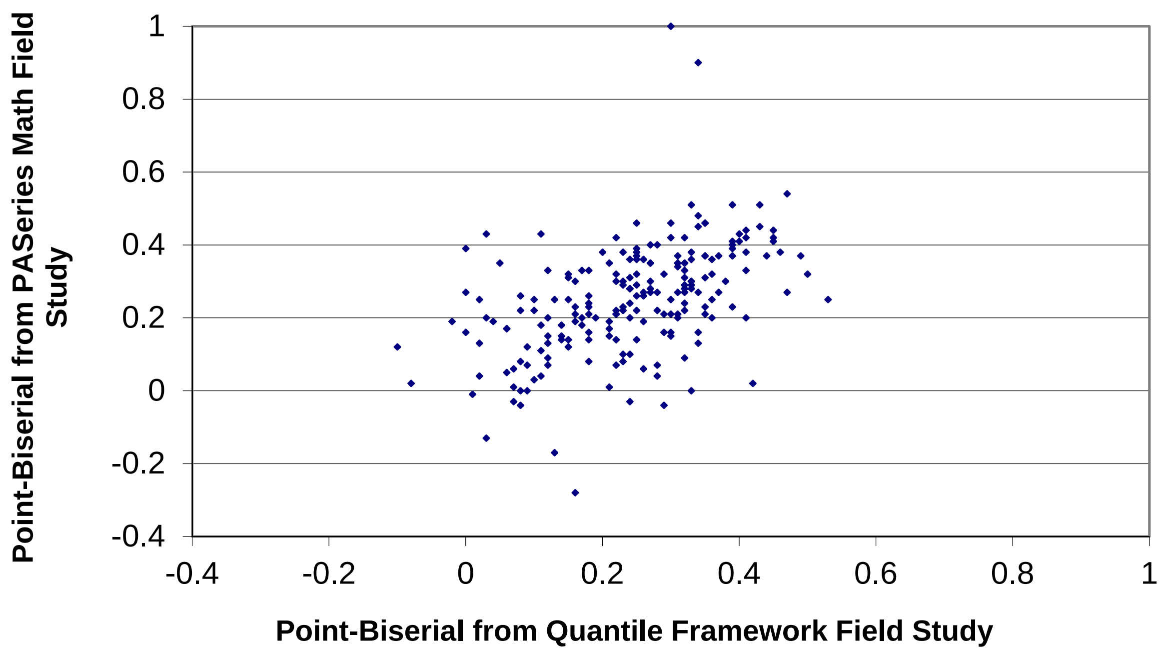


Figure 8. Comparison of the point-biserial correlations of the standardization set of items across two field studies.



The *PASeries* Math field study included 23,987 students who provided their grade level. *Table 7* shows the descriptive statistics for the sample by grade level. A monotonically increasing Quantile measure is observed across the grade levels.

Table 7. Mean and median Quantile measures for students with complete data from the *PASeries* Math field study (*N* = 23,987).

Grade Level	*N*	Mean Quantile measure	Median Quantile measure
3	4,703	370.46	370
4	4,478	592.29	598
5	3,871	696.54	690
6	2,813	788.32	771
7	3,555	827.24	816
8	3,481	884.81	874
9	1,086	970.24	967

Relationship of Quantile Measures to other Measures of Mathematical Ability. Scores from tests purporting to measure the same construct, for example "mathematical ability," should be moderately correlated (Anastasi, 1982). *Table 8* presents the results from field studies conducted with The Quantile Framework for Mathematics. For each of the tests listed, student mathematics scores were correlated with Quantile measures from the Quantile Framework field study.

Table 8. Results from studies conducted with The Quantile Framework for Mathematics.

Standardized Test	Grades in Study	*N*	Correlation Between Test Score and Quantile measure
RIT and Measures of Academic Progress (MAP by NWEA)	4 & 5	94	0.69
North Carolina End-of-Grade Tests (Mathematics)	4 & 5	341	0.73

Quantile Framework Linked to other Measures of Mathematics Understanding. The Quantile Framework for Mathematics has been linked to several standardized tests of mathematics achievement. When assessment scales are linked, a common frame of reference can be used to interpret the test results. This frame of reference can be "used to convey additional normative information, test-content information, and information that is jointly normative and content-based. For many test uses … [this frame of reference] conveys information that is more crucial than the information conveyed by the primary score scale" (Petersen, Kolen, and Hoover, 1989, p. 222).

Table 9 presents the results from linking studies conducted with the Quantile Framework. For each of the tests listed, student mathematics scores can also be reported as Quantile measures. This dual reporting provides a rich, criterion-related frame of reference for interpreting the standardized test scores. When a student takes one of the standardized tests, in addition to receiving her or his norm-referenced test results, s/he can receive information related to the specific QSCs that s/he is ready to receive instruction.

Table 9. Results from linking studies conducted with the Quantile Framework.

Standardized Test	Grades in Study	*N*	Correlation Between Test Score and Quantile measure
Mississippi Curriculum Test, Mathematics (MCT)	2 – 8	7,039	0.89
TerraNova (CTB/McGraw-Hill)	3, 5, 7, 9	6,356	0.92
Texas Assessment of Knowledge and Skills (TAKS)	3 – 11	14,286	0.69 to 0.78*
Proficiency Assessments for Wyoming Students (PAWS)	3, 5, 8, and 11	3,923	0.87
Progress Towards Standards (PTS3)	3-8 and 10	8,544	0.86 to 0.90*
Progress in Maths (PiM – GL Assessments)	1 – 8	3,183	0.71 to 0.81*
North Carolina End-of-Grade/End-of-Course Tests (NC EOG/NC EOC)	3, 5, 7, A1, G, and A2	5,069	0.88 to 0.90*
Kentucky Core Content Tests (KCCT)	3 - 8 and 11	12,660	0.80 to 0.83*
Oklahoma Core Competency Tests (OCCT)	3 – 8	5,649	0.81 to 0.85*
Iowa Assessments	2, 4, 6, 8, and 10	7,365	0.92
ReadiStep (The College Board)	8	2,183	0.83
Virginia Standards of Learning (SOL)	3-8, A1, G, and A2	12,470	0.86 to 0.89*
Kentucky Performance Rating for Educational Progress (K-PREP)	3 - 8	6,859	0.81 to 0.85*

Notes: * TAKS, PTS3, PiM, NCEOC, KCCT, OCCT, K-PREP, and SOL were not vertically scaled; separate linking equations were derived for each grade/course.

Multidimensionality of the Quantile Framework. Test dimensionality is defined as the minimum number of abilities or constructs measured by a set of test items. A construct is a theoretical representation of an underlying trait, concept, attribute, process, and/or structure that a test purports to measure (Messick, 1993). A test can be considered to measure one latent trait, construct, or ability (in which case it is called unidimensional); or a combination of abilities (in which case it is referred to as multidimensional). The dimensional structure of a test is intricately tied to the purpose and definition of the construct to be measured. It is also an important factor in many of the model(s) used in data analyses. Though many of the models assume unidimensionality, this assumption cannot be strictly met because there are always other cognitive, personality, and test-taking factors that have some level of impact on test performance (Hambleton and Swaminathan, 1985).

Study 1 – Comparison of Mathematics with Reading. The multidimensionality of the Quantile scale was examined using the Principal Components Analysis of Residuals in Winsteps (PRCOMP=S). The items were renamed with the strand number first for ease in review of the output. A three-step process was undertaken in order to examine the results and provide a context for interpreting the results.

The first step in the process was to run the Principal Components Analysis on all Quantile Framework field study items (*N* = 898). Next, the residual matrix was factor analyzed. *Table 10* shows the output from the analysis. The variance that is unexplained by the first factor (the Rasch measurement model) is 0.2% of the residual variance or 2.5 items of information. Based upon this set of data, it cannot be concluded that mathematics achievement as measured by the Quantile scale is multidimensional. The results supported the use of a unidimensional item response model on the items.

Table 10. Principal components analysis and distribution of variance explained by the model with the Quantile Framework field-study mathematics items (*N* = 685).

Source	Standardized Residual Variance (in Eigenvalue units)	Empirical	Modeled
Total Variance in Observations	1327.4	100.0%	100.0%
Variance Explained by Measures	642.4	48.4%	49.9%
Unexplained Variance (Total)	685.0	51.6%	50.1%
Unexplained Variance Explained by 1st Factor of the Residual Matrix	2.5	0.2%	

Next, the items were ordered by factor loading. Based on an examination of the item names with strand listed first, there did not appear to be any effect of strand. Only 6 items out of the 685 unique items had loadings above 0.30 on the first residual factor. These six items were all level 10 (Geometry) items and were from both strands 2 (Geometry) and 3 (Algebra).

To better understand the values produced in the first analysis, a second analysis was undertaken. The Level 5 (Grade 5) Quantile items were analyzed separately. The results are presented in *Table 11*.

Table 11. Principal components analysis and distribution of variance explained by the model with the Grade 5 Quantile Framework field-study mathematics items (N = 65).

Source	Standardized Residual Variance (in Eigenvalue units)	Empirical	Modeled
Total Variance in Observations	118.1	100.0%	100.0%
Variance Explained by Measures	53.1	45.0%	45.9%
Unexplained Variance (Total)	65.0	55.0%	54.1%
Unexplained Variance Explained by 1st Factor of the Residual Matrix	1.8	1.5%	

Three examples in the research literature describe the investigation of reading as a unidimensional construct: the 1940s Davis Study (Davis, 1944; Thurstone, 1946), the 1970s Anchor Study (Rentz and Bashaw, 1975, 1977; Jaeger, 1973; Loret, Seder, Bianchini, and Vale, 1974), and five 1980s and 1990s studies examining research conducted by ETS (Kirsch & Jungeblut and their colleagues, 1993, 1994; Reder, 1996; Salganik & Tal, 1989; Zwick, 1987). Other more recent examples include Harvey Goldstein's research with PISA (November 17, 2003), research on the development of the North Carolina End-of-Grade Tests (NCDPI, 1996), and research with the 2003 Maryland School Assessment—Reading. All of the studies confirm the assumption of unidimensionality of the reading assessments. Since most research concludes that reading is a unidimensional construct, for comparison purposes, a set of reading grade 5 reading items was also analyzed. The results are presented in *Table 12*.

Table 12. Principal components analysis and distribution of variance explained by the model with Grade 5 reading comprehension items (*N* = 54).

Source	Standardized Residual Variance (in Eigenvalue units)	Empirical	Modeled
Total Variance in Observations	137.1	100.0%	100.0%
Variance Explained by Measures	83.1	60.6%	62.1%
Unexplained Variance (Total)	54.0	39.4%	37.9%
Unexplained Variance Explained by 1st Factor of the Residual Matrix	2.0	1.5%	

The Rasch model explains 60.6% of the variance in the reading comprehension items. Along with the results presented in *Tables 11* and *12*, these data are consistent with the use of a unidimensional item response theory model for each of the analyses (reading and mathematics).

Finally, items from strands 2 (geometry) and 3 (algebra) were analyzed. It was hypothesized, that if multi-dimensionality were to be evidenced in the data, this would be the most likely contrast. The Winsteps analysis using all 296 of the strand 2 and 3 items in all of the forms did not appear to have any connectivity (common item) problems.

Table 13. Principal components analysis and distribution of variance explained by the model with the Strand 2 and 3 Quantile Framework field-study mathematics items (*N* = 296).

Source	Standardized Residual Variance (in Eigenvalue units)	Empirical	Modeled
Total Variance in Observations	644.7	100.0%	100.0%
Variance Explained by Measures	348.7	54.1%	55.5%
Unexplained Variance (Total)	296.0	45.9%	44.5%
Unexplained Variance Explained by 1st Factor of the Residual Matrix	2.3	0.4%	

Given the larger number of items in the analyses (296 in *Table 13* compared to 65 when only the Grade 5 items were examined in *Table 11*), the Rasch model explains 54.1% of the variance in the geometry (strand 2) and algebra (strand 3) items. The results presented in *Tables 10* and *11* are consistent with the interpretation of a single construct for each of the analyses (reading and mathematics).

Study 2 – Burg 2007. A study conducted by Burg (2007) analyzed the dimensional structure of mathematical achievement tests aligned to the NCTM content standards. Since there is no consensus within the measurement community on a single method to determine dimensionality, Burg employed four different methods for assessing dimensionality: (1) exploring the conditional covariances (DETECT), (2) assessment of essential unidimensionality (DIMTEST), (3) item factor analysis (NOHARM), and (4) principal component analysis (WINSTEPS). All four approaches have been shown to be effective indices of dimensional structure. Burg analyzed Grades 3 through 8 data from the Quantile Framework field study previously described.

Each set of on-grade items for a test form from Grades 3 through 8 were analyzed for possible sources of dimensionality related to the five mathematical content strands. The analyses were also used to compare test structures across grades. The results indicated that although mathematical achievement tests for Grades 3 through 8 are complex and exhibit some multidimensionality, the sources of dimensionality are not related to the content strands. The complexity of the data structure, along with the known overlap of mathematical skills, suggests that mathematical achievement tests could represent a fundamentally unidimensional construct. Therefore, while these sub-domains of mathematics are useful for organizing instruction, developing curricular materials such as textbooks, and describing the organization of items on assessments, they do not describe a significant psychometric property of the test or impact the interpretation of the test results. Mathematics, as measured by the Quantile Framework, can be described as one construct with various sub-domains.

Furthermore, these findings support the NCTM Connections Standard, which states that all students (prekindergarten through Grade 12) should be able to make and use connections among mathematical ideas and see how the mathematical ideas interconnect. Mathematics can be best described as an interconnection of overlapping skills with a high degree of correlation across the mathematical topics, skills, and strands.

Study 3 – Hennings and Simpson 2012. Results from Hennings and Simpson (2012) also suggest that the mathematics assessments used in MetaMetrics' linking studies are functionally unidimensional. Data from a Quantile Framework linking study involving the end-of-grade tests from a Southeastern state was examined. Scored student responses to items on the combined Quantile Linking Test and the state end-of-grade test were used. The end-of-grade tests had three polytomous items worth two points each on the forms for Grades 3 through 8, and one polytomous item worth four points on the forms for Grades 4 through 8. The remaining items on both tests were dichotomous and scored 0/1. *Table 14* shows the number of students and the number of items, combined and by test, for each grade.

Table 14. Number of items included in analyses

Grade	N of Students	Quantile Linking Test	End-of-Grade Test	Total
3	897	40	47	87
4	1,161	42	48	90
5	1,029	46	48	94
6	1,327	44	48	92
7	1,475	43	48	91
8	933	47	48	95

The polychoric item correlation matrix was analyzed for each test and grade. Because the principal components method of factor extraction in SAS does not require a positive-definite correlation matrix as input, principal component analyses were conducted instead of factor analyses.

The results support treating the data as unidimensional. The first component was dominant in all analyses. The first eigenvalue accounted for greater than 20% of the total variance in the analyses. Ratios of first-to-second eigenvalues ranged from approximately 6 to slightly over 9 (Gorsuch, 1983; Reckase, 1979). Secondary dimensions, i.e., the second and third components, accounted for approximately 5 - 6.5% of the total variance for each grade. *Table 15* lists the eignevalues for the first five principal components by grade, *Table 16* shows the ratios of first-to-second eigenvalues, and *Table 17* shows the proportion of variance accounted for by the first five principal components for each grade.

Table 15. Eigenvalues for the first five principal components.

	Principal Components				
Grade	**1**	**2**	**3**	**4**	**5**
3	24.152	3.463	2.411	2.253	2.011
4	23.252	3.637	2.257	1.894	1.829
5	22.770	3.222	2.407	2.239	1.935
6	21.400	3.058	2.297	2.185	1.866
7	23.919	3.922	2.442	1.744	1.648
8	24.572	2.654	2.152	2.076	1.914

Table 16. Ratio of the first-to-second eigenvalues by grade.

Grade	Ratio
3	6.975
4	6.394
5	7.066
6	6.997
7	6.099
8	9.257

Table 17. Proportion of variance explained for the first five principal components by grade.

	Principal Components				
Grade	**1**	**2**	**3**	**4**	**5**
3	0.278	0.040	0.028	0.026	0.023
4	0.258	0.040	0.025	0.021	0.020
5	0.242	0.034	0.026	0.024	0.021
6	0.233	0.033	0.025	0.024	0.020
7	0.263	0.043	0.027	0.019	0.018
8	0.259	0.028	0.023	0.022	0.020

The NC READY EOG Mathematics/EOC Algebra I/Integrated I - Quantile Framework Linking Process

Description of the Assessments

North Carolina READY EOG Mathematics and EOC Algebra I/Integrated I Assessments. North Carolina READY EOG Mathematics and EOC Algebra I/Integrated I Assessments measure students' proficiency based upon the Common Core State Standards for Mathematics (CCSSM) adopted by North Carolina in 2010. The EOG assessments are administered annually to students in Grades 3 through 8. The Algebra I/Integrated I assessment is administered at the end of the course to students enrolled in Algebra I or Integrated Math I. Each assessment consists of items that were written for specific content standards and demand one or more of the eight Standards for Mathematical Practice that are described in the CCSSM at every grade level (NCDPI, 2013c).

The NC Ready EOG Mathematics for Grades 3 and 4 consist of 54 items with 27 calculator inactive items and 27 calculator active items. The structure of the Grades 3 and 4 assessments consist entirely of multiple-choice items with four-response options. For the Grade 5 assessment, the calculator inactive section includes 19 multiple-choice and 8 gridded-response items and the calculator active section includes 27 multiple-choice items. For the NC Ready EOG Mathematics at Grades 6, 7, and 8, the calculator-inactive section consists of 7 multiple-choice and 11 gridded-response items that require students to insert numeric answers. The calculator-active section has 42 multiple-choice items (NCDPI, 2013e). The NC READY EOG Mathematics assessments were not vertically scaled across grades. Each test has scale scores that range from 400 to 500. These scale scores cannot be compared directly from grade to grade.

Since the CCSSM is subdivided into domains, which are large groups of related standards, the test items reflect a distinct distribution from each domain. The following table distinguishes these allocations at the identified grade levels (NCDPI, 2013c).

Table 18. Summary of the NC READY EOG Mathematics assessment blueprint targets for test development.

Domain	**Grade 3**	**Grade 4**	**Grade 5**
Operations and Algebraic Thinking	30-35%	12-17%	5–10%
Number and Operations in Base Ten	5-10%	22-27%	22–27%
Number and Operations-Fractions	20-25%	27-32%	47–52%
Measurement and Data	22-27%	12-17%	10–15%
Geometry	10-15%	12-17%	2–7%
Total	**100%**	**100%**	**100%**

Table 18 (continued). Summary of the NC READY EOG Mathematics assessment blueprint targets for test development.

Domain	**Grade 6**	**Grade 7**	**Grade 8**
Ratios and Proportional Relationships	12-17%	22–27%	NA
The Number System	27-32%	7–12%	2-7%
Expressions and Equations	27-32%	22–27%	27-32%
Functions	NA	NA	22-27%
Geometry	12-17%	22–27%	20-25%
Statistics and Probability	7-12%	12–17%	15-20%
Total	**100%**	**100%**	**100%**

The NC READY EOC Algebra I/Integrated I contains 60 items with approximately 80% four-choice multiple-choice items and 20% gridded-response items that require students to insert numeric answers (NCDPI, 2013e). Ten of the NC READY EOC Algebra I/Integrated items are embedded into the test as field-test items. Each of the remaining 50 items count as one point toward the student score. The NC READY EOC Algebra I/Integrated I scale scores range from 200 to 300, and these scale scores are on a separate scale.

At the high school course level, the CCSSM categorizes the standards by conceptual categories rather than by a set of standards for each course. As a result, states have the option to determine their own sequence of the CCSSM with the intention of completing the entire set of CCSSM standards by the end of the third year of high school study. *Table 19* shows the distribution of the high school conceptual categories for the NC READY EOC Algebra I/Integrated I assessment.

Table 19. Conceptual category distributions for Algebra I/Integrated I EOC.

Conceptual Category	**Algebra I/Integrated I**
Number and Quantity	5-10%
Algebra	22-27%
Functions	35-40%
Geometry	10-15%
Statistics and Probability	15-20%
Total	**100%**

Assessment results will be used both for school and district accountability under the NC READY Accountability Model and for Federal reporting purposes (NCDPI, 2013c).

The Quantile Framework for Mathematics. The Quantile Framework was developed to assist teachers, parents, and students in identifying strengths and weaknesses in mathematics and forecast growth in overall mathematical achievement. Items and mathematical content are calibrated using the Rasch IRT model. The Quantile scale ranges from “EM” (Emerging Mathematician, 0Q and below) to above 1600Q. The Quantile Framework was developed to assess how well a student (1) understands the natural language of mathematics, (2) knows how to read mathematical expressions and employ algorithms to solve decontextualized problems, and, (3) knows why conceptual and procedural knowledge is important and how and when to apply it. The Quantile Framework Item Bank consists of multiple-choice items aligned with first grade content through Geometry, Algebra II, and Pre-calculus content and field tested with a national sample of students during the winter of 2004.

The Quantile Linking Test was constructed by aligning the items from the NC READY EOG Mathematics assessments for grades 3, 4, 6, and 8 with the Quantile Framework taxonomy of Quantile Skills and Concepts (QSCs). Based upon these target test reviews, previously tested items were used to develop each grade-level linking test. Each Quantile Linking Test reflects comparable material that is tested at each identified grade level of the NC READY EOG Mathematics. The Quantile Linking Tests for Grades 3 and 4 have 44 items (rather than the 54 items on the NC READY EOG Mathematics assessments) because of the 10 field-test items included in the NC READY EOG Mathematics assessments. The Quantile Linking Tests for Grades 6 and 8 have 50 items (rather that the 60 items on the NC READY EOG Mathematics assessments) because of the 10 field-test items included in the NC READY EOG Mathematics assessments.

The items used for the linking tests predominantly match the QSCs that were identified for each item in the target test. When an exact QSC match did not occur, the linking test used a different QSC that satisfied one or more of the following conditions:

1. The test item used a QSC that addressed the same North Carolina Core Standard as the target item.
2. The test item used a QSC that was a prerequisite to the matched QSC in the target test.
3. The test item was more appropriate for grade level or student expectations based on North Carolina Core Standards.

The Quantile Linking Tests for Grades 3 and 4 consisted of 44 multiple-choice items. The distribution of the content strands closely matched the distribution of the North Carolina Core domains for each grade level.

Table 20. Distributions of content strands Grades 3 and 4 Quantile Linking Tests.

Content Strand	Grade 3		Grade 4	
	Percent of Items	Number of Items	Percent of Items	Number of Items
Numbers and Operations	56%	25	63%	28
Geometry	5%	2	18%	8
Algebra/Patterns & Functions	14%	6	5%	2
Data Analysis & Probability	9%	4	5%	2
Measurement	16%	7	9%	4
Total	**100%**	**44**	**100%**	**44**

The Grade 3 Quantile Linking Test consisted of 9 calculator-inactive items and 35 calculator-active items. The Grade 4 test consisted of 11 calculator-inactive items and 33 calculator-active items.

The content of these tests did not require a reference sheet with formulas. In addition, no ancillary materials such as rulers or protractors were necessary. Calculators that are suggested for student use on this test were a four-function calculator that did ***not*** include the fraction key. Calculators were provided by the student or the school district for this assessment administration.

The Quantile Linking Tests for Grades 6 and 8 consisted of 50 multiple-choice items. The distribution of the content strands closely matched the distribution of the domains from the North Carolina Core standards.

Table 21. Distributions of content strands for Grades 6 and 8 Quantile Linking Tests.

Content Strand	Grade 6		Grade 8	
	Percent of Items	Number of Items	Percent of Items	Number of Items
Numbers and Operations	52%	26	16%	8
Geometry	4%	2	20%	10
Algebra/Patterns & Functions	18%	9	48%	24
Data Analysis & Probability	14%	7	12%	6
Measurement	12%	6	4%	2
Total	**100%**	**50**	**100%**	**50**

None of the items on the Grades 6 and 8 Quantile Linking Tests required ancillary materials or tools such as protractors, rulers, or compasses. These Quantile Linking Tests did include a formula sheet as a reference point for students to determine the formula necessary to solve a problem. Calculators were to be used only during the calculator-active sections of the linking tests. Grade 6 students could use a four-function or scientific calculator; and it was advisable to use the calculators they were accustom to using during instruction, but use must abide by the North Carolina restrictions for calculators. Grade 8 students could use a graphing calculator that is within the North Carolina calculator requirements. Calculators were provided by the students or by the school district.

The Algebra I/Integrated I Quantile Linking Test consisted of 50 items. The distribution of the content strands closely matched the distribution of the Conceptual Categories distribution based upon the alignment study of the NC Ready EOC Algebra I/Integrated I with the Quantile Framework taxonomy.

Table 22. Distributions of content strands Quantile Linking Test Algebra I/Integrated I.

Content Strands	Algebra I/Integrated I	Number of Items
Numbers and Operations	10%	5
Geometry	6%	3
Algebra/Patterns & Functions	62%	31
Data Analysis & Probability	16%	8
Measurement	6%	3
Total	**100%**	**50**

The Grade 3 linking test had 5 items in common with the Grade 4 linking test. The Grade 4 linking test had 12 items in common with one or more grade levels of Quantile Linking Tests. The Grades 6 and 8 linking tests each had approximately 12 items linked to one or more grade levels of the Quantile Linking Tests. The Algebra I/Integrated I EOC assessment had 11 items linked to Grade 8 and one of those items was also linked to Grade 6. These linked items were used to develop a continuum in the vertical scale for measuring student growth.

Each Quantile Linking Test had a mean Quantile measure that aligned with the NC READY EOG Mathematics/EOC Algebra I/Integrated I assessments content (Grade 3, 408Q; Grade 4, 626Q; Grade 6, 783Q; Grade 8, 965Q; and Algebra I/Integrated I, 1047Q). To the extent possible, the grade level at which each item on the Quantile Linking Test was initially calibrated matched the grade level of the NC READY EOG Mathematics/EOC Algebra I/Integrated I assessments. An exception to this guideline occurred when an item was to be used as an across-grade linking item and was selected from a higher or lower grade level.

Evaluation of the Quantile Linking Tests. After administration, the Quantile Linking Tests items were reviewed. The raw score descriptive statistics for all items and all students that took the Quantile Linking Tests are presented in *Table 23*.

Table 23. Descriptive statistics for the Quantile Linking Tests raw scores.

Grade/Course	*N**	Raw Score Mean (SD)	Minimum Score		Maximum Score	
			Observed	Possible	Observed	Possible**
3	2,109	30.92 (8.1)	0	0	44	44
4	2,201	25.72 (7.9)	0	0	44	44
6	2,310	28.58 (9.6)	0	0	48	49
8	1,916	27.56 (8.7)	3	0	49	50
Alg I/Int I	2,538	24.88 (9.5)	1	0	49	50
Total	11,074					

* N-size reflects the removal of 142 students for missing, unusable, or duplicate students.
** One item was removed from Grade 6.

Based on the item examination, one item was removed from the Grade 6 analysis, because of a printing error in the test booklet. Selected item statistics for the Quantile Linking Tests are presented in *Table 24*. While some items retained on the tests had low point-biserial correlations, the items performed adequately (average ability measure for the correct answer was highest compared to the average ability measures of the three distractors from Winsteps analyses).

Table 24. Item statistics from the development of the Quantile Linking Tests.

Grade/Course	*N** (Persons)	*N*** (Items)	Percent Correct Mean (Range)	Point-Biserial Range	Coefficient Alpha
3	2,109	44	70 (35 - 96)	0.17 - 0.61	0.900
4	2,201	44	58 (10 - 95)	0.10 - 0.55	0.882
6	2,310	49	58 (18 - 94)	0.11 - 0.57	0.905
8	1,916	50	55 (14 - 91)	0.03 - 0.49	0.875
Alg I/Int I	2,538	50	50 (14 - 84)	0.13 - 0.50	0.898
Total	11,074				

* N-size reflects the removal of 142 students for missing, unusable, or duplicate students.
** One item was removed from Grade 6.

Coefficient Alphas for each of the five Quantile Linking Tests, one for each grade/course, ranged from 0.875 to 0.905. These values indicate strong internal consistency reliability for each of the five tests and high consistency across the five tests.

Study Design

A single-group/common person design was chosen for this study (Kolen and Brennen, 2004). This design is most useful "when (1) administering two forms to examinees is operationally possible, (2) differential order effects are not expected to occur, and (3) it is difficult to obtain participation of a sufficient number of examinees in an equating study that uses the random groups design" (pp. 16–17). The Quantile Linking Tests were administered between April 29, 2013 and May 15, 2013, within two weeks of the administration of the NC READY EOG Mathematics/EOC Algebra I/Integrated I assessments.

Analysis of the NC READY EOG Mathematics/EOC Algebra I/Integrated I assessment/Quantile Linking Test Sample

The sample of students for the study was identified by the North Carolina Department of Public Instruction. The participating schools were located from across North Carolina with a total of 120 schools from 61 districts participating in the linking study.

Table 25 presents the number of students tested in the linking study and the percentage of students with complete data (both a NC READY EOG Mathematics/EOC Algebra I/Integrated I scale score and a Quantile Linking Test Quantile measure). A total of 10,903 students (Grades 3, 4, 6, 8, and Algebra I/Integrated I), or 98.9%, had both test scores. This sample will be referred to as the matched sample.

Table 25. Number of students sampled and number of students in the matched sample.

Grade/Course	NC READY EOG Math/EOC *N* Received	Quantile Linking Test *N*	Matched *N*	Matched Percent
3	104,035	2,090	2,069	99.0
4	111,463	2,197	2,181	99.3
6	112,688	2,308	2,283	98.9
8	109,639	1,901	1,868	98.3
Alg I/Int I	119,717	2,531	2,502	98.9
Total	557,542	11,027	10,903	98.9

All students and items were submitted to a Winsteps (Linacre, 2011) analysis using a logit convergence criterion of 0.0001 and a residual convergence criterion of 0.003.

To account for individual differences in motivation when responding to the two assessments, the sample set was trimmed. By grade, test scores from each of the assessments were rank ordered and then converted to percentiles. For each student, the difference in percentiles between the two assessments was examined. A screen of a 25-percentile-point difference was selected for all tests. This helped to minimize the number of students removed from the sample and maintain the characteristics of the distribution, while at the same time removing students that were obvious outliers on one or both of the assessments.

For the final sample of students used in the study, students in the matched sample with the following score patterns were removed:

- Accommodations that effect the construct being measured
  - AssistiveTechnology
  - Cranmer Abacus
- 100% correct on the Quantile Linking Test,
- Missing total score on the NC READY EOG Mathematics/EOC Algebra I/Integrated I assessment,
- Misfit to the Rasch model, or
- Showed greater than a 25-percentile-rank difference between the NC READY EOG Mathematics/EOC Algebra I/Integrated I scale scores and Quantile Linking Test Quantile measures within grade.

Table 26 shows, for each grade, the number of students (*N*) in the final sample and the percent each grade *N*-count represents of the original matched sample. Of the 10,903 students in the matched sample, 8,720 (80%) remained in the final sample. The table also summarizes the number of students (by grade) removed from analysis, and the reason for their removal.

Table 26. Comparison of matched sample and final sample and the reason for student removal.

Matched Sample		*N* Removed by Reason				Final Sample	
Grade/ Course	*N*	Accommodated Students	Misfit to Rasch	Scores*	Percentile Rank Difference	*N*	Percent of Matched Sample
3	2,069	2	97	15	251	1,704	82.4
4	2,181	4	177	5	280	1,715	78.6
6	2,283	2	24	1	376	1,880	82.3
8	1,868	0	22	0	340	1,506	80.6
Alg I/Int I	2,502	0	40	0	547	1,915	76.5
Total	10,903	8	360	21	1,794	8,720	80.0

* Note: Students with a 100% correct on the linking test or with an invalid NC READY EOG Mathematics/EOC Algebra I/Integrated I assessment score.

Table 27 presents the demographic characteristics of all students in the NC READY EOG Mathematics/EOC Algebra I/Integrated I state sample, the matched sample, and the final sample of students included in this study. The three samples are very similar.

Table 27. Percentage of students in the NC READY EOG Mathematics/EOC Algebra I/Integrated I state sample, the matched sample, and the final sample for selected demographic characteristics.

Student Characteristic	Category	State Sample N=557,542	Matched Sample N=10,903	Final Sample N=8,720
Grade or Course	3	18.7	19.0	19.5
	4	20.0	20.0	19.7
	6	20.2	20.9	21.6
	8	19.7	17.1	17.3
	Alg I/Int I	21.5	22.9	22.0
Gender	Female	49.3	49.6	49.8
	Male	50.6	50.4	50.2
	Unknown/not avail	0.1	0.0	0.0
Race/Ethnicity	American Indian	1.5	0.9	1.0
	Asian	2.7	2.4	2.4
	Black	25.5	28.6	27.9
	Hispanic	13.9	14.9	14.6
	Pacific Islander	0.1	0.1	0.1
	White	52.6	49.5	50.5
	Two or more	3.7	3.5	3.6
	N/A	0.1	0.1	0.1
LEP Status	Currently identified	6.2	6.7	6.8
	Exit by committee	0.0	0.0	0.0
	Exits LEP	5.1	5.7	5.7
	Never identified	88.5	87.5	87.4
	No Status	0.1	0.1	0.1
	Parental refusal of IPT testing	0.0	0.0	0.0
Student/Disability	Exited within 2 years	2.0	1.6	1.5
	Yes	9.6	9.7	9.8
	No	88.5	88.8	88.7

Student Characteristic	Category	State Sample N=557,542	Matched Sample N=10,903	Final Sample N=8,720
EC Code	Autism	0.5	0.4	0.4
	Deaf-Blindness	0.0	0.0	0.0
	Deafness	0.0	0.0	0.0
	Developmental Delay	0.1	0.1	0.0
	Hearing Impairment	0.1	0.1	0.1
	Intell. Disability - Mild	0.2	0.1	0.1
	Intell. Disability - Moderate	0.0	0.0	0.0
	Intell. Disability - Severe	0.0	0.0	0.0
	Orthopedic Impairment	0.0	0.0	0.0
	Other Health Impairment	2.3	2.6	2.6
	Serious Emotional Disability	0.4	0.3	0.3
	Specific Learning Disability	5.5	5.5	5.6
	Speech or Language Impairment	2.3	2.1	2.1
	Traumatic Brain Injury	0.0	0.0	0.0
	VI	0.0	0.0	0.0
	Multiple Disabilities	0.0	0.0	0.0
	Not Provided	88.5	88.8	88.7
Plan 504	Yes	1.2	1.0	1.0
	No	98.8	99.0	99.0
Word To Word Bilingual	Yes	0.0	0.0	0.0
	No	100.0	100.0	100.0
Acad/Intell Gifted - Reading	Yes	11.9	12.0	12.8
	No	88.1	88.0	87.2

Table 28 presents the descriptive statistics for the NC READY EOG Mathematics/EOC Algebra I/Integrated I scale score matched sample as well as the matched sample Quantile Linking Test Quantile measure. Evaluating the Quantile measures on the NC

READY EOG Mathematics/EOC Algebra I/Integrated I assessments and the Quantile Linking Tests show very comparable results. The correlations between the matched sample NC READY EOG Mathematics/EOC Algebra I/Integrated I scale scores and the matched sample Quantile measures range between 0.726 and 0.815. Based upon these correlations, it can be concluded that the two tests are measuring similar mathematics constructs.

Table 28. Descriptive statistics for the NC READY EOG Mathematics/EOC Algebra I/Integrated I scale scores and Quantile measures and the Quantile Linking Test, matched sample (*N* = 10,903).

Grade/ Course	*N*	Matched Sample NC READY EOG Mathematics/EOC Algebra I/Integrated I Scale Score Mean (SD)	Matched Sample Quantile Linking Test Quantile Measure Mean (SD)	*r*
3	2,069	449.55 (9.5)	641.96 (228,6)	0.815
4	2,181	449.01 (9.4)	718.73 (203.0)	0.794
6	2,283	449.40 (9.4)	866.78 (204.9)	0.797
8	1,868	447.93 (8.4)	1003.70 (183.3)	0.777
Alg I/Int I	2,502	251.65 (9.7)	1040.62 (202.5)	0.726
Total	10,903			

Table 29 presents the descriptive statistics of the final sample NC READY EOG Mathematics/EOC Algebra I/Integrated I assessment scale scores and the Quantile Linking Test Quantile measures. The correlations between the two scores range from 0.872 to 0.900. These correlations between the two scores are strong and higher than the matched sample.

Table 29. Descriptive statistics for the NC READY EOG Mathematics/EOC Algebra I/Integrated I scale scores and the Quantile Linking Test Quantile measures, final sample (*N* = 8,720).

Grade/ Course	*N*	Final Sample NC READY EOG Mathematics/EOC Algebra I/Integrated I Scale Score Mean (SD)	Final Sample Quantile Linking Test Quantile Measure Mean (SD)	*r*
3	1,704	449.21 (9.6)	637.16 (218.0)	0.900
4	1,715	449.74 (9.3)	738.67 (192.9)	0.890
6	1,880	449.82 (9.6)	884.12 (204.6)	0.896
8	1,506	448.36 (8.6)	1018.31 (187.0)	0.893
Alg I/Int I	1,915	251.90 (9.8)	1057.98 (205.3)	0.872
Total	8,720			

Figures 9 through *18* shows the relationship between the NC READY EOG Mathematics/EOC Algebra I/Integrated I scale scores and the Quantile Linking Test Quantile measures for the matched and final samples for each grade/course. The matched samples show more scatter than the final samples. In each grade/course, it can be seen that there is a linear relationship between the NC READY EOG Mathematics/ EOC Algebra I/Integrated I scale scores and the final sample Quantile measures reinforcing the use of linear equating.

Figure 9. Scatter plot of the NC READY EOG Mathematics scale scores and the Quantile Linking Test Quantile measures for the Grade 3 matched sample (*N* = 2,069).

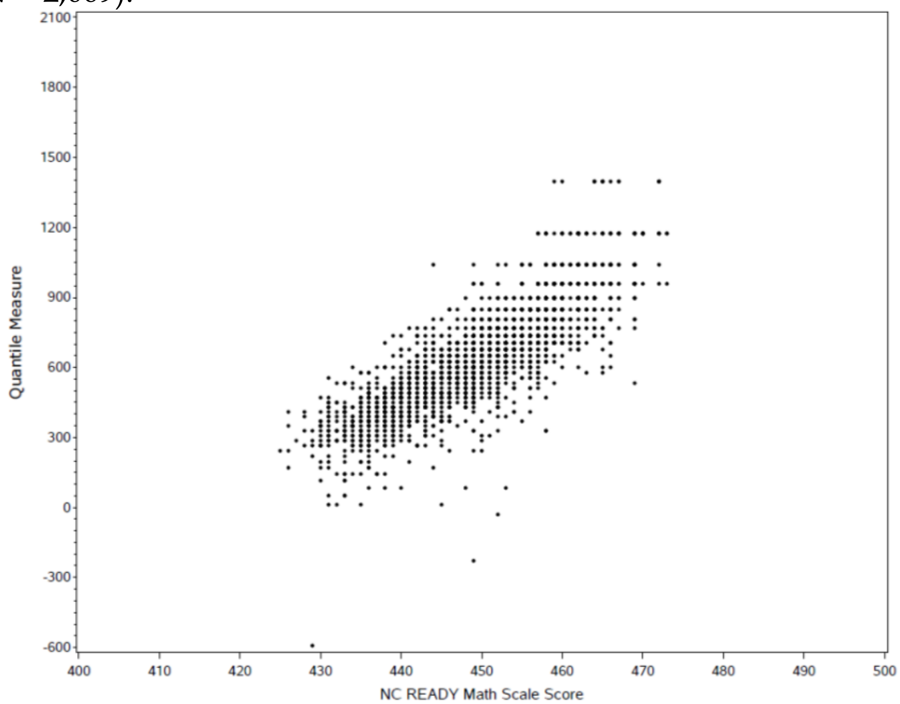


Figure 10. Scatter plot of the NC READY EOG Mathematics scale scores and the Quantile Linking Test Quantile measures for the Grade 3 final sample (*N* = 1,704).

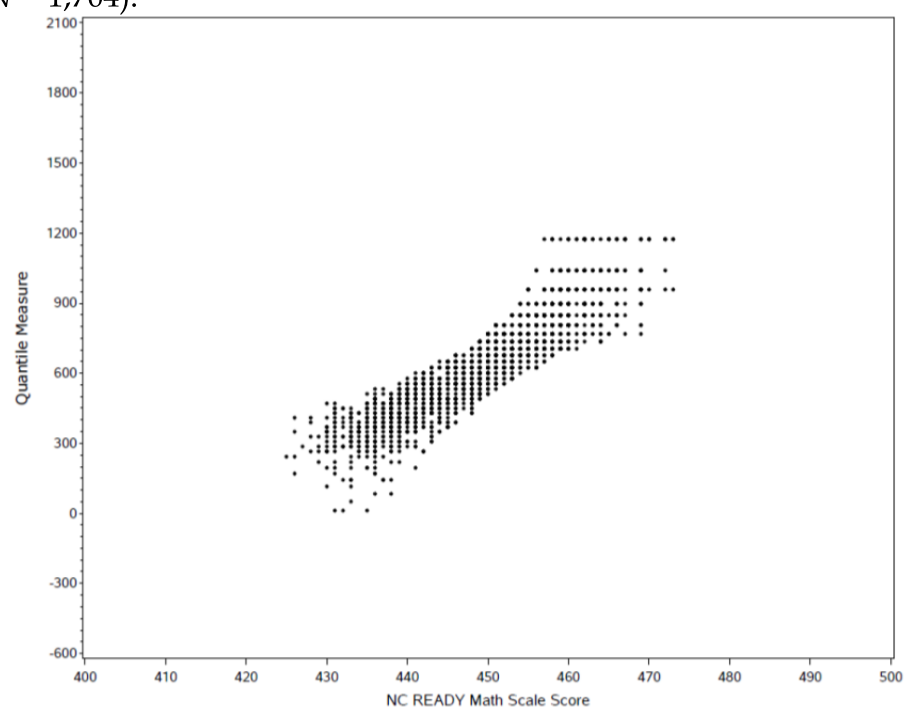


Figure 11. Scatter plot of the NC READY EOG Mathematics scale scores and the Quantile Linking Test Quantile measures for the Grade 4 matched sample (*N* = 2,181).

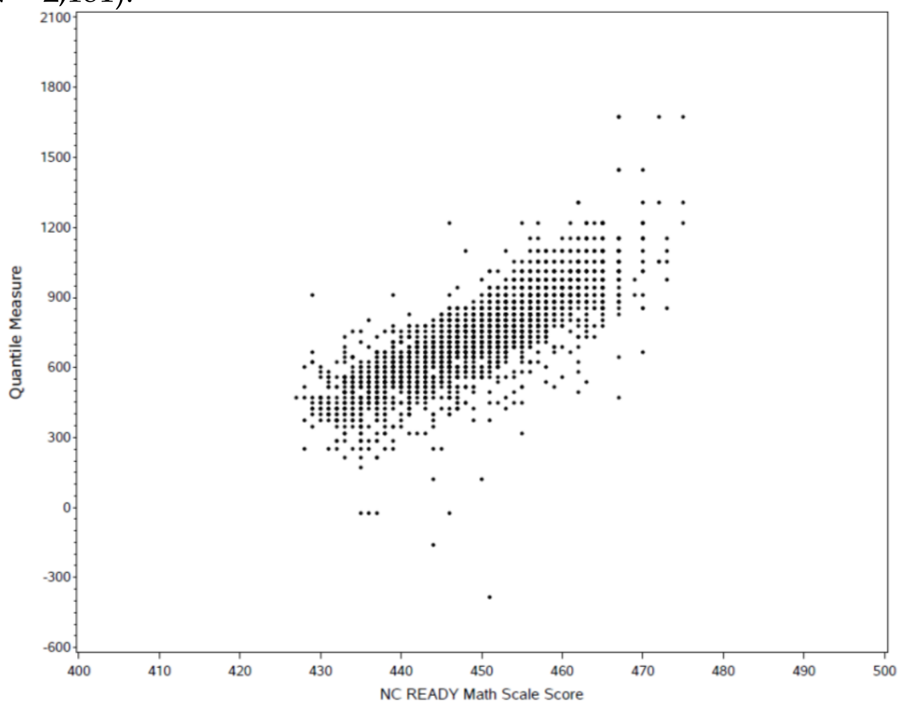


Figure 12. Scatter plot of the NC READY EOG Mathematics scale scores and the Quantile Linking Test Quantile measures for the Grade 4 final sample (*N* = 1,715).

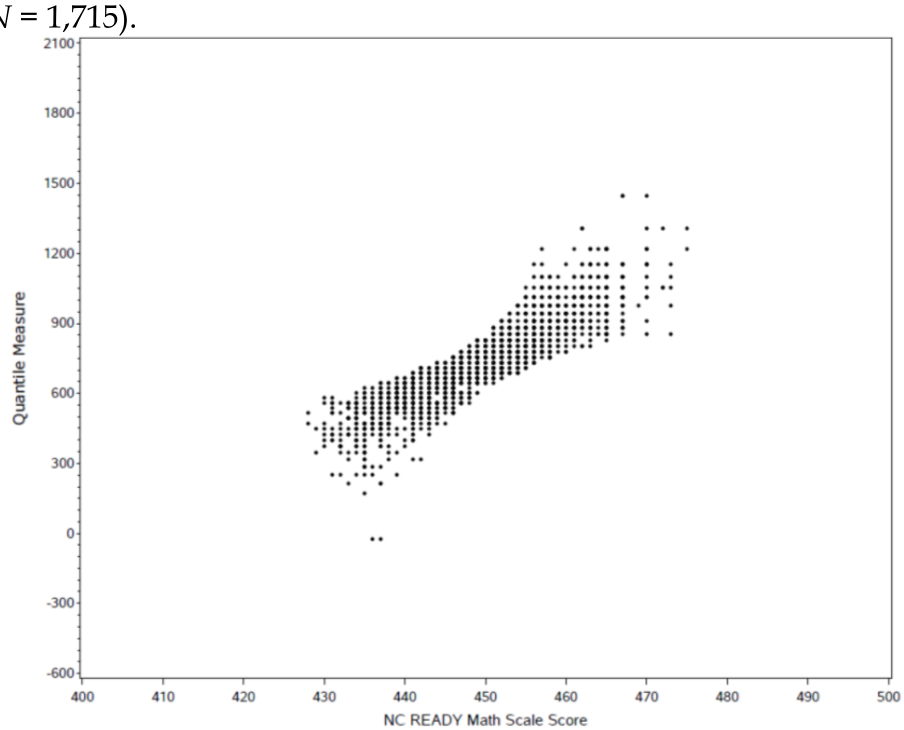


Figure 13. Scatter plot of the NC READY EOG Mathematics scale scores and the Quantile Linking Test Quantile measures for the Grade 6 matched sample (*N* = 2,283).

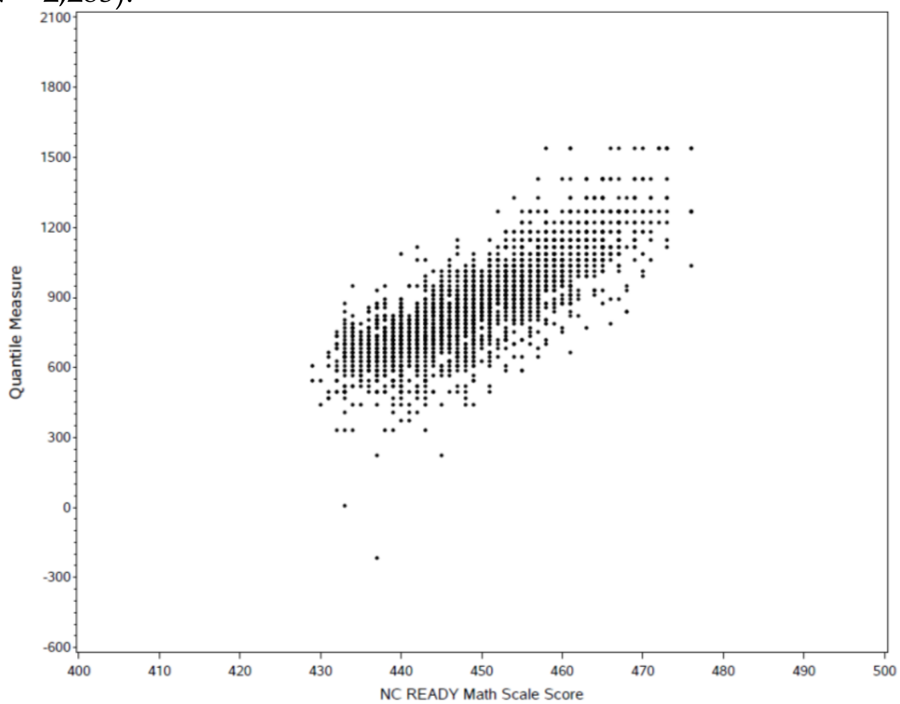


Figure 14. Scatter plot of the NC READY EOG Mathematics scale scores and the Quantile Linking Test Quantile measures for the Grade 6 final sample (*N* = 1,880).

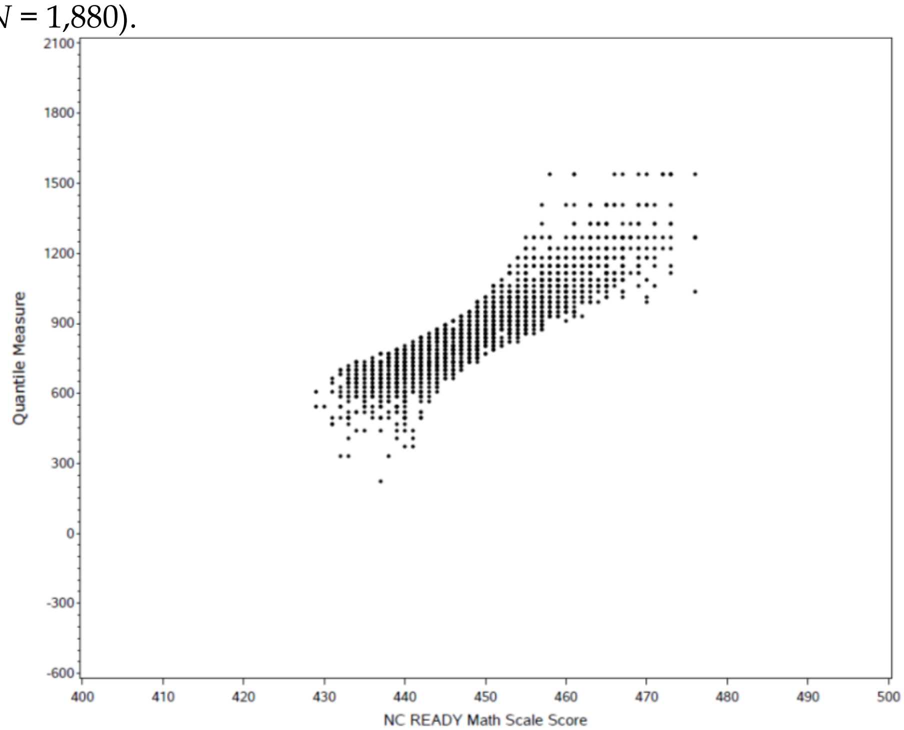


Figure 15. Scatter plot of the NC READY EOG Mathematics scale scores and the Quantile Linking Test Quantile measures for the Grade 8 matched sample (*N* = 1,868).

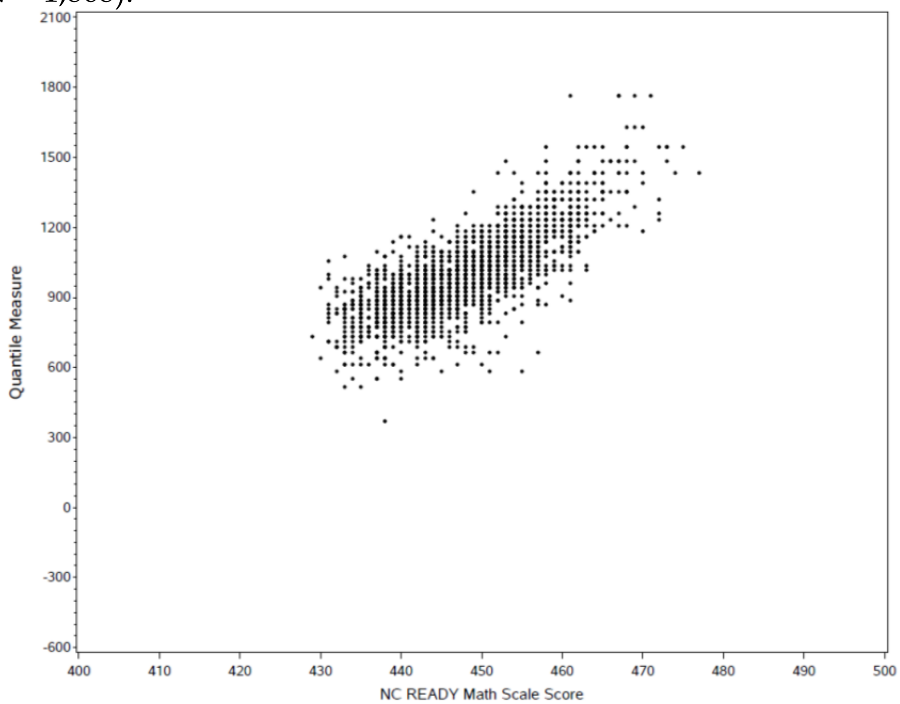


Figure 16. Scatter plot of the NC READY EOG Mathematics scale scores and the Quantile Linking Test Quantile measures for the Grade 8 final sample (*N* = 1,506).

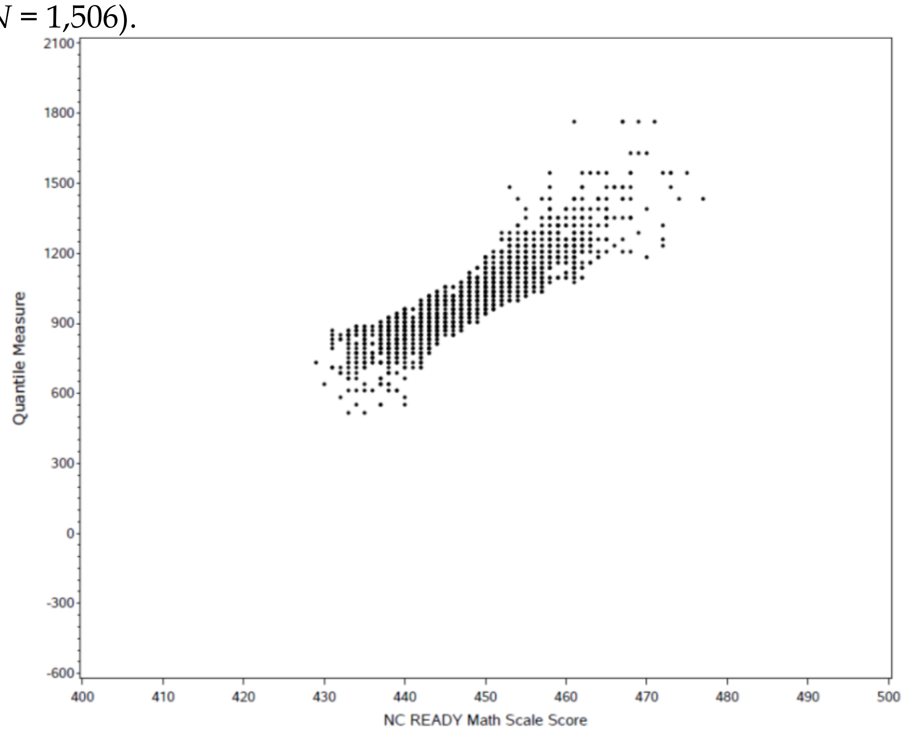


Figure 17. Scatter plot of the NC READY EOG EOC Algebra I/Integrated I scale scores and the Quantile Linking Test Quantile measures for the Algebra I/Integrated I matched sample (*N* = 2,502).

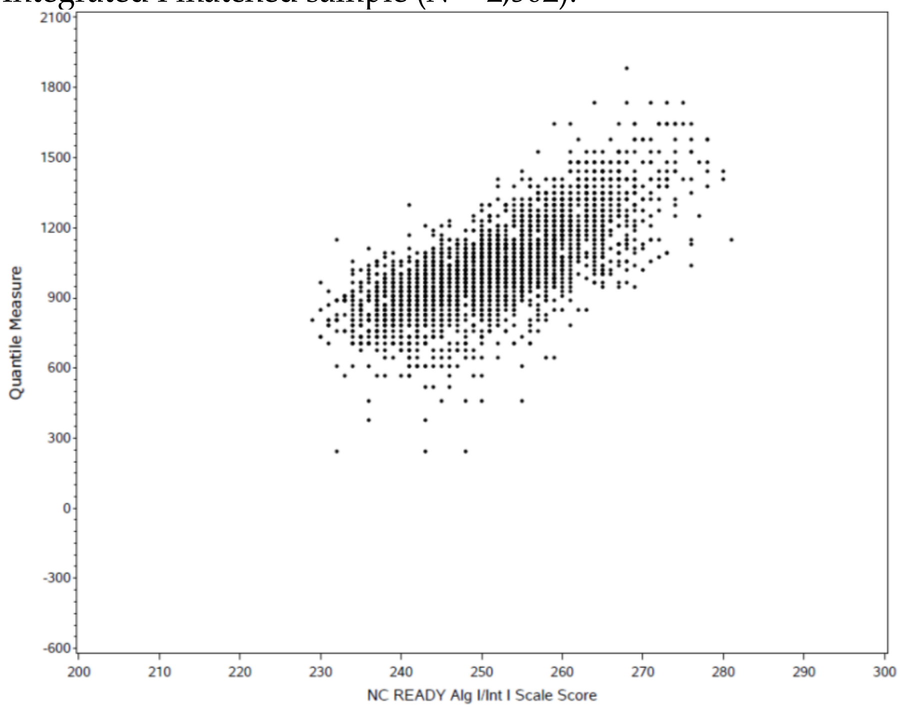
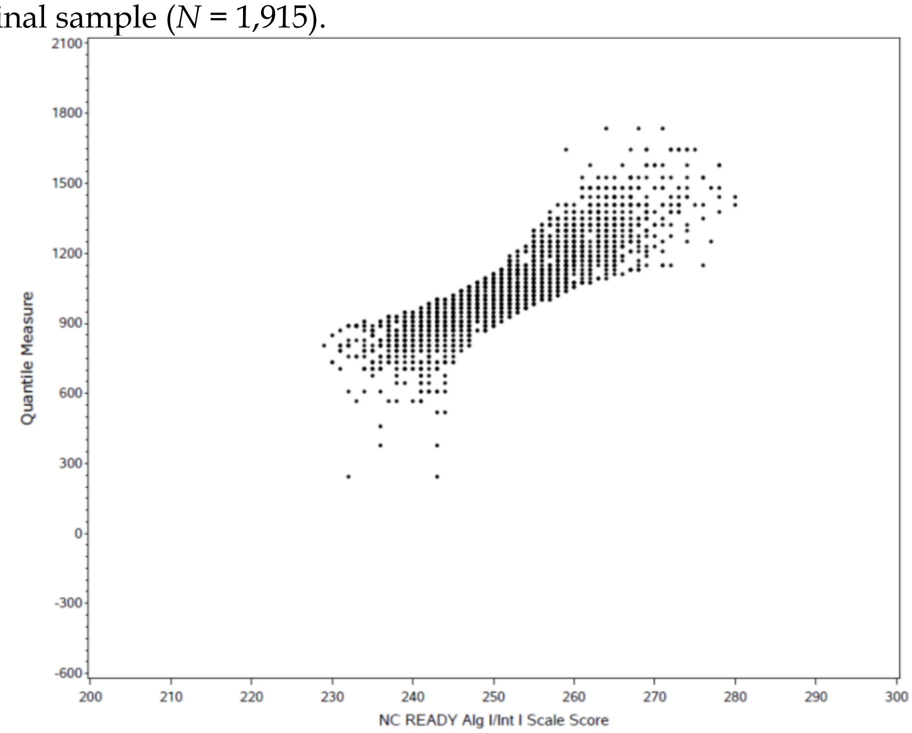


Figure 18. Scatter plot of the NC READY EOC Algebra I/Integrated I scale scores and the Quantile Linking Test Quantile measures for the Algebra I/Integrated I final sample (*N* = 1,915).



Linking the NC READY EOG Mathematics/EOC Algebra I/Integrated I Scale with the Quantile Scale

Linking in general means "putting the scores from two or more tests on the same scale" (National Research Council, 1999, p.15). This study was designed to provide information that could be used to match students' mathematical achievement with instructional resources—to identify the materials, concepts, and skills a student should be matched with for successful mathematical instruction, given their performance on the NC READY EOG Mathematics/EOC Algebra I/Integrated I assessments.

Linking Analyses. Two score scales (e.g., the Quantile Scale and the NC READY EOG Mathematics/EOC Algebra I/Integrated I assessment scales) can be linked using linear equating when the underlying item response models used to develop assessments are different. The linear equating method is most appropriate when (1) sample sizes are small; (2) test forms have similar difficulties; and (3) simplicity in conversion tables or equations, in conducting analyses, and in describing procedures are desired (Kolen and Brennan, 2004).

In linear equating, a transformation is chosen such that scores on two tests are considered to be equated if they correspond to the same number of standard deviations above (or below) the mean in some group of examinees (Angoff, 1984, cited in Petersen, Kohen, and Hoover, 1989; Kolen and Brennan, 2004). Given scores x and y on tests X and Y, the linear relationship is

$$\frac{(x-\mu_X)}{\sigma_X} = \frac{(y-\mu_y)}{\sigma_y} \qquad \text{(Equation 6)}$$

and the linear transformation l_x (called the SD line in this report) used to transform scores on test Y to scores on text X is

$$x = l_x(y) = \left(\frac{\sigma_X}{\sigma_y}\right)y + \left(\mu_x - \frac{\mu_y \sigma_X}{\sigma_y}\right) \qquad \text{(Equation 7)}$$

Linear equating using an SD-line approach is preferable to linear regression because the tests are not perfectly correlated. With less than perfectly reliable tests, linear regression is dependent on which way the regression is conducted: predicting scores on test X from scores on test Y or predicting scores on test Y from scores on test X. The SD line provides the symmetric linking function that is desired.

The final linking equation between the NC READY EOG Mathematics/EOC Algebra I/Integrated I scale scores and the Quantile scale can be written as:

$$\text{Quantile measure} = \text{Slope(NC READY EOG Mathematics/EOC Algebra I/Integrated I scale score)} + \text{Intercept} \quad \text{(Equation 8)}$$

where the slope is the ratio of the standard deviations of the NC READY EOG Mathematics/EOC Algebra I/Integrated I scale scores and Quantile Linking Test Quantile measures. These values can be found in *Table 29*.

Using the final sample data described in *Table 29*, the linear linking functions relating the NC READY EOG Mathematics/EOC Algebra I/Integrated I scale scores and Quantile measures for all students in the sample are presented in *Table 30*. Separate linking functions were developed for each grade/course of the NC READY EOG Mathematics/EOC Algebra I/Integrated I assessment since they are not on a vertical scale.

Because the original design for the NC READY mathematics assessments was to report results using a vertical scale across grades, no Quantile data was collected for Grades 5 and 7. During the calibration of the NC READY mathematics items for Grades 3 through 8 it was determined that a vertical scale could not be fitted (personal communication with NCDPI). Consequently, the Quantile measure equations needed to be estimated. Using a regression analysis, the Quantile means for Grades 5 and 7 were estimated using the means from the other grades' final samples. The standard deviations for Grades 5 and 7 were calculated using a pooled variance formula of the other grade's final sample data. The NC READY EOG Mathematics Grades 5 and 7 scale score means and standard deviations were calculated using the state data. The usual SD formulas for Grades 5 and 7 were derived using the means and standard deviations determined above.

Conversion tables were developed for each grade in order to express the NC READY EOG Mathematics/EOC Algebra I/Integrated I scores in the Quantile metric and were delivered to the North Carolina Department of Public Instruction in electronic format.

Table 30. Linear linking equation coefficients used to predict Quantile measures from the NC READY EOG Mathematics/EOC Algebra I/Integrated I scale scores.

Grade/Course	Slope	Intercept
3	22.740744	-9578.224
4	20.801171	-8616.395
5	21.092335	-8694.573
6	21.357151	-8722.812
7	20.836926	-8439.688
8	21.748657	-8733.002
Alg I/Int I	20.895137	-4205.586

Table 31 contains the capped Quantile measures by grade/course. The measures that are reported for an individual student should reflect the purpose for which they will be used. If the purpose is instructional, then the scores should be capped at the upper bound of measurement error (e.g., at the 95^{th} percentile point). In an instructional environment, all scores at or below 0Q should be reported as "EM" (Emerging Mathematician); no student should receive a negative Quantile measure.

Table 31. Capped values of the Quantile measure by grade/course.

Grade/Course	Capped Quantile Measure
3	975Q
4	1075Q
5	1125Q
6	1200Q
7	1325Q
8	1450Q
Alg I/Int I	1475Q

Validity of the NC READY EOG Mathematics/EOC Algebra I/Integrated I assessment – Quantile Link

Table 32 presents the descriptive statistics and effect size statistics of the NC READY EOG Mathematics/EOC Algebra I/Integrated I Quantile measures as well as the Quantile Linking Test Quantile measures for the final sample.

Table 32. Descriptive statistics and effect size statistics for the final sample NC READY EOG Mathematics/EOC Algebra I/Integrated I Quantile measures and the Quantile Linking Test Quantile measures.

Grade	*N*	Final Sample NC READY EOG Mathematics/EOC Algebra I/Integrated I Quantile Measure Mean (SD)	Final Sample Quantile Linking Test Quantile Measure Mean (SD)	Effect Size
3	1,704	637.16 (218.0)	637.15 (218.0)	0.000035
4	1,715	738.67 (192.9)	738.74 (192.9)	-0.000369
6	1,880	884.12 (204.6)	884.10 (204.6)	0.000099
8	1,506	1018.31 (187.0)	1018.30 (187.0)	0.000047
Alg I/Int I	1,915	1057.98 (205.3)	1057.99 (205.3)	-0.000035
Total	8,720			

The Hedges' *g* effect size shows the relationship between two variables or, in this case, between the NC READY EOG Mathematics/EOC Algebra I/Integrated I Quantile measure and the Quantile Linking Test Quantile measure. A guideline to use for interpretation of the effect size is:

Table 33. Interpretation chart for effect size.

Small	0.20
Medium	0.50
Large	0.80

For the five comparisons in *Table 32*, effect sizes were minimal for all comparisons indicating no significant difference between the NC READY EOG Mathematics/EOC

Algebra I/Integrated I Quantile measures and the Quantile Linking Test Quantile measures. This is because each grade/course has a unique linear equation.

Table 34 contains the percentile ranks of the Quantile Linking Test Quantile measures and the NC READY EOG Mathematics/EOC Algebra I/Integrated I assessment Quantile measures (based on the final sample). The criterion of a half standard deviation (100Q) on the Quantile scale was used to determine the size of the difference. In examining the values, the measures are very similar across the distributions. This supports the use of Quantile measures on the NC READY EOG Mathematics/EOC Algebra I/Integrated I assessments.

Table 34. Comparison of the Quantile measures for selected percentile ranks for the final sample Quantile Linking Test and the NC READY EOG Matematics/EOC Algebra I/Integrated I assessment.

Grade 3		
Percentile Rank	Linking Test Quantile Measure	NC READY EOG Math Sample Quantile Measure
1	170	200
5	308	269
10	369	337
25	470	473
50	624	655
75	806	814
90	958	928
95	1040	973
99	1174	1087

Grade 4		
Percentile Rank	Linking Test Quantile Measure	NC READY EOG Math Sample Quantile Measure
1	286	349
5	448	432
10	516	474
25	601	599
50	731	744
75	854	869
90	975	994
95	1053	1056
99	1219	1160

Table 34 (continued). Comparison of the Quantile measures for selected percentile ranks for the final sample Quantile Linking Test and the NC READY EOG Mathematics/EOC Algebra I/Integrated I assessment.

Grade 6		
Percentile Rank	Linking Test Quantile Measure	NC READY EOG Math Sample Quantile Measure
1	468	503
5	586	568
10	645	610
25	735	717
50	874	888
75	1013	1037
90	1146	1166
95	1268	1251
99	1407	1358

Grade 8		
Percentile Rank	Linking Test Quantile Measure	NC READY EOG Math Sample Quantile Measure
1	639	662
5	732	728
10	793	793
25	887	880
50	1017	1010
75	1138	1141
90	1259	1271
95	1353	1315
99	1545	1467

Algebra I/Integrated I		
Percentile Rank	Linking Test Quantile Measure	NC READY EOC Math Sample Quantile Measure
1	608	642
5	758	726
10	827	788
25	908	893
50	1038	1060
75	1188	1206
90	1348	1332
95	1408	1394
99	1577	1520

Performance standards provide a common meaning of test scores throughout a state or nation concerning what is expected at various levels of competence. The North Carolina Department of Public Instruction established four achievement levels: Level 1, Level 2, Level 3, and Level 4. As an example, the four achievement levels for the Grade 3 NC READY EOG Mathematics Assessment are (NCDPI, 2013b):

Level 1: Students performing at this level have **limited command** of the knowledge and skills contained in the *Common Core State Standards* (*CCSS*) for Mathematics assessed at grade 3 and are likely to need intensive academic support to engage successfully in further studies in this content area.
Level 2: Students performing at this level have **partial command** of the knowledge and skills contained in the *Common Core State Standards* (*CCSS*) for Mathematics assessed at grade 3 and are likely to need additional academic support to engage successfully in further studies in this content area.
Level 3: Students performing at this level have **solid command** of the knowledge and skills contained in the *Common Core State Standards* (*CCSS*) for Mathematics assessed at grade 3 and are academically prepared to engage successfully in further studies in this content area.
Level 4: Students performing at this level have **superior command** of the knowledge and skills contained in the *Common Core State Standards* (*CCSS*) for Mathematics assessed at grade 3 and are academically well prepared to engage successfully in further studies in this content area.

The four achievement levels for NC READY EOC Algebra I/Integrated I Assessment are (NCDPI, 2013a):

Level 1: Students performing at this level have a **limited command** of the knowledge and skills contained in the *Common Core State Standards* (*CCSS*) for Mathematics assessed at the end of Math I and will need academic support to engage successfully in more rigorous studies in this content area. They will also need continued academic support to become prepared to engage successfully in credit-bearing, first-year Mathematics courses without the need for remediation.
Level 2: Students performing at this level have a **partial command** of the knowledge and skills contained in the *Common Core State Standards* (*CCSS*) for Mathematics assessed at the end of Math I and will likely need academic support to engage successfully in more rigorous studies in this content area. They will also likely need continued academic support to become prepared to engage successfully in credit-bearing, first-year Mathematics courses without the need for remediation.
Level 3: Students performing at this level have **solid command** of the knowledge and skills contained in the *Common Core State Standards* (*CCSS*) for Mathematics assessed at the end of Math I and are academically prepared to engage successfully in more rigorous studies in this content area. They are also on track to become academically prepared to engage successfully in credit-bearing, first-year Mathematics courses without the need for remediation.

Level 4: Students performing at this level have a **superior command** of the knowledge and skills contained in the *Common Core State Standards* (*CCSS*) for Mathematics assessed at the end of Math I and are academically well-prepared to engage successfully in more rigorous studies in this content area. They are also on-track to become academically prepared to engage successfully in credit-bearing, first-year Mathematics courses without the need for remediation.

Table 35 presents the achievement level cut scores on the NC READY EOG Mathematics/EOC Algebra I/Integrated I assessments and the associated Quantile measures. The values in the table are the cut scores associated with the bottom score for each category.

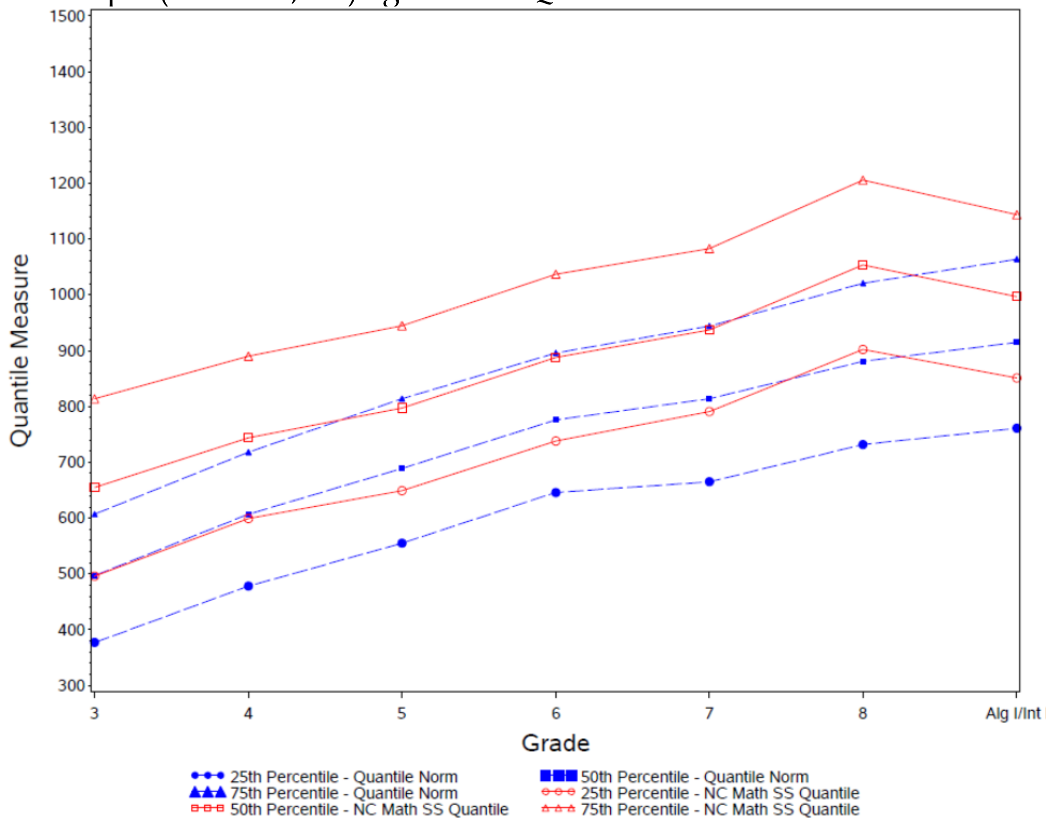
Table 35. Performance level cut scores on the NC READY EOG Mathematics/EOC Algebra I/Integrated I assessment and the associated Quantile measures.

Grade/ Course	Level 2		Level 3		Level 4	
	NC READY EOG Mathematics/ EOC Algebra I/Integrated I Scale Score	Quantile Measure	NC READY EOG Mathematics/ EOC Algebra I/Integrated I Scale Score	Quantile Measure	NC READY EOG Mathematics/ EOC Algebra I/Integrated I Scale Score	Quantile Measure
3	443	495Q	451	680Q	460	885Q
4	444	620Q	451	765Q	460	950Q
5	444	670Q	451	820Q	460	1010Q
6	447	825Q	453	950Q	461	1125Q
7	447	875Q	453	1000Q	461	1165Q
8	447	990Q	454	1140Q	463	1335Q
Alg I/Int I	247	955Q	253	1080Q	264	1310Q

The next graph shows the Quantile measures for the NC READY EOG Mathematics/EOC Algebra I/Integrated I assessments Quantile measures from the final sample and the Quantile norms. These norms were created based on linking studies conducted with the Quantile Framework. The sample's distribution of scores from this study was similar to the distribution of scores on norm-referenced assessments and other standardized measures of mathematics achievement. The results compared favorably with other mathematics measures which reinforced MetaMetrics' confidence in the Quantile norms.

As can be seen in *Figure 19*, the Quantile measures for the NC READY EOG Mathematics/EOC Algebra I/Integrated I assessments are higher than the Quantile measure norms. This indicates that the final sample in this study is more able than the samples used for the Quantile norms.

Figure 19. Selected Percentiles (25th, 50th, and 75th) plotted for the NC READY EOG Mathematics/EOC Algebra I/Integrated I Quantile measures for the final sample (*N* = N=8,720) against the Quantile measure norms.



The following box and whisker plots (*Figures 20, 21, 22,* and *23*) show the progression of scores (the *y*-axis) from grade to grade (the *x*-axis). (Note: Alg I/Int I is presented as Grade 9.) For each grade, the box refers to the interquartile range. The line within the box indicates the median and the • indicates the mean. The end of each whisker shows the minimum and maximum values of the Quantile Linking Tests Quantile measures and the NC READY EOG Mathematics/EOC Algebra I/Integrated I assessments Quantile measures for each grade (the *y*-axis). The Quantile measures are on a vertical scale and *Figures 20, 21, 22,* and *23* demonstrate this by showing that as the grade increases so do the Quantile scores on the NC READY EOG Mathematics/EOC Algebra I/Integrated I assessments. The pattern of Quantile measures is the same for each figure. *Figure 23* includes the performance levels of Level 2, Level 3, and Level 4 set by North Carolina.

Figure 20. Box and whisker plot of the Quantile Linking Tests Quantile measures by grade/course, final sample (*N* =8,720).

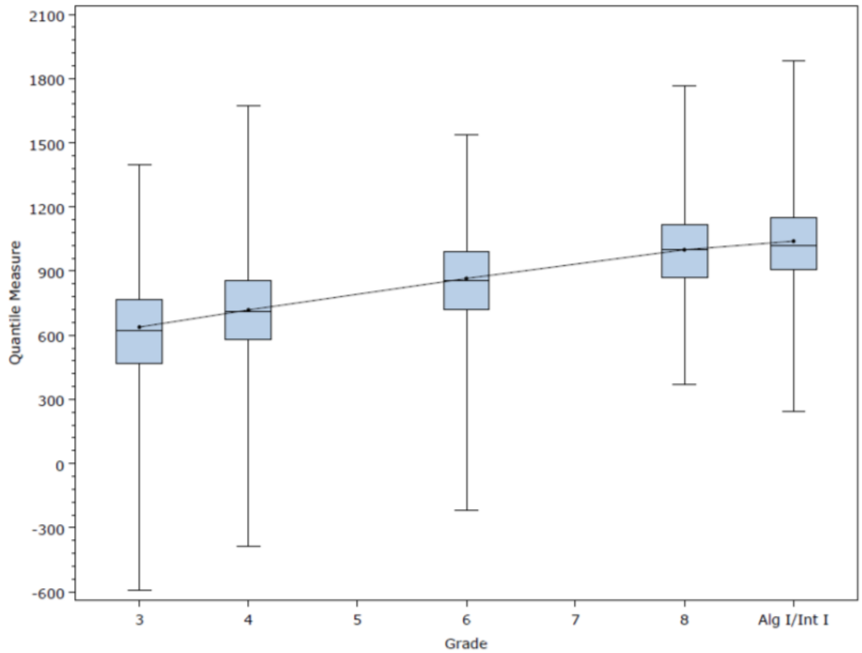


Figure 21. Box and whisker plot of the NC READY EOG Mathematics/EOC Algebra I/Integrated I Quantile measures by grade/course, matched sample (*N* = 10,903).

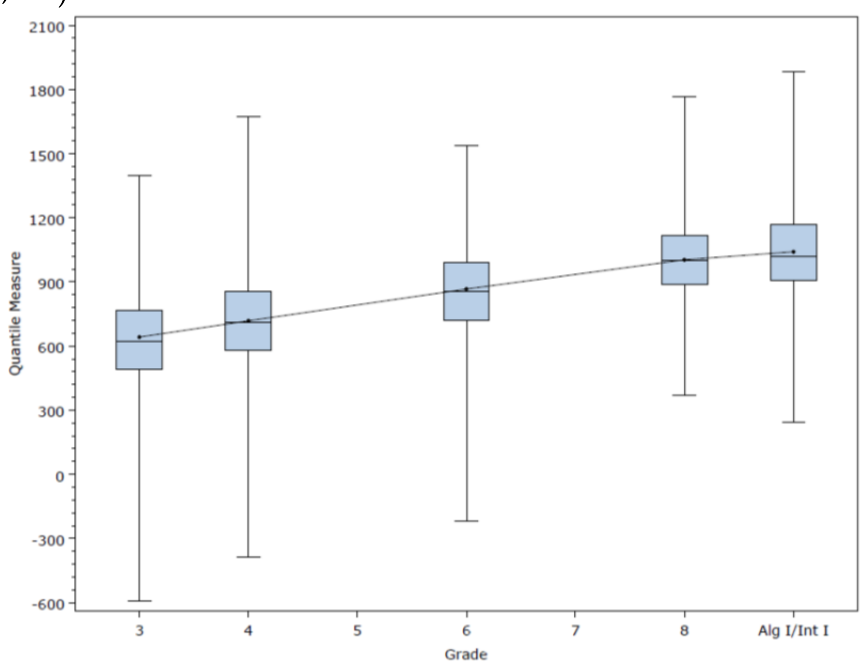


Figure 22. Box and whisker plot of the NC READY EOG Mathematics /EOC Algebra I/Integrated I Quantile measures by grade/course, final sample (*N* = 8,720).

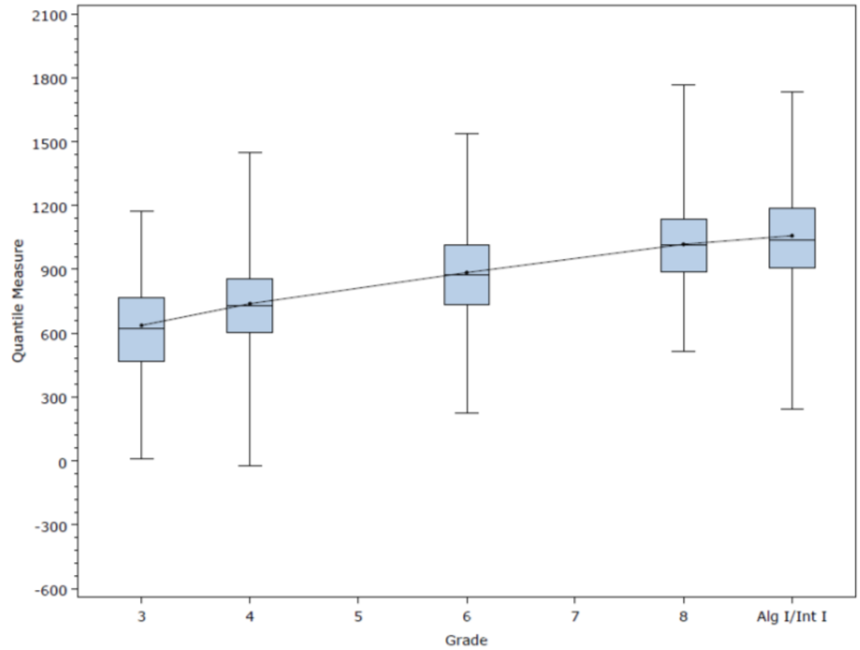
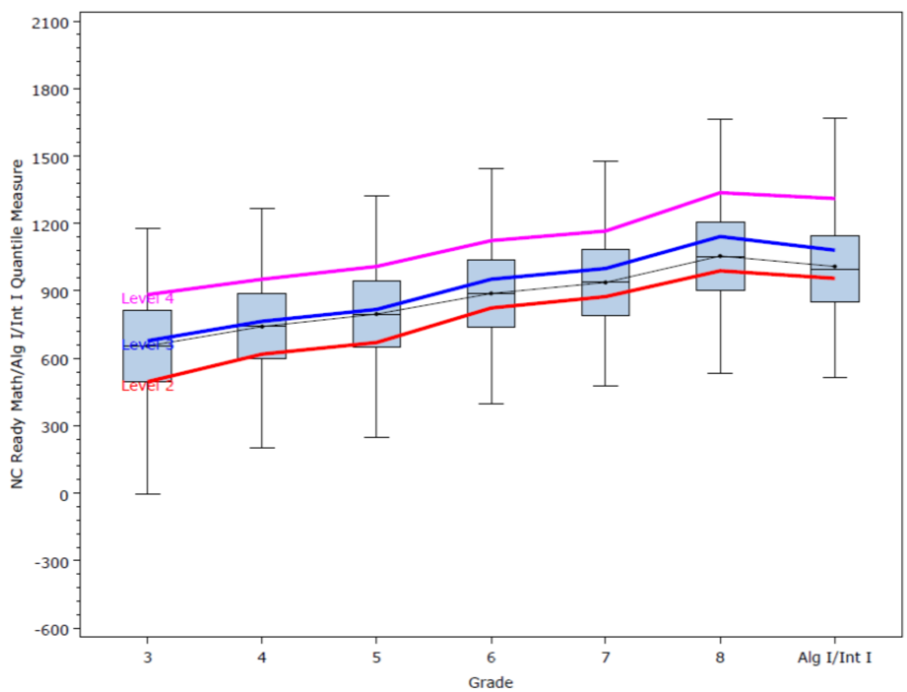


Figure 23. Box and whisker plot of the NC READY EOG Mathematics /EOC Algebra I/Integrated I Quantile measures with the performance standards by grade/course, state sample (*N* = 780,377).



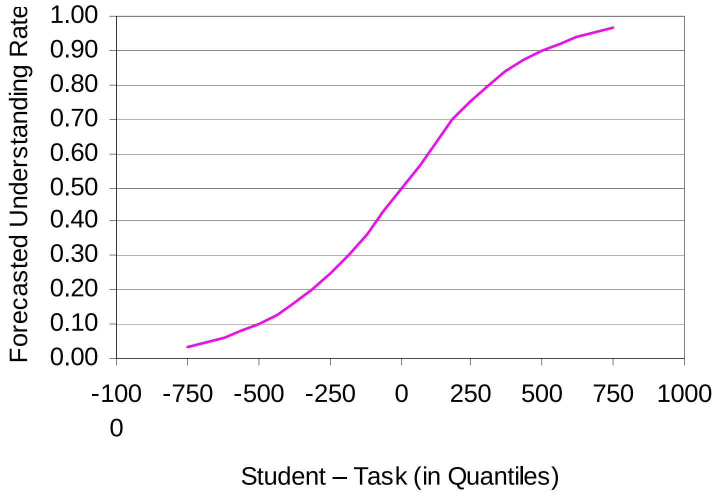
Quantile Framework and Instruction

Quantile measures are available from many norm-referenced and criterion-referenced assessments, in addition to state tests and instructional products. Students who take a mathematics achievement test that is linked with the Quantile Framework or one that reports directly in the Quantile metric will receive a Quantile measure. Educators can use these Quantile measures to match students, by *readiness level*, to level-appropriate instructional materials and forecast understanding. For example, a student with a Quantile measure of 500Q should be ready for instruction of mathematics problems at a demand level of 500Q.

Differentiated Instruction. A Quantile measure for materials is a number indicating the mathematical demand of the material in terms of the concept/application solvability. The Quantile measure for an individual student is the level at which he or she is ready for instruction (50% competency with the material) and has knowledge of the prerequisite mathematical concepts and skills necessary to succeed. The Quantile scale ranges from Emerging Mathematician (0Q and below) to above 1600Q. The Quantile measure does not relate to a specific grade, per se, so the score is developmental as it spans the mathematics continuum from kindergarten mathematics through the content typically taught in Algebra II, Geometry, Trigonometry, and Pre-calculus. The measure can be used by a teacher to determine what mathematical instruction the student is likely to be ready for next.

Figure 24 shows the general relationship between the student-task discrepancy and forecasted understanding. When the student measure and the task mathematical demand are the same (difference of 0Q), then the forecasted understanding, or success rate, is modeled as 50% and the student is likely ready for instruction on the skill or concept.

Figure 24. Relationship between student mathematical demand discrepancy and forecasted understanding (success rate).



An appropriate instructional range for the Quantile measure of a student is 50Q above and 50Q below the Quantile measure of the student (44% - 56% competency). This range identifies the "learning frontier" of mathematics skills in which a student has the prerequisite knowledge and skills needed to understand the instruction and will likely have success with tasks related to the skill/concept after this introductory instruction.

Quantile measures provide reliable, actionable results because instruction and assessment are described using the same metric. When instruction is measured at a unique mathematical level of understanding and any form of assessment can be reported using the same scale, equal levels of achievement are observed.

By understanding the interaction between student measures and resource measures (e.g., textbook lessons, instructional materials), any level of understanding can be used as a benchmark. An individual can modulate his or her own likely success rate by lowering the difficulty of the task (i.e., increase to 90% understanding) or increasing the difficulty of the task (i.e., lower to 40% understanding) depending on the situation (refer to *Figure 14*). This flexibility allows the teacher, parent, or student the ultimate control to modulate the fit between person and task.

The primary utility of the Quantile Framework is its ability to forecast what will likely happen when students confront resources and instruction on specific mathematical skills and concepts. With every application by teacher, student, or parent there is a test of the framework's accuracy. The framework makes a point prediction every time a resource or lesson is chosen for a student. Anecdotal evidence suggests that the Quantile Framework predicts as intended. That is not to say that there is an absence of error in forecasted understanding. There is error in resource measures based on QSC (mathematical skills and concepts) measures, student measures, and their difference modeled as forecasted understanding. However, the error is sufficiently small that the judgments about students, resources, and understanding rates are useful.

The subjective experience of 25%, 50%, and 75% understanding/success as reported by students varies greatly. A 1000Q student being instructed on 1000Q QSCs (50% understanding) has a successful instructional experience—he has the background knowledge needed to learn and apply the new information. Teachers working with such a student report that the student can engage with the skills and concepts that are the focus of the instruction and, as a result of the instruction, are able to solve problems utilizing those skills. In short, such students appear to understand what they are learning. A 1000Q student being instructed on 1200Q QSCs (25% understanding) encounters so many unfamiliar skills and difficult concepts that the learning is frequently lost. Such students report frustration and seldom engage in instruction at this level of understanding. Finally, a 1000Q student being instructed on 800Q QSCs (75% understanding) reports that he is able to engage with the skills and concepts with minimal instruction, is able to solve complex problems related to the skills and concepts, is able to connect the skills and concepts with skills and concepts from other strands, and experiences fluency and automaticity of skills.

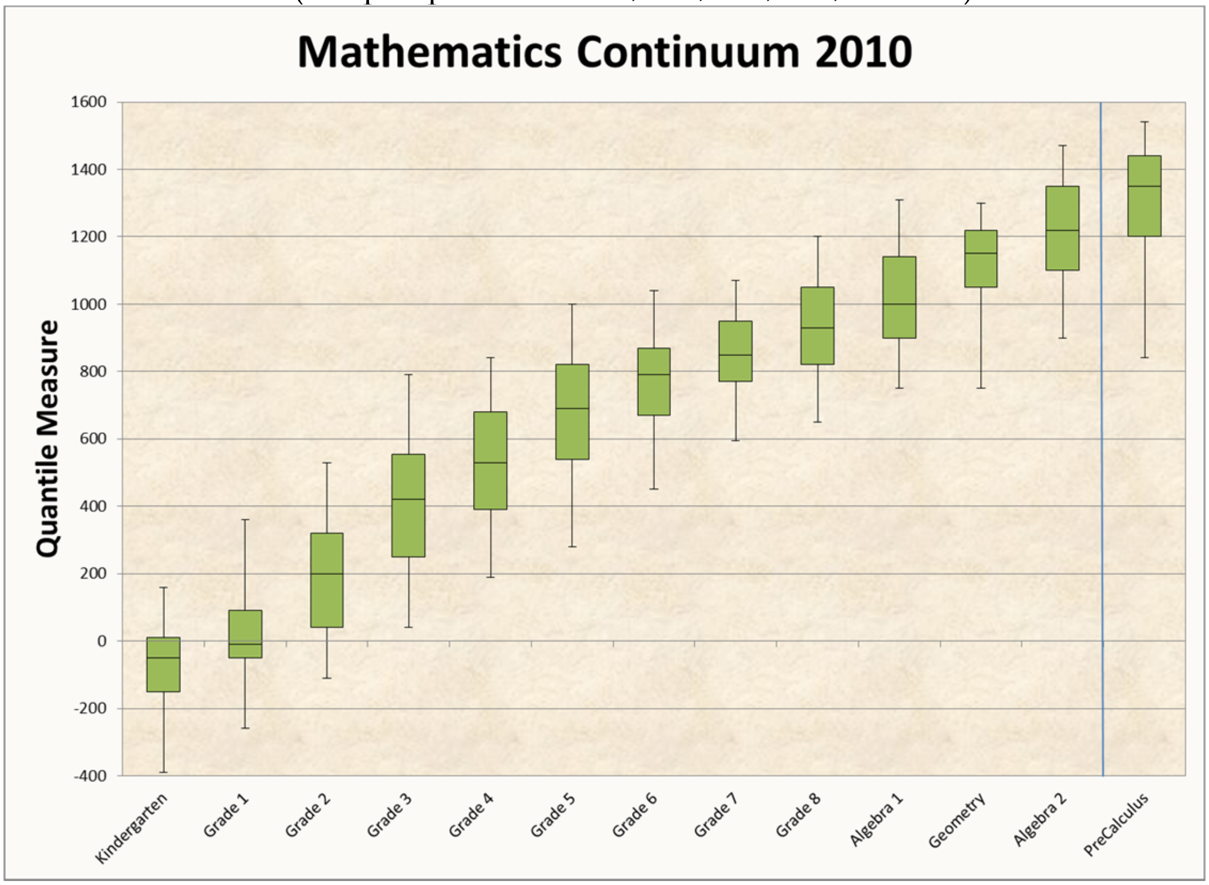
Quantile Framework and the CCSS. There is increasing recognition of the importance of bridging the gap that exists between K-12 and higher education and other postsecondary endeavors. Many state and policy leaders have formed task forces and policy committees such as P-20 councils. The Common Core State Standards (CCSS) for Mathematics were designed to enable all students to become college and career ready by the end of high school while acknowledging that students are on many different pathways to this goal: "One of the hallmarks of the Common Core State Standards for Mathematics is the specification of content that all students must study in order to be college and career ready. This 'college and career ready line' is a minimum for all students" (NGA Center & CCSSO, 2010b, p. 4). The CCSS for Mathematics suggest that "college and career ready" means completing a sequence that covers Algebra I, Geometry, and Algebra II (or equivalently, Integrated mathematics 1, 2 and 3) during the middle school and high school years; and, leads to a student's promotion into more advanced mathematics by their senior year. This has led some policy makers to generally equate the successful completion of Algebra II as a working definition of college and career ready. Exactly how and when this content must be covered is left to the states to designate in their implementations of the CCSS for Mathematics throughout K-12 (NGA Center & CCSSO, 2010a, p. 84).

The *mathematical demand* of a mathematical textbook (in Quantile measures) quantitatively defines the level of mathematical achievement that a student needs in order to be ready for instruction on the mathematical content of the textbook. Assigning QSC(s) and Quantile measures to a textbook is done through a calibration process.

Textbooks are analyzed at the lesson level and the calibrations are completed by subject matter experts (SMEs) experienced with the Quantile Framework and with the mathematics taught in mathematics classrooms. The intent of the calibration process is to determine the mathematical demand presented in the materials. Textbooks contain a variety of activities and lessons. In addition, some textbook lessons may include a variety of skills. Only one Quantile measure is calculated per lesson and is obtained through analyzing the Quantile measures of the QSCs that have been mapped to the lesson. This Quantile measure represents the composite task demand of the lesson.

MetaMetrics has calibrated more than 41,000 instructional materials (e.g., textbook lessons, instructional resources) across the K-12 mathematics curriculum (Smith and Turner, 2012). *Figure 25* shows the continuum of calibrated textbook lessons from Kindergarten through Pre-calculus where the median of the distribution for Pre-calculus is 1350Q. The range between the first quartile and the median of the first three chapters of Pre-calculus textbooks is from 1200Q to 1350Q. This range describes an initial estimate of the mathematical achievement level needed to be ready for mathematical instruction corresponding to the "college and career readiness" standard in the Common Core State Standards for Mathematics.

Figure 25. A continuum of mathematical demand for Kindergarten through Pre-calculus textbooks (box plot percentiles: 5th, 25th, 50th, 75th, and 95th).



This information describing college and career readiness in mathematics can be used to interpret the NC READY EOG Mathematics/EOC Algebra I/Integrated I performance standards. For each grade the "proficient" (Level 3) range of Quantile measures as defined by the NC READY EOG Mathematics/EOC Algebra I/Integrated I assessments is compared to the mathematical demands in the next grade/course. As can be seen in *Figure 26*, almost all students scoring at the "proficient" level should be prepared for the mathematical demands of the next grade/course. The Algebra I/Integrated I students at the proficient level are less ready for the next course work.

Figure 26. NC READY EOG Mathematics/EOC Algebra I/Integrated I "proficient" ranges (expressed as Quantile measures) compared with the mathematical demands of the next grade/course, by grade or course.

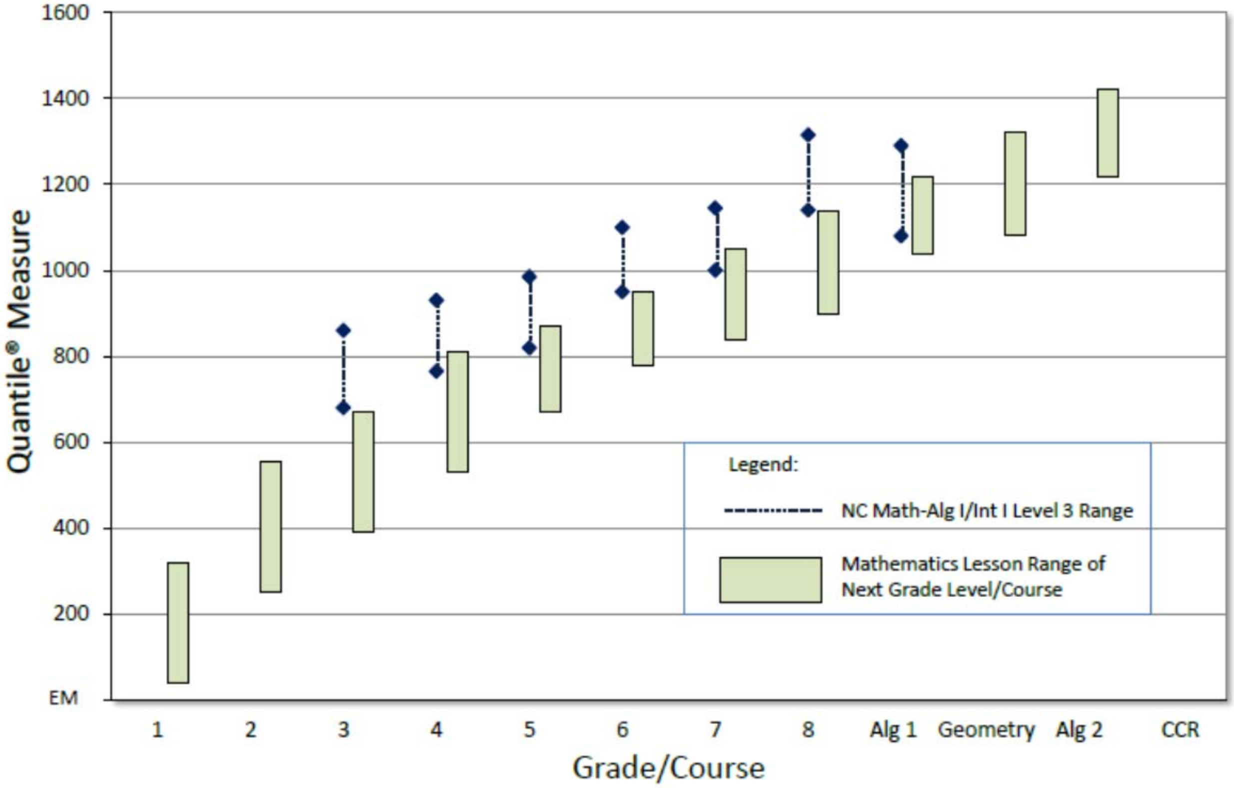
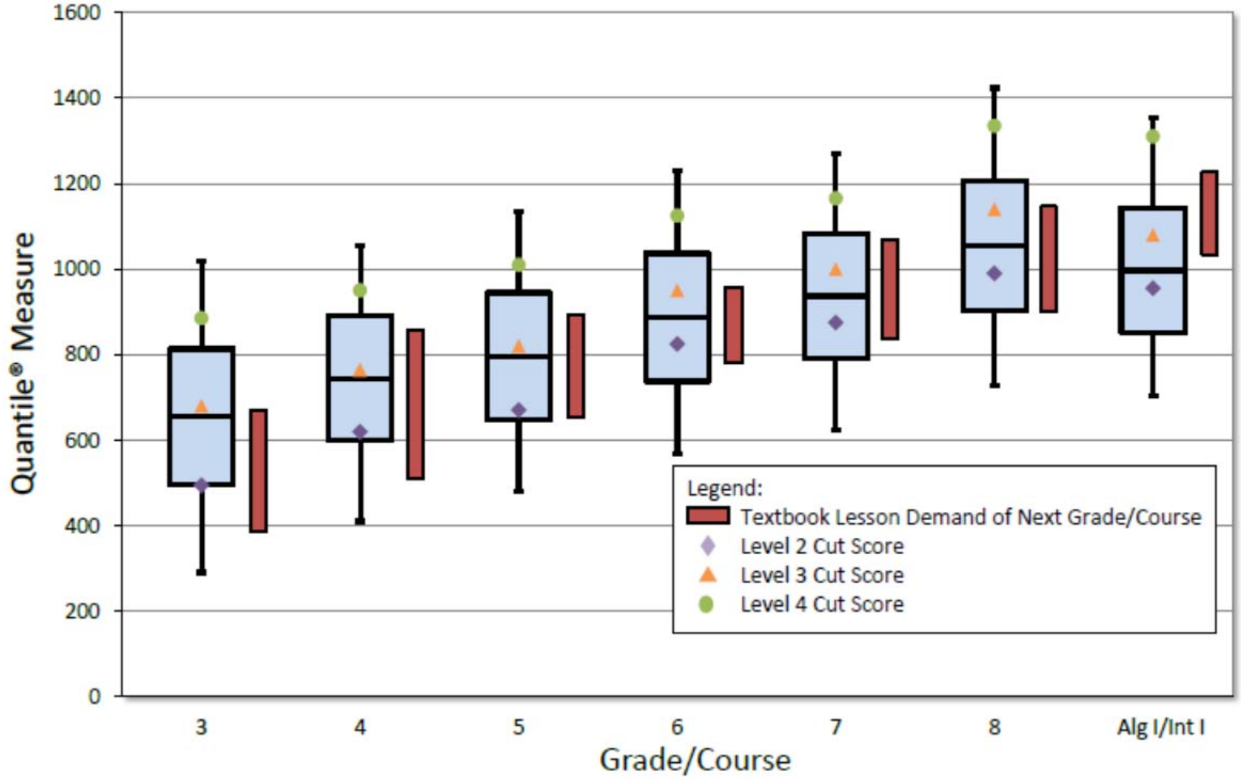


Figure 27 shows that the spring 2013 student performance on the NC READY EOG Mathematics/EOC Algebra I/Integrated I assessments at each grade/course level is "on track" for college and career readiness in Grades 3 through 8. In comparing the performance of students in Algebra I/Integrated I, some students will need encouragement with supplemental materials at the next course. Students can be matched with mathematics materials that are at or above the recommendations in the Common Core State Standards for each grade/course.

Figure 27. NC READY EOG Mathematics/EOC Algebra I/Integrated I 2012-2013 student performance expressed as Quantile measures.



In 2009, MetaMetrics and the North Carolina Department of Public Instruction conducted a study to relink the NCEOG/EOC Mathematics Tests with the Quantile scale (MetaMetrics, 2010). The minimum score considered "proficient" (Level 3) at each grade level on the NCEOG/EOC Mathematics is presented in *Table 36*. In 2013, NCDPI transitioned their assessment program to the NC READY EOG Mathematics/EOC Algebra I/Integrated I assessment to align with the Common Core State Standards in Mathematics and to describe student mathematics performance in relation to college and career readiness. One outcome of this change was to set the performance standards for NC READY EOG Mathematics/EOC Algebra I/Integrated I at a higher level. For comparison purposes, the minimum "proficient" score for the NC READY EOG Mathematics/EOC Algebra I/Integrated I assessment is also repeated from *Table 35*. The Quantile scale can be used as an external "yardstick" to evaluate this change in the mathematical demand on the North Carolina Mathematics assessments. The information in *Table 36* shows that the NC READY EOG/EOC Mathematics standards are demanding more of students in terms of mathematical ability in 2013.

Table 36. Minimum "Level 3" Quantile measure on NCEOG/EOC Mathematics (2009) and NC READY EOG Mathematics/EOC Algebra I/Integrated I (2013).

Grade	"Proficient" Level 3 Cut Score (2009)	"Proficient" Level 3 Cut Score (2013)
3	515Q	680Q
4	645Q	765Q
5	775Q	820Q
6	795Q	950Q
7	860Q	1000Q
8	900Q	1140Q
Alg I/Int I	1020Q	1080Q

Conclusions, Caveats, and Recommendations

Forging a link between scales is a way to add value to one scale without having to administer an additional test. Value can be in the form of any or all of the following:

- increased *interpretability* (e.g., "Based on this test score, what mathematical skills and concepts does my child actually know?"),
- increased *diagnostic capability* (e.g., "Based on this test score, what are the student's weaknesses?"), or
- increased *instructional use* (e.g., "Based on these test scores, I need to modify my instruction to include these skills.").

The link that has been established between the NC READY EOG Mathematics/EOC Algebra I/Integrated I assessments and the Quantile Framework permits students to be matched with resources and materials that provide an appropriate level of challenge while avoiding frustration. The result of this purposeful match may be that students will be less fearful of mathematics, and, thereby become better mathematical thinkers. The real power of the Quantile Framework is in examining the growth in mathematical achievement of students—wherever the student may be in the development of his or her mathematical skills and concepts. Students can be matched with resources and materials for which they are forecasted to experience 50% understanding, therefore, they are ready for instruction on the topic. As a student's mathematical achievement grows, he or she can be matched with more demanding skills and concepts. And, as the skills and concepts become more demanding, then the student grows.

The development of the link between the scores on the NC READY EOG Mathematics/EOC Algebra I/Integrated I assessments and the Quantile scale has been described and evaluated in this study. There are many factors that can affect the linking process. In this study two of the factors include:

- sample characteristics (e.g., gender, ethnicity), and
- relationship of sample distribution characteristics to the distribution characteristics of the state.

Conventions for Reporting. Quantile measures are reported as a number followed by a capital "Q" for "Quantile." There is no space between the measure and the "Q" and measures of 1,000 or greater are reported without a comma (e.g., 1050Q). All Quantile person measures should be rounded to the nearest 5Q to avoid over interpretation of the measures. As with any test score, uncertainty in the form of measurement error is present.

Next Steps. To utilize the results from this study, Quantile measures need to be incorporated into the NC READY EOG Mathematics/EOC Algebra I/Integrated I assessment results processing and interpretation frameworks. Suggested resources need to be developed for ranges of students. Care must be taken to ensure that the resources and materials on the lists are also developmentally appropriate for the students. The Quantile measure is one factor related to understanding and is a good starting point in the selection process of materials and resources for a specific student. Other factors such as student developmental level, motivation, and interest; amount of background knowledge possessed by the student; and characteristics of the resources and skills also need to be considered when matching resources and instruction with a student.

In this era of student-level accountability and high-stakes assessment, differentiated instruction—the attempt "on the part of classroom teachers to meet students where they are in the learning process and move them along as quickly and as far as possible in the context of a mixed-ability classroom" (Tomlinson, 1999)—is a means for all educators to help students succeed. Differentiated instruction promotes high-level and powerful curriculum for all students, but varies the level of teacher support, task complexity, pacing, and avenues to learning based on student readiness, interest, and learning profile. One strategy for managing a differentiated classroom suggested by Tomlinson is the use of multiple resources and supplementary materials that can be identified with the aid of the Quantile Framework. Equipped with a student's Quantile measure, teachers can connect him or her to textbook lessons, worksheets, games, websites, and trade books that have appropriate Quantile measures (Smith, no date; Smith and Turner, 2012). By incorporating Quantile measures into the planning of mathematics instruction, it becomes possible to forecast with greater probability how successfully students are likely to understand the material presented to them. Teachers can provide instruction on QSCs with Quantile measures below the targeted instruction when students are not ready for that instruction by focusing on prerequisite QSCs. On the other hand, teachers can focus enrichment activities on the impending QSCs.

Two resources are available on the Quantile Framework website – Quantile Teacher Assistant and Math@Home (Smith, no date; Smith and Turner, 2012). In order to support instruction with the many resources connected with the Quantile Framework, the Quantile Teacher Assistant (QTA) was developed to simplify and gather all relevant information. When using the QTA (http://qta.quantiles.com/), teachers can identify a specific state objective and determine the knowledge base. In addition, teachers can differentiate instruction by indicating the range of Quantile measures for their students in their classrooms. Math@Home (http://mah.quantiles.com/) activities reinforce mathematical skills covered in the previous school year and lay the groundwork for what will be taught when students return to class in the fall. By incorporating fun family games into everyday activities, students can practice mathematical skills year-round and parents can feel more confident about helping their children with mathematics.

MetaMetrics, in partnership with The Council of Chief State School Officers, has begun coordinating a national, state-led summer mathematics initiative to bolster student mathematics achievement during summer break. The Summer Math Challenge is designed to raise national awareness of the summer loss epidemic (Cooper, Nye, Charlton, Lindsay, and Greathouse, 1996), share compelling research on the importance of targeted mathematics activities, and provide access to a variety of free resources to support mathematics instruction and the initiative as a whole.

The 2013 "Summer Math Challenge" was a six-week, e-mail-based initiative designed to help students on summer vacation fight "summer slide" in mathematics skills. The initiative was designed to combat summer math slide by helping students retain mathematics skills acquired during the previous school year. The initiative targeted Grades 3 through 6 by reinforcing mathematics concepts presented from Grades 2 through 5 aligned with the Common Core State Standards. Participants received targeted instructional materials for a weekly concept along with personalized e-mail activity suggestions and resources that supported each concept. Twelve SEA chiefs requested assistance in launching a 2013 Summer Math initiative in conjunction with the CCSSO Chief's Summer Reading Challenge. North Carolina promoted the Summer Math Challenge through e-mail newsletters to educators. The "Chief's Summer Math Challenge" Flyer provides an overview of the CCSSO Chief's Math Challenge and MetaMetrics' 2013 Support to SEA leaders (URL: https://d1jt5u2s0h3gkt.cloudfront.net/m/cms_page_media/135/Chief's%20Summer%20Math%20Challenge%20Overview_2.pdf).

The following is a list of suggestions that can be used to leverage a student's Quantile measure in the classroom:

- Start class with warm-up problems and activities related to the prerequisite skills from a knowledge cluster.
- Enhance major themes of mathematics by building a bank of skills at varying levels that not only support a theme but also provide a way for all students to participate in the theme successfully. For example, consider how addition progresses from single numbers to multi-digit numbers, and then moves to decimals and fractions.
- Sequence mathematical skills according to their difficulty as much as possible.
- Develop a mathematics folder that goes home with students and returns weekly for review. The folder can contain examples of practice skills within a student's range, applications of topics outside the classroom, reports of recent assessments, and a parent form to record the amount of time spent working mathematics problems at home.

- Choose skills lower in a student's Quantile range when factors make the student view mathematics as more challenging, threatening, or unfamiliar. Select skills at or above a student's range to stimulate growth, when a topic holds high interest for a student, or when additional support such as background teaching or peer tutoring is provided.
- Develop individualized lists of skills that are tailored to provide appropriately challenging and curriculum suitable for all students.

Below are some suggestions related to leveraging a student's Quantile measure at home:

- Ensure that each child gets plenty of mathematical practice, concentrating on skills within his or her Quantile range. Parents can ask their child's teacher to print a list of appropriate skills or search the mathematics skill database on the Quantile website.
- Communicate with the child's teachers about the child's mathematical needs and accomplishments. They can use the Quantile scale to describe their assessment of the child's mathematical achievement.
- When a new topic proves too challenging for a child, use activities or other materials from the Web site to help. Review the prerequisite QSCs to ensure that gaps or misconceptions are not interfering with the current topic.
- Celebrate a child's mathematical accomplishments. The Quantile Framework provides an easy way for students to track their own growth. Parents and children can set goals for mathematics—spending so much time daily working on mathematical problems, discussing situational topics such as statistics from a newspaper or discounts at the store, reading a book about a mathematical topic, trying new kinds of Web sites and games, or working a certain number of mathematics problems per week. When children reach the goal, make it an occasion!

References

American Educational Research Association, American Psychological Association, and National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

Anastasi, A. (1982). *Psychological Testing* (5th ed.). New York: MacMillan Publishing Company, Inc.

Bond, T.G. & Fox, C.M. (2001). *Applying the Rasch model: Fundamental measurement in the human sciences*. Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.

Burg, S.S. (2007). *An investigation of dimensionality across grade levels and effects on vertical linking for elementary grade mathematics achievement tests*. University of North Carolina, Chapel Hill, NC.

Camilli, G. & Shepard, L.A. (1994). *Methods for identifying biased test items*. Thousand Oaks, CA: Sage Publications, Inc.

Cochran, W.G. & Cox, G.M. (1957). *Experimental Designs*. New York: John Wiley & Sons.

Cooper, H., Nye, B., Charlton, K., Lindsay, J., & Greathouse, S. (1996). The effects of summer vacation on achievement test scores: A narrative and meta-analytic review. *Review of Educational Research, 66*(3), 227-268.

Davis, F. (1944). Fundamental factors of comprehension in reading, *Psychometrika*, *9,* 185–197.

Dorans, N.J. & Holland, P.W. (1993). DIF detection and description: Mantel-Haenszel and standardization. In P.W. Holland and H. Wainer (Eds.), *Differential Item Functioning* (pp. 35-66). Hillsdale, NJ: Lawrence Erlbaum.

Emenogu, B.C. & Childs, R.A. (2005). Curriculum, translation, and differential functioning of measurement and geometry items. *Canadian Journal of Education*, *28*(1 & 2), 128-146.

Geary, D.C., & Hamson, C.O. (2002). Improving the mathematics and science of achievement of American children: Psychology's role. *American Psychological Association Online.* Retrieved May 23, 2002, from http://www.apa.org/ed/geary.html.

Goldstein, H. (2003). International comparisons of student attainment: Some issues arising from the PISA study. Retrieved December 14, 2004, from www.ioe.ac.uk/hgpersonal.

Gorsuch, R.L. (1983). *Factor analysis*. Mahwah, NJ: Lawrence Erlbaum Associates Inc.

Haladyna, T.M. (1994). *Developing and validating multiple-choice test items*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Hambleton, R.K. & Swaminathan, H. (1985). *Item response theory: Principles and applications*. Kluwer-Nijhoff Publishing.

Hambleton, R.K., Swaminathan, H., & Rogers, H.J. (1991). *Fundamentals of item response theory*. Newbury Park, CA: Sage Publications, Inc.

Hattie, J. (1985). Methodology review: Assessing unidimensionality of tests and items. *Applied Psychological Bulletin*. *9*, 139-164.

Hennings, S. S. & Simpson, M.A. (2012, October). *Quantile® Linking Test and a state assessment: How many dimensions?* Presentation at the 2012 semi-annual meeting of MetaMetrics' Technical Advisory Committee. Durham, NC: MetaMetrics, Inc.

Holland, P.W. & Wainer, H. (1993). *Differential item functioning*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Hudnutt, B. (2012). *Weaving mathematical connections from counting to calculus: Knowledge clusters and The Quantile® Framework for Mathematics*. Durham, NC: MetaMetrics, Inc.

Jaeger, R. M. (1973). The national test equating study in reading: The anchor test study. *Measurement in Education*, *4*, 1–8.

Kirsch, I. S., Jungeblut, A., Jenkins, L., & Kolstad, A. (1993). *Adult literacy in America: A first look at the results of the National Adult Literacy Survey*. Washington, DC: National Center for Education Statistics, U.S. Department of Education.

Kirsch, I. S., Jungeblut, A., & Mosenthal, P.B. (1994). *Moving toward the measurement of adult literacy*. Paper presented at the March NCES Meeting, Washington, DC.

Kolen, M.J. & Brennan, R.L. (2004). *Test equating, scaling, and linking: Methods and practices*. (2nd ed.) New York: Springer Science + Business Media, LLC.

Linacre, J.M. (2011). WINSTEPS (Version 3.73) [Computer Program]. Chicago: Author.

Lord, F.M. (1980). *Applications of item response theory to practical testing problems*. New York: Erlbaum Associates.

Loret, P. G., Seder, A., Bianchini, J. C., & Vale, C. A. (1974). *Anchor test study final report: Project report* (Vols. 1–30). Berkeley, CA: Educational Testing Service. (ERIC Document Nos. ED 092 601-ED 092 631).

Messick, S. (1993). Validity. In R.L. Linn (Ed.), *Educational Measurement* (3rd ed., pp. 13–104). New York: American Council on Education and Macmillan Publishing Company.

MetaMetrics, Inc. (October 29, 1998). *Linking.* Unpublished manuscript. Durham, NC: Author.

MetaMetrics, Inc. (2010). *Linking the North Carolina EOG/EOC Tests in Mathematics with the Quantile Framework*. Durham, NC: Author.

MetaMetrics, Inc. (2005). *PASeries* mathematics technical manual. Durham, NC: Author.

MetaMetrics, Inc. (2008). *The Quantile® Framework for Mathematics norms* [Unpublished normative data]. Durham, NC: Author.

National Council of Teachers of Mathematics (NCTM). (2000). *Principles and standards for school mathematics.* Reston, VA: Author.

National Governors Association Center for Best Practices (NGA Center) & the Council of Chief State School Officers (CCSSO). (2010a). *Common core state standards for mathematics.* Retrieved from http://www.corestandards.org/assets/CCSSI_Math%20Standards.pdf.

National Governors Association Center for Best Practices (NGA Center) & the Council of Chief State School Officers (CCSSO). (2010b). *Common core state standards for mathematics: Appendix A.* Retrieved from http://www.corestandards.org/assets/CCSSI_Mathematics_Appendix_A.pdf.

National Research Council. (1999). *Uncommon measures: Equivalence and linkage among educational tests*. Washington, D.C.: National Academy Press.

National Research Council. (2001). *Adding it up: Helping children learn mathematics*. J. Kilpatrick, J. Swafford, and B. Findell (Eds.). Mathematics Leaning Study Committee, Center for Education, Division of Behavioral and Social Sciences and Education. Washington, DC: National Academy Press.

National Research Council. (2002). *Helping children learn mathematics*. Mathematics Learning Study Committee, J. Kilpatrick and J. Swafford, Editors. Center for Education, Division of Behavioral and Social Sciences and Education. Washington, DC: National Academy Press.

North Carolina Department of Public Education. (2013a). *Achievement level descriptors for North Carolina end-of-course tests.* Retrieved on November 6, 2013 from http://sbepolicy.dpi.state.nc.us/Policies/GCS-C-036.asp?Acr=GCS&Cat=C&Pol=036

North Carolina Department of Public Education. (2013b). *Achievement level descriptors for North Carolina end-of-grade tests.* Retrieved on November 6, 2013 from http://sbepolicy.dpi.state.nc.us/Policies/GCS-C-033.asp?Acr=GCS&Cat=C&Pol=033

North Carolina Department of Public Instruction. (2013c). *Common core state standards (CCSS) for mathematics: North Carolina assessment specifications summary, READY EOG assessments, Grades 3-8, READY EOC algebra I/integrated I assessments.* Retrieved on October 31, 2013 from http://www.dpi.state.nc.us/docs/acre/assessment/math.pdf

North Carolina Department of Public Instruction. (2013d). *End-of-course assessment: Number of items and total test time (minutes)*. Retrieved on December 6, 2013 from http://www.ncpublicschools.org/docs/accountability/testing/eoc/eocadmininfo12.pdf

North Carolina Department of Public Instruction. (2013e). *End-of-grade assessment: Number of items and total test time (minutes)*. Retrieved on December 6, 2013 from http://www.ncpublicschools.org/docs/accountability/testing/eog/eogadmininfo13.pdf

North Carolina Department of Public Instruction. (2013f). *North Carolina READY end-of-course assessments*. Retrieved on October 24, 2013 from http://www.ncpublicschools.org/docs/accountability/policyoperations/assessbriefs/assessbriefeoc13.pdf

North Carolina Department of Public Instruction. (2013g). *North Carolina READY end-of-grade assessments*. Retrieved on October 24, 2013 from http://www.ncpublicschools.org/docs/accountability/policyoperations/assessbriefs/assessbriefeog13.pdf

Petersen, N.S., Kolen, M.J., & Hoover, H.D. (1989). Scaling, Norming, and Equating. In R.L. Linn (Ed.), *Educational Measurement* (3rd ed. pp. 221-262). New York: American Council on Education and Macmillan Publishing Company.

Poznanski, J.B. (1990). A meta-analytic approach to the estimation of item difficulties. Unpublished doctoral dissertation, Duke University, Durham, NC.

Price, L. R., Lurie, A., & Wilkins, C. (2001). EQUIPERCENT: A SAS program for calculating equivalent scores using the equipercentile method. *Applied Psychological Measurement, 25*, 332-341.

Rasch, G. (1980). *Probabilistic models for some intelligence and attachment tests*. Chicago: The University of Chicago Press. (First published in 1960).

Reckase, M. D. (1979). Unifactor latent trait models applied to multifactor tests: Results and implications. *Journal of Educational Statistics, 4*, 207-230.

Reder, S. (1998). Dimensionality and construct validity of the NALS assessment. In M.C. Smith (Ed.), *Literacy for the twenty-first century: Research, policy, practices and the National Adult Literacy Survey*. Westport, CT: Preager Publishing.

Rentz, R. R., & Bashaw, W. L. (1975). *Equating reading tests with the Rasch model* (Vol. 1-2). Athens, GA: University of Georgia, Educational Research Laboratory.

Rentz, R. R., & Bashaw, W. L. (1977). The National Reference Scale for Reading: An application of the Rasch model. *Journal of Educational Measurement*, *14,* 161–179.

Roussos, L., Schnipke, D., & Pashley, P. (1999). A generalized formula for the Mantel-Haenszel differential item functioning parameter. *Journal of Behavioral and Educational Statistics*, 24, 293-322.

Salganik, L. H., & Tal, J. (1989). A Review and Reanalysis of the ETS/NAEP Young *Adult Literacy Survey*. Washington, DC: Pelavin Associates.

Salvia, J. & Ysseldyke, J.E. (1998). *Assessment* (7th ed.). Boston: Houghton Mifflin Company.

SAS Institute, Inc. (1985). The FREQ procedure. In *SAS Users Guide: Statistics, Version 5 Edition*. Cary, NC: Author.

Sitter, R.R. (1992). Comparing three bootstrap methods for survey data. *The Canadian Journal of Statistics*, *20*(2), 135-154.

Smith, M. (2010). *The need for differentiating mathematics instruction* (MetaMetrics Policy Paper). Durham: MetaMetrics, Inc.

Smith, M., & Turner, J. (no date). *Supporting differentiated math instruction in a Common Core world*. Durham: MetaMetrics, Inc.

Smith, M., & Turner, J. (2012, February). *A mathematics problem: How to help students achieve success in mathematics through college and beyond* (MetaMetrics Policy Brief). Durham, NC: MetaMetrics, Inc.

Starr, L. (2002). Math wars! *Education World*. Retrieved January 27, 2003, from http://www.eduationworld.com/a_curr/curr071.shmtl.

Stenner, A.J. (1990). Objectivity: Specific and general. *Rasch Measurement Transactions*, *4*, 111.

Thurstone, L. L. (1946). Note on a reanalysis of Davis' Reading Tests. *Psychometrika, 11*(2), 185.

Tomlinson, C.A. (1999). *The differentiated classroom*. Alexandria, VA: Association for Supervision and Curriculum Development.

Wright, B.D. & Linacre, J.M. (1994, August). *The Rasch model as a foundation for the Quantile Framework*. Unpublished manuscript.

Wright, B.D. & Stone, M.H. (1979). *Best Test Design*. Chicago: MESA Press.

Zwick, R. (1987). Assessing the Dimensionality of the NAEP Reading Data. *Journal of Educational Measurement, 24,* 293–308.

Appendix A

THE QUANTILE® FRAMEWORK FOR MATHEMATICS MAP

Linking assessment with mathematics instruction

Imagine empowering and accelerating students' learning in mathematics by better differentiating instruction and monitoring growth in student ability. With the Quantile Framework, educators can help achieve this goal by identifying level-appropriate mathematical tasks for students and track their progress!

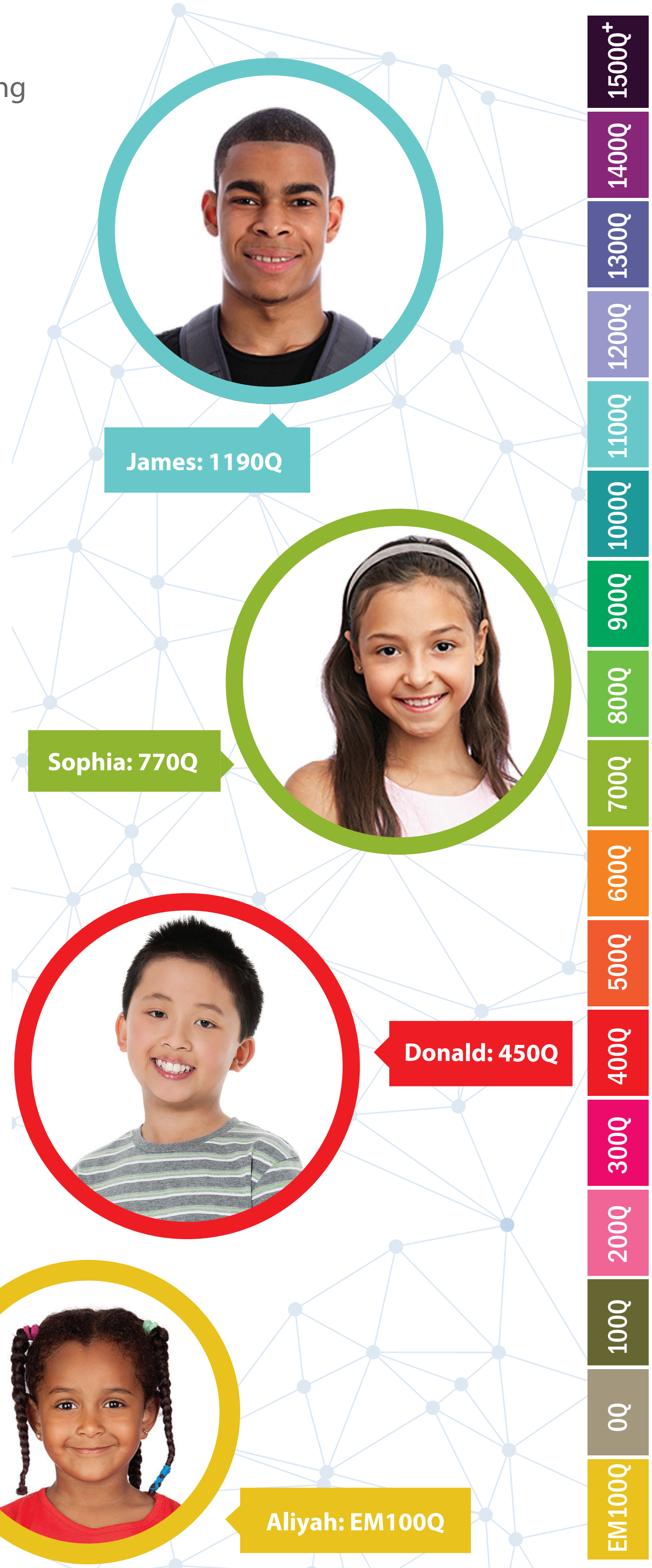
HOW IT WORKS

The Quantile Framework for Mathematics is a unique measurement system that uses a common scale and metric to assess a student's mathematical achievement level and the difficulty of specific skills and concepts. The Quantile Framework describes a student's ability to solve mathematical problems and the demand of the skills and concepts typically taught in kindergarten mathematics through Algebra II, Geometry, Trigonometry and Precalculus. The Quantile Map provides educators with a sampling of primary mathematical skills and concepts from over 500 Quantile Skills and Concepts (QSCs) throughout the Quantile scale. This sampling of QSCs ranges from EM (Emerging Mathematician) for early, foundational mathematical skills and concepts to 1500Q for more advanced skills and concepts. As the difficulty, or demand of the skill increases, so does the Quantile measure.

HOW TO USE IT

With the Quantile Framework, educators can explore the interconnectedness of mathematical skills and concepts and identify those elements that are critical for progressing student learning. Educators are better able to inform their instruction on how to best teach a skill or concept by pinpointing which skills build upon each other. The skill mapping of mathematical concepts enables educators to build an instructional path that best fits their students' unique abilites. Both students and QSCs receive a Quantile measure. Numerous tests report Quantile student measures including many state end-of-year assessments, national norm-referenced assessments and math programs. On the QSC side, more than 580 textbooks, 64,000 lessons and 3,100 downloadable resources have received Quantile measures.

Quantile measures provide educators with the information they need to identify gaps in mathematical knowledge, as well as serve as a guide for progressing to more advanced topics. Every QSC is part of a knowledge cluster that shows relationships and connections between mathematical skills and offers their relative difficulty among different skills. Both the *prerequisite* and *impending* skills are elements of knowledge clusters and serve as building blocks that support students' success. Educators can advance student learning by using prerequisite and impending skills to build mathematical knowledge and understanding. Prerequisite skills help educators see the pieces of the puzzle that make up a skill or concept, showing what needs to be understood first. Impending skills are skills and concepts that build upon a focus skill and allow educators to see a trajectory of knowledge across grades and content strands.



For more information, free resources, and to search the Math Skills Database, visit Quantiles.com.

High School Example
James

Heritage High School | Geometry Course

Quantile Measure: 1190Q

James is exploring theorems about lines and angles in his Geometry class. In his current learning path, the focus skill being taught is *use properties, definitions, and theorems of angles and lines to solve problems related to adjacent, vertical, complementary, supplementary, and linear pairs of angles.* This focus skill is part of a knowledge cluster that contains prerequisite and impending skills. Working with prerequisite skills can help students struggling to learn and impending skills can help students progress to the next level of learning.

Since James' Quantile measure is within the range of the focus skill being taught (his Quantile measure +/- 50Q), James will be ready for this type of instruction. With his mathematical ability being at the same level as the focus skill, learning will be optimal. Once James is performing well with the focus skill, he will be better prepared to learn the impending skills connected with this focus skill.

1250Q
IMPENDING SKILL
Use definitions and theorems of angles formed when a transversal intersects parallel lines.

1220Q
IMPENDING SKILL
Use properties, definitions, and theorems of polygons to solve problems related to the interior and exterior angles of a convex polygon.

1160Q
FOCUS SKILL
Use properties, definitions, and theorems of angles and lines to solve problems related to adjacent, vertical, complementary, supplementary, and linear pairs of angles.
CCSS G.CO.9

1010Q
PREREQUISITE SKILL
Define and identify alternate interior, alternate exterior, corresponding, adjacent and vertical angles.

1020Q
PREREQUISITE SKILL
Define and identify complementary and supplementary angles.

800Q
PREREQUISITE SKILL
Write a linear equation or inequality to represent a given number or word problem; solve.

ALGEBRA & ALGEBRAIC THINKING | NUMBER SENSE | NUMERICAL OPERATIONS | MEASUREMENT | GEOMETRY | DATA ANALYSIS, STATISTICS & PROBABILITY

THE **QUANTILE® FRAMEWORK** FOR MATHEMATICS MAP

Linking assessment with mathematics instruction

Middle School Example
Sophia

Heritage Middle School | Grade 6

Quantile Measure: 770Q

Sophia is using variables to represent mathematical expressions in her math class. In her current learning path, the focus skill being taught is *translate between models or verbal phrases and algebraic expressions*. This focus skill is part of a knowledge cluster that contains prerequisite and impending skills. Working with prerequisite skills can help students struggling to learn and impending skills can help students progress to the next level of learning.

Since Sophia's Quantile measure is within the range of the focus skill being taught (her Quantile measure +/- 50Q), Sophia will be ready for this type of instruction. With her mathematical ability being at the same level as the focus skill, learning will be optimal. Once Sophia is performing well with the focus skill, she will be better prepared to learn the impending skills connected with this focus skill.

810Q
IMPENDING SKILL
Write an equation to describe the algebraic relationship between two defined variables in number and word problems, including recognizing which variable is dependent.

800Q
IMPENDING SKILL
Identify parts of a numerical or algebraic expression.

800Q
IMPENDING SKILL
Write a linear equation or inequality to represent a given number or word problem; solve.

750Q
FOCUS SKILL
Translate between models or verbal phrases and algebraic expressions.
CCSS 6.EE.6

620Q
PREREQUISITE SKILL
Translate between models or verbal phrases and numerical expressions.

430Q
PREREQUISITE SKILL
Describe the meaning of an unknown in the context of a word problem.

MetaMetrics.

For more information, visit Quantiles.com.

ALGEBRA & ALGEBRAIC THINKING | NUMBER SENSE | NUMERICAL OPERATIONS | MEASUREMENT | GEOMETRY | DATA ANALYSIS, STATISTICS & PROBABILITY

THE **QUANTILE® FRAMEWORK FOR MATHEMATICS** MAP

Linking assessment with mathematics instruction

Late Elementary Example
Donald

Heritage Elementary School | Grade 4

Student Quantile Measure: 450Q

Donald is learning about line graphs with very large data values. In his current learning path, the focus skill being taught is *organize, display, and interpret information in graphs containing scales that represent multiple units.* This focus skill is part of a knowledge cluster that contains prerequisite and impending skills. Working with prerequisite skills can help students struggling to learn and impending skills can help students progress to the next level of learning.

Since Donald's Quantile measure is within the range of the focus skill being taught (his Quantile measure +/- 50Q), Donald will be ready for this type of instruction. With his mathematical ability being at the same level as the focus skill, learning will be optimal. Once Donald is performing well with the focus skill, he will be better prepared to learn the impending skills connected with this focus skill.

800Q ▲
IMPENDING SKILL
Identify and use appropriate scales and intervals in graphs and data displays.

480Q ▲
FOCUS SKILL
Organize, display, and interpret information in graphs containing scales that represent multiple units.
CCSS 3.MD.3

480Q ▲
IMPENDING SKILL
Organize, display, and interpret information in bar graphs.

470Q ▲
IMPENDING SKILL
Organize, display, and interpret information in line graphs.

200Q ▲
PREREQUISITE SKILL
Organize, display, and interpret information in line plots and tally charts.

110Q ★
PREREQUISITE SKILL
Skip count by 2s, 5s and 10s beginning at any number.

90Q ★
PREREQUISITE SKILL
Skip count by 3s, 4s, 6s, 7s, 8s, and 9s.

EM10Q ▲
PREREQUISITE SKILL
Organize, display, and interpret information in picture graphs and bar graphs using grids.

◆ ALGEBRA & ALGEBRAIC THINKING | ★ NUMBER SENSE | ■ NUMERICAL OPERATIONS | ● MEASUREMENT | ⏢ GEOMETRY | ▲ DATA ANALYSIS, STATISTICS & PROBABILITY

Early Elementary Example
Aliyah

Heritage Elementary School | Kindergarten

Quantile Measure: EM100Q

Aliyah is exploring unknown-addend problems in her class. In her current learning path, the focus skill being taught is *know and use related addition and subtraction facts.* This focus skill is part of a knowledge cluster that contains prerequisite and impending skills. Working with prerequisite skills can help students struggling to learn and impending skills can help students progress to the next level of learning.

Since Aliyah's Quantile measure is within the range of the focus skill being taught (her Quantile measure +/- 50Q), Aliyah will be ready for this type of instruction. With her mathematical ability being at the same level as the focus skill, learning will be optimal. Once Aliyah is performing well with the focus skill, she will be better prepared to learn the impending skills connected with this focus skill.

EM25Q ■
IMPENDING SKILL

Model the concept of subtraction using numbers less than or equal to 10.

EM80Q ■
FOCUS SKILL

Know and use related addition and subtraction facts.

CCSS 1.0A.4

EM110Q ◆
PREREQUISITE SKILL

Identify missing addends for addition facts.

EM260Q ■
PREREQUISITE SKILL

Model the concept of addition for sums to 10.

MetaMetrics.

For more information, visit Quantiles.com.

◆ ALGEBRA & ALGEBRAIC THINKING | ★ NUMBER SENSE | ■ NUMERICAL OPERATIONS | ● MEASUREMENT | GEOMETRY | DATA ANALYSIS, STATISTICS & PROBABILITY