TECHNICAL MANUAL

Quantile® Framework for Mathematics

Development and Validity Evidence

Bringing Meaning to Measurement

800 Taylor Street,
Suite 102
Durham, NC 27701

MetaMetricsInc.com

January 2022

The Quantile Framework® for Mathematics
Development and Validity Evidence

Technical Manual

MetaMetrics
800 Taylor Street, Suite 102
Durham, North Carolina 27701
https://www.quantiles.com/

January 2022

Table of Contents

List of Tables

List of Figures

The Quantile Framework for Mathematics

The Quantile Framework is a scale that describes a student's mathematical achievement. Similar to how degrees on a thermometer measure temperature, the Quantile Framework uses a common metric—the Quantile—to scientifically measure a student's ability to reason mathematically, monitor a student's readiness for mathematics instruction, and locate a student on its taxonomy of mathematical skills, concepts, and applications.

The Quantile Framework uses this common metric to measure many different aspects of education in mathematics. The same metric can be applied to measure the materials used in instruction, to calibrate the assessments used to monitor instruction, and to interpret the results that are derived from the assessments. The result is an anchor to which resources, concepts, skills, and assessments can be connected.

There are dozens of mathematics tests that measure a common construct and report results in proprietary, nonexchangeable metrics. Not only are all of the tests using different units of measurement, but all use different scales on which to make measurements. Consequently, it is difficult to connect the test results with materials used in the classroom. The alignment of materials and linking of assessments with the Quantile Framework provides educators, parents, and students a common vocabulary to communicate and improve mathematics learning. The benefits of having a common metric include being able to:

- Develop individual multiyear growth trajectories that denote a developmental continuum from the early elementary level to Statistics and Calculus. The Quantile scale is vertically constructed, so the meaning of a Quantile measure is the same regardless of grade level.
- Monitor and report student growth that meets the needs of state accountability systems.
- Help classroom teachers make day-to-day instructional decisions that foster acceleration and growth toward algebra readiness and through the next several years of secondary mathematics.
- Build links between mathematics curricula and major mathematics tests.
- 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.

In developing the Quantile Framework, the following tasks were undertaken:

- The development of a structure of mathematics that spans the developmental continuum from first-grade content through Algebra I, Geometry, Algebra II, Statistics and Calculus, or Math 1 through Math 3 content.
- The production of a bank of items that have been field tested.
- The development of the Quantile scale (multiplier and anchor point) based on the calibrations of the field-test items.

- The validation of the measurement of mathematics ability as defined by the Quantile Framework.

Each of these tasks is described in the subsequent sections.

Structure of the Quantile Framework for Mathematics

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 specific domain of knowledge and skills.

During the past 10 years, one of the key shifts in mathematics is the call for greater rigor in mathematics instruction. Rigor is defined as the pursuit of "conceptual understanding, procedural skills and fluency, and application with equal intensity" (National Governor's Association and Council of Chief State School Officers, 2014).

- *Conceptual understanding*. The standards call for conceptual understanding of key concepts, such as place value and ratios. Students must be able to access concepts from a number of perspectives in order to see mathematics as more than a set of mnemonics or discrete procedures.

- *Procedural skills and fluency*. The standards call for speed and accuracy in calculation. Students must practice core functions, such as single-digit multiplication, in order to have access to more complex concepts and procedures. Fluency must be addressed in the classroom or through supporting materials.

- *Application*. The standards call for students to use mathematics in situations that require mathematical knowledge. Correctly applying mathematical knowledge depends on students having a solid conceptual understanding and procedural fluency.

When developing the Quantile Framework, MetaMetrics recognized that in order to adequately address the scope and complexity of mathematics, multiple proficiencies and competencies must be assessed. The Quantile Framework is an effort to recognize and define a developmental context of mathematics instruction. This notion is consistent with the National Council of Teachers of Mathematics' (NCTM) conclusions about the importance of school mathematics for college and career readiness (Larson, 2011).

Mathematical strands. A strand is a major subdivision of mathematical content. Strands describe what students should know and be able to do. The National Council of Teachers of Mathematics' (NCTM) publication *Principles and Standards for School Mathematics* (2000, hereafter NCTM Standards) outlined ten standards—five content standards and five process standards. These content standards are Number and Operations, Algebra, Geometry, Measurement, and Data Analysis and Probability. The process standards are Communications, Connections, Problem Solving, Reasoning, and Representation.

College- and career-readiness standards for mathematics identify critical areas of mathematics that students are expected to learn each year from kindergarten through high school (National Governors Association Center for Best Practices [NGA Center] & the Council of Chief State School Officers [CCSSO], 2010a, 2010b). The critical areas in kindergarten through Grade 8 are divided into domains which differ at each grade level and include Counting and Cardinality, Operations and Algebraic Thinking, Number and Operations in Base Ten, Number and Operations-Fractions, Ratios and Proportional Relationships, The Number System, Expressions and Equations, Functions, Measurement and Data, Statistics and Probability, and Geometry. The standards for Grades 9–12 are organized by six conceptual categories: Number and Quantity, Algebra, Functions, Modeling, Geometry, and Statistics and Probability.

The six strands of the Quantile Framework bridge the Content Standards of the NCTM Standards and the domains specified in the college- and career-readiness standards for mathematics.

- *Algebra and Algebraic Thinking.* The use of symbols and variables to describe the relationships between different quantities is covered by algebra. By representing unknowns and understanding the meaning of equality, students develop the ability to use algebraic thinking to make generalizations. Algebraic representations can also allow the modeling of an evolving relationship between two or more variables.

- *Number Sense.* Students with number sense are able to understand a number as a specific amount, a product of factors, and the sum of place values in expanded form. These students have an in-depth understanding of the base-ten system and understand the different representations of numbers.

- *Numerical Operations.* Students perform operations using strategies and standard algorithms on different types of numbers but can also use estimation to simplify computation and to determine how reasonable their results are. This strand also encompasses computational fluency.

- *Measurement.* The description of the characteristics of an object using numerical attributes is covered by measurement. The strand includes using the concept of a unit to determine length, area and volume in the various systems of measurement, and the relationship between units of measurement within and between these systems.

- *Geometry.* The characteristics, properties, and comparison of shapes and structures are covered by geometry, including the composition and decomposition of shapes. Not only does geometry cover abstract shapes and concepts, but it provides a structure that can be used to observe the world.

- *Data Analysis, Statistics, and Probability.* The gathering of data and interpretation of data are included in data analysis, probability, and statistics. The ability to apply knowledge gathered using mathematical methods to draw logical conclusions is an essential skill addressed in this strand.

The Quantile Skill and Concept. Within the Quantile Framework, a Quantile Skill and Concept, or QSC, describes a specific mathematical skill or concept a student can acquire. These QSCs are arranged in an orderly progression to create a taxonomy called the Quantile scale. Examples of QSCs include:

- Know and use addition and subtraction facts to 10 and understand the meaning of equality.
- Use addition and subtraction to find unknown measures of non-overlapping angles.
- Determine the effects of changes in slope and/or intercepts on graphs and equations of lines.

During the spring of 2003, the QSCs used within the Quantile Framework were developed for Grades 1 through 8, Grade 9 (Algebra I) and Grade 10 (Geometry). The framework was extended to Algebra II and revised during the summer/fall of 2003. In the summer/fall of 2007, the content was extended to include material typically taught in Kindergarten and Grade 12 (Precalculus). And in the summer/fall of 2019 the framework was once again revised to include Statistics and Calculus.

The first step in developing a content taxonomy was to review the curricular frameworks from a variety of sources (e.g., National Council of Teachers of Mathematics [NCTM], National Assessment of Educational Progress: 2005 Pre-Publication Edition, North Carolina, California, Florida, Illinois, and Texas). 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, Algebra II, Statistics, and Calculus. Each QSC consists of a description of the content, a unique identification number, the grade at which it typically first appears, and the strand with which it is associated.

The Quantile Framework for Mathematics Map (Appendix A) presents a visual representation of the construct of mathematics ability. The map is organized by the six 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 three large, national studies (Quantile Framework field study, *PASeries* Mathematics field study described later in this document, the 2019 Quantile extension to include Statistics and Calculus), and from additional field studies as new QSCs are proposed and investigated.

Quantile Scale Development

The second step in the process of developing The Quantile Framework for 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. Each QSC developed during the design of the Quantile Framework was aligned to a strand and 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). Each item was associated with a strand and a QSC. In the development of the Quantile Framework item bank, the reading demand of the items was kept as low as possible to ensure that the items were testing mathematics achievement and not reading. Additional Statistics and Calculus items were developed and field tested in 2019 and are included in the Quantile Item Bank.

Item writing and review. 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 make 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 address these issues and identify areas to avoid when developing items. The following areas are covered: violence and crime, sources of common phobias, negative emotions such as death and family issues, offensive language, drugs/alcohol/tobacco, sex/attraction, race/ethnicity, class, gender, religion, supernatural/magic, parent/family, politics, animal cruelty and hunting, environmental issues, brand names, and junk food. These materials were developed based on material published by McGraw-Hill (Guidelines for Bias-Free Publishing, 1983) 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 approved, approved with edits, or deleted.

Field-test design and linking. The next stage in the development of the Quantile item bank was the field-testing of all of the items. First, individual test items were compiled into leveled assessments and distributed to groups of students. The data gathered from these assessments were then analyzed using a variety of statistical methods. The final result was a bank of test items appropriately placed within the Quantile scale, suitable for determining the mathematical achievement of students on this scale. Assessment forms were developed for 10 levels for the purposes of field testing. 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.

The final field tests were composed of 685 unique items. Besides the 660 items mentioned above, two sets of 12 linking items were developed to serve as below-level items for Grade 2 and above-level items for Algebra II. Two additional Algebra II items were developed to ensure coverage of all the QSCs at that level.

Linking the test levels vertically (across grades) employed a common-item test design (design in which items are used on multiple forms). In this design, multiple tests are given to nonrandom 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).

Quantile Framework field study and analysis. 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., urban; suburban; and small town, city, or rural communities). See *Table 1* for information about the sample at each grade level and the total sample. See *Table 2* for test administration forms by level.

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)

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

Students who were administered Levels 2 through 11 test forms were provided with rulers and students who were administered Levels 3 through 11 test forms were provided with protractors. For students who were administered Levels 5 through 8 and 10 and 11 test forms, formulas were provided on the back of the test booklet. Administration time was approximately 45 minutes at each level. Students who were administered a Level 2 test form had the option of having the test read aloud and marked in the test booklet if that was typical of instruction.

Field-test analyses. At the conclusion of the field test, complete data from a total of 9,678 students was analyzed. Data were deleted if test level or test form was not indicated or the answer sheet was blank. 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, test form level and ID 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	*p*-value Mean (Range)	Correct Response Point-Biserial Correlation Mean (Range)	Incorrect Responses Point-Biserial Correlation Mean (Range)
2	90	0.58 (0.12 – 0.95)	0.32 (-0.15 – 0.56)	-0.21 (-0.43 – 0.12)
3	90	0.53 (0.11 – 0.93)	0.26 (-0.08 – 0.52)	-0.22 (-0.54 – 0.02)
4	90	0.55 (0.12 – 0.92)	0.24 (-0.21 – 0.50)	-0.22 (-0.48 – 0.12)
5	90	0.54 (0.12 – 0.95)	0.28 (-0.05 – 0.50)	-0.23 (-0.45 – 0.05)
6	90	0.52 (0.04 – 0.86)	0.24 (-0.08 – 0.45)	-0.22 (-0.46 – 0.09)
7	90	0.44 (0.10 – 0.77)	0.29 (-0.12 – 0.56)	-0.21 (-0.46 – 0.25)
8	90	0.43 (0.10 – 0.81)	0.26 (-0.15 – 0.50)	-0.20 (-0.45 – 0.13)
9	90	0.40 (0.10 – 0.79)	0.21 (-0.19 – 0.52)	-0.19 (-0.53 – 0.22)
10	88	0.51 (0.01 – 0.97)	0.19 (-0.26 – 0.53)	-0.21 (-0.55 – 0.18)
11	90	0.53 (0.09 – 0.98)	0.26 (-0.09 – 0.51)	-0.22 (-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.

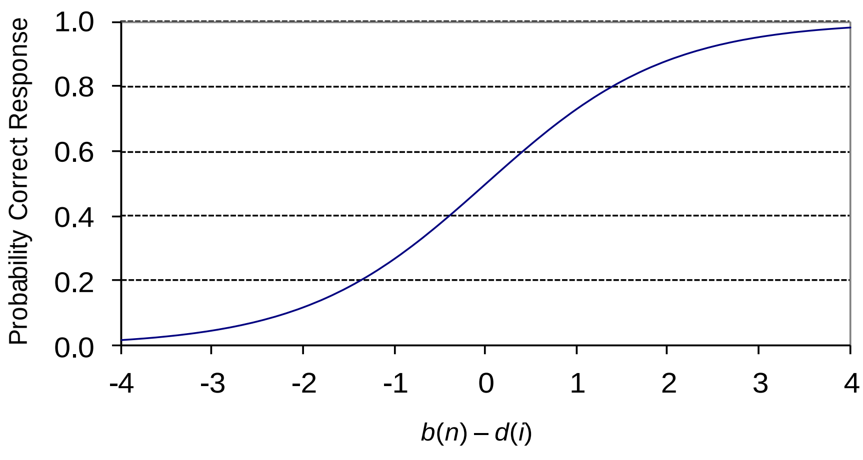
The conversion of observations into measures can be accomplished using the Rasch (1980) model, which states a requirement for the way that item calibrations and observations (count of correct items) interact in a probability model to produce measures. The Rasch IRT model expresses the probability that 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.

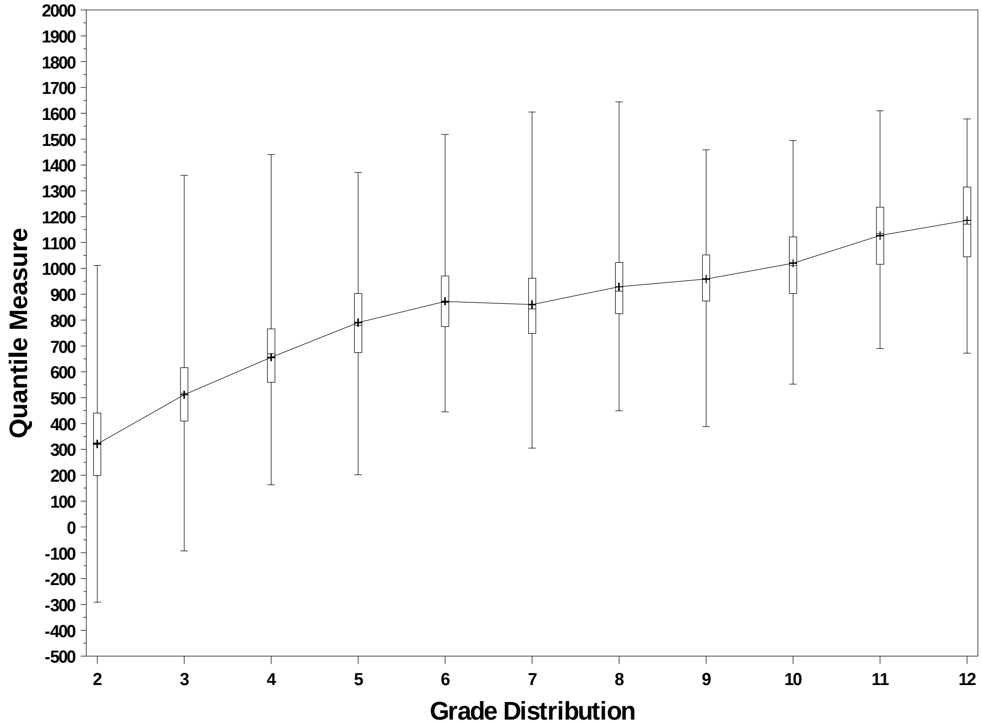
All students and items were submitted to a Winsteps analysis using a logit convergence criterion of 0.0001 and a residual convergence criterion of 0.001. Items that a student skipped were treated as missing, rather than being treated as incorrect. Only students who responded to at least 20 items were included in the analyses (22 students were omitted, 0.22%). The Quantile measure comes from multiplying the logit value by 180 and is anchored at 656Q. The multiplier and the anchor point will be discussed in a later section. *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* Mathematics described later in this document exhibit a monotonically increasing function).

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

Grade Level	*N*	Quantile measure Mean (SD)	Quantile measure Median
2	1,275	321 (189.1)	323
3	1,339	511 (157.7)	516
4	1,427	655 (157.5)	667
5	1,337	790 (167.7)	771
6	959	872 (153.0)	865
7	1,244	861 (174.2)	841
8	1,004	929 (157.6)	910
9	482	959 (152.8)	953
10	251	1020 (162.9)	1005
11	200	1127 (178.6)	1131
12	138	1186 (189.2)	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*	Logit Value Median	Quantile measure Mean (Median)
2	1,167	-2.800	289 (292)
3	1,260	-1.650	502 (499)
4	1,352	-0.780	653 (656)
5	1,289	0.000	795 (796)
6	937	0.430	881 (874)
7	1,181	0.370	878 (863)
8	955	0.810	951 (942)
9	466	1.020	983 (980)
10	244	1.400	1044 (1048)
11	191	2.070	1160 (1169)
12	134	2.295	1220 (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).

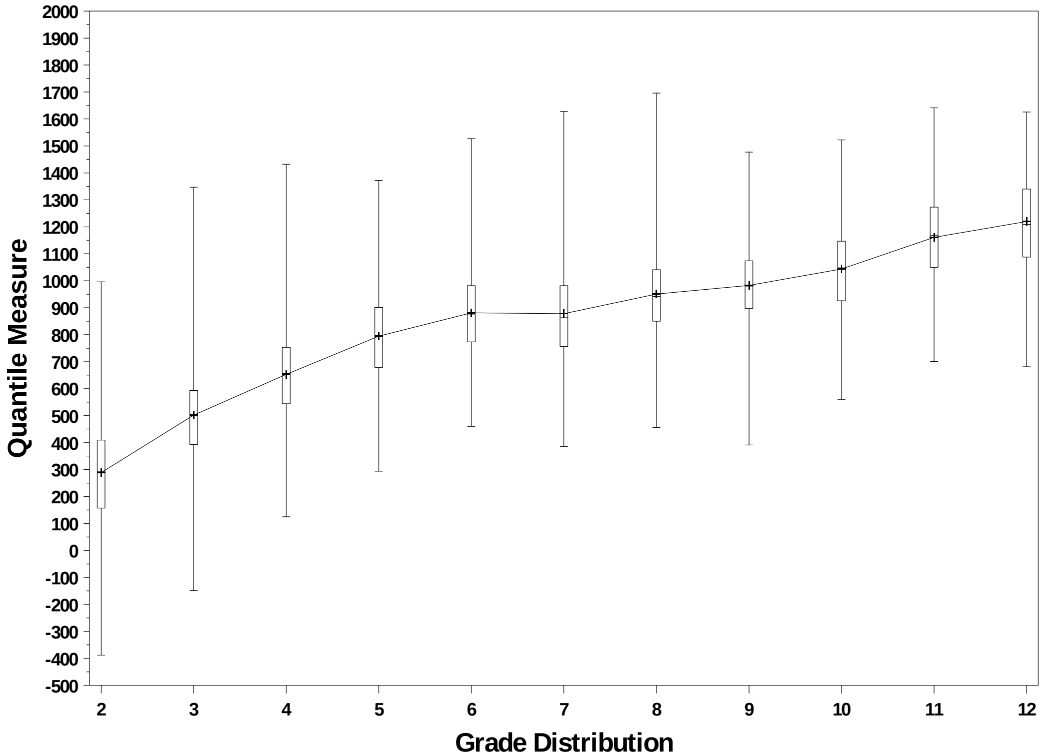
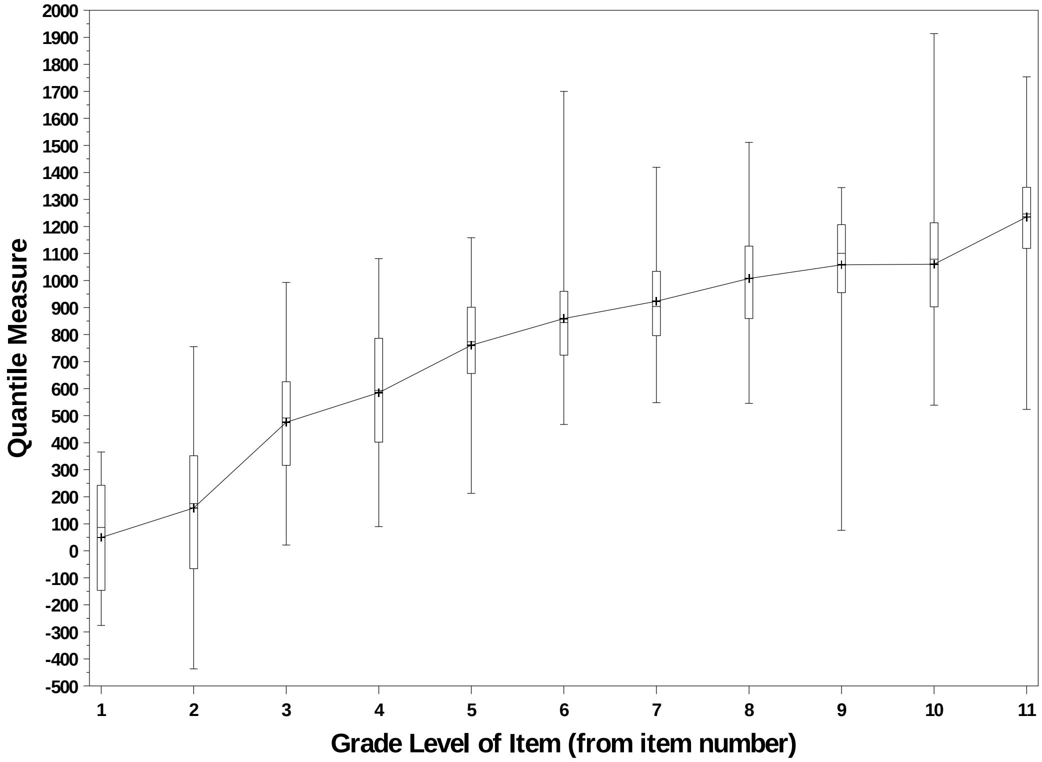


Figure 4 shows the distribution of item difficulties based on the final sample of students. For this analysis, missing data were treated as "skipped" items and not counted as wrong. There is a gradual increase in difficulty when items are sorted by level of test for which the items were written. This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



The field testing of the items written for the Quantile Framework indicates a strong correlation between the grade level of the item and the item difficulty.

The Specification of the Quantile Scale

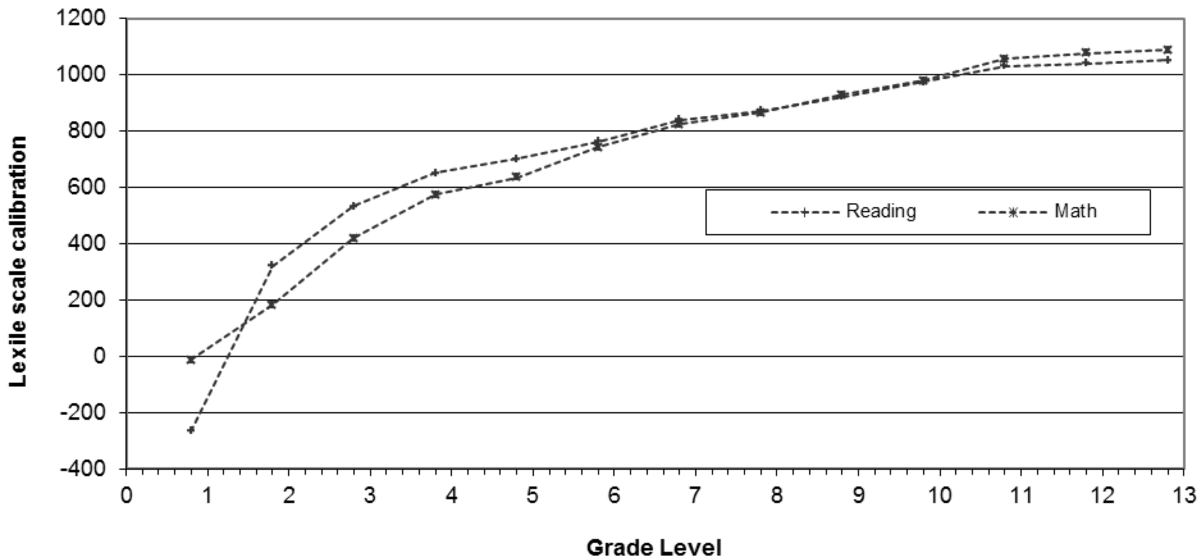
In developing the Quantile scale, two features of the scale were needed: (1) scale multiplier (conversion factor) and (2) anchor point.

As described in the previous section, 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 item 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 used to go from logits to Quantile measures. 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 reading scale.*

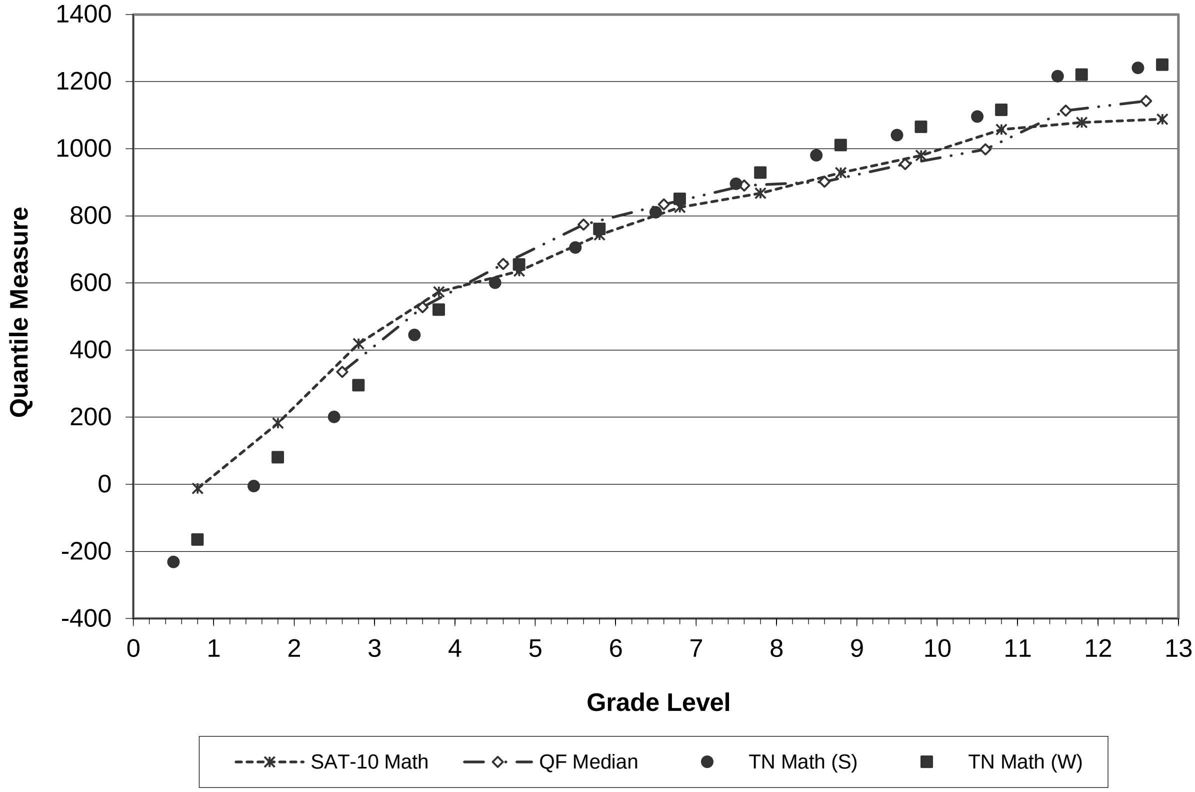


Based on an examination of *Figure 5*, it was concluded that the same conversion factor of 180 that is used with the Lexile scale could be used with the Quantile scale. Both sets of data exhibited a similar pattern across grades.

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 Quantile scale is 700Q. 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).

Therefore, 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) Measures

The next step in the development process was to use the Quantile Framework to estimate the Quantile measure of each QSC. Having a measure for each QSC on the Quantile scale will allow the difficulty of skills and concepts and the complexity of other resources to be evaluated. The Quantile measure of a QSC estimates the solvability, or a prediction of how difficult the skill or concept will be for a learner.

The QSCs are assembled into Knowledge Clusters along a content continuum. Recall that the Quantile Framework is a content taxonomy of mathematical skills and topics. Knowledge Clusters are a family of skills, like building blocks, that depend one upon the other to connect and demonstrate how comprehension of a mathematical topic is 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 trajectory) needed to solve increasingly complex problems (Hudnutt, 2012).

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

For the development of Knowledge Clusters, certain terminology was developed to describe the relationships between QSCs.

- A **focus QSC** is the skills and concept that are the focus of instruction.
- A **prerequisite QSC** is a QSC that describes a skill or concept that provides a building block necessary for another QSC. For example, adding single-digit numbers is a prerequisite for adding two-digit numbers.
- A **supporting QSC** is a QSC that describes associated skills or knowledge that assists and enriches the understanding of another QSC. For example, two supporting QSC are multiplying two fractions and determining the probability of compound events.
- An **impending QSC** describes a skill or concept that will further augment understanding, building on another QSC. An impending QSC for using division facts is simplifying equivalent fractions.

Each focus QSC was classified with prerequisite QSCs and supporting QSCs, or was identified as a foundational QSC. As a part of the taxonomy, QSCs are either a single link in a chain of skills that lead to the understanding of larger mathematical concepts, or they are the first step toward such an understanding. A QSC that is classified as foundational requires only general readiness to learn.

The SMEs examined each QSC to determine where the specific QSC comes in the content continuum based on their classroom experience, instructional resources (e.g., textbooks), and other curricular frameworks (e.g., NCTM Standards). The process called for each SME to independently review the QSC and develop a draft Knowledge Cluster. The second step consisted of the 3-5 SMEs meeting and reviewing the draft clusters. Through discussion and consensus, the SMEs developed the final Knowledge Cluster for the QSC.

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

The Quantile Framework is a theory-referenced measurement system of mathematical understanding. As such, a QSC Quantile measure represents the "typical" difficulty of all items that could be written to represent the QSC and the collection of items can be thought of as an *ensemble* of the all of the items that could be developed for a specific skill or concept. During 2002, Stenner, Burdick, Sanford, and Burdick (2006) conducted a study to explore the "ensemble" concept to explain differences across reading items with The Lexile Framework for Reading. The theoretical Lexile reading measure of a piece of text is the mean theoretical difficulty of all items associated with the text. Stenner and his colleagues state that the "Lexile Theory replaces statements about individual items with statements about ensembles. The ensemble interpretation enables the elimination of irrelevant details. The extra-theoretical details are taken into account jointly, not individually, and, via averaging, are removed from the data text explained by the theory" (p. 314). The result is that when making text-dependent generalizations, text readability can be measured with high accuracy and the uncertainty in expected comprehension is largely due to the unreliability in reader measures.

To determine the Quantile measure of a QSC, actual performance by examinees is used. While expert judgment alone could have been used to scale the QSCs, empirical scaling is 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:

- Met psychometric quality criteria (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; and
- Appropriate for instruction of concept (e.g., first night's homework; from the A and B sections of the lesson problems in textbooks) 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 was defined as the mean Quantile measure of items that met the criteria.

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 pre-requisite and supporting 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 (e.g., 200Q-290Q, 800Q-890Q).

In 2007, with the extension of the Quantile Framework to include Kindergarten and Precalculus, the Quantile measures of the QSCs were reviewed. Where additional items had been tested and the data was available, estimated QSC Quantile measures were calculated. In 2014, a large data set was analyzed to examine the relationship between the original QSC Quantile measures and empirical QSC means from the items administered. The overall correlation between QSC Quantile measures and empirically estimated Quantile measures was 0.98 (N = 7,993 students). Based on the analyses, 12 QSCs were identified with larger-than-expected deviations given the “ensemble” interpretation of a QSC Quantile measure. Each QSC was reviewed in terms of the items that generated the data, linking studies where the QSC was employed, and data from other assessments developed using the Quantile Framework. Of the 12 QSCs identified, it was concluded that the Quantile measure of nine of the QSCs should be recalculated. Five of the QSCs targeted were for Kindergarten and Grade 1 and the new data set provided enough data to calculate an empirical Quantile measure (the Quantile measure for the QSC had previously been estimated). The remaining four QSC Quantile measures were updated to reflect current curricular or pedagogical practices and technological advances because the type of “typical” item and the technology used to assess the skill or concept had shifted from the time that the QSC Quantile measure was established in 2004 (QSCs: 79, 654, 180, and 217). Three of the QSC Quantile measures were not changed (QSC: 134, 604, 408) because (1) some of the items did not reflect the intent of the QSC, or (2) not enough items were tested to indicate that the Quantile measure should be recalculated.

In 2019, the Quantile Framework taxonomy was extended to include advanced statistics and calculus. A total of 74 QSCs were developed (29 Advanced Statistics and 45 Calculus). Five to six items were developed for each new QSC to span the range of content and cognitive complexity and then field tested. A total of 1,170 students enrolled in advanced mathematics or AP calculus or statistics classes participated in the field study. All items were analyzed for psychometric quality and calibrated to the Quantile scale. QSC measures were estimated based on the mean and standard deviation of the item difficulties. QSCs with large item difficulty standard deviations, or, insufficient number of items with adequate psychometric quality (e.g., p-values below .1 or above .95, or point-measure correlation below 0.16) were not estimated (QSCs: 2033, 2037, and 2067). Based on the analysis, a total of 71 QSCs were added to the Quantile Framework and the QSC Quantile measures ranged from 1070Q to 1670Q.

Reporting Quantile Measures

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. A computed Quantile measure of 772.5Q would be represented as 773Q. If the purpose is instructional, then the Quantile measures should be capped at the upper bound of measurement error (e.g., at the 95^{th} percentile of the national Quantile norms) to ensure developmental appropriateness of the instructional material. MetaMetrics expresses these measures used for instructional purposes as “Reported Quantile Measures” and recommends that they be used on individual score reports. The grade level caps used for reporting Grades 2—8 Quantile measures are shown in *Table 6.*

In an instructional environment, all scores below 0Q should be reported as “EMxxxQ”; no student should receive a negative Quantile measure. A Quantile student 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 achievement as the number moves away from zero. The smaller the number following the EM code, the more advanced the student is. 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 student is more advanced than a 150Q student. The lowest reported value below 0Q is EM400Q.

Some assessments report a Quantile range, which is 50Q above and 50Q below the student’s actual Quantile measure. The Quantile range takes into account measurement error found in the tests and in the Quantile measures of the skills/concepts. If a student attempts material above his or her Quantile range, the level of challenge may be too great for the student to be able to construct an understanding of the skill or concept. Likewise, material below the student’s Quantile range may provide the student with little challenge.

Table 6. Maximum reported Quantile measures, by grade.

Grade/Level	Quantile Cap
Kindergarten	600Q
Grade 1	675Q
Grade 2	725Q
Grade 3	975Q
Grade 4	1075Q
Grade 5	1125Q
Grade 6	1200Q
Grade 7	1325Q
Grade 8	1450Q
Grade 9	1475Q
Grade 10	1500Q
Grade 11	1575Q
Grade 12	1650Q

Validity Evidence for 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 2014 *Standards for Educational and Psychological Testing* (American 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 for proposed uses of tests" (p. 11). In other words, a valid test measures what it is supposed to measure.

In applying this definition to the Quantile Framework, the question that should be asked is "What evidence supports the use of the Quantile Framework to describe mathematics skill and concept complexity and student ability?" Stenner, Smith, and Burdick state that "[t]he process of ascribing meaning to scores produced by a measurement procedure is generally recognized as the most important task in developing an educational or psychological measure, be it an achievement test, interest inventory, or personality scale" (1983). 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-identification validity. This global form of validity encompassing content-description and criterion-prediction validity may be evaluated for The Quantile Framework for Mathematics by examining how well Quantile measures relate to other measures of mathematical achievement.

Relationship of Quantile Measures to Other Measures of Mathematical Understanding

Scores from tests purporting to measure the same construct, for example "mathematical achievement," should be moderately correlated (Anastasi, 1982). The Quantile Framework for Mathematics has been linked with numerous 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 7 presents the results from linking studies conducted with the Quantile Framework. For each of the tests listed, student mathematics scores were reported using the test's scale, as well as by Quantile measures. This dual reporting provides a rich, criterion-related frame of reference for interpreting the standardized test scores. Each student who takes a standardized test that has been linked to the Quantile Framework can receive, in addition to norm- or criterion-referenced test results, information related to the specific QSCs on which he or she is ready to be instructed. *Table 6* also shows that measures derived from the Quantile Framework are more than moderately correlated to other measures of mathematical understanding.

Table 7. 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	4,253	0.92
Proficiency Assessments for Wyoming Students (PAWS)	3, 5, 8 11	2,616 537	0.87 0.91
Progress Towards Standards (PTS3)	3-8 and 10	8,544	0.86 to 0.90*
Comprehensive Testing Progressing (CPT 4 – ERB)	3, 5, and 7	802	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
Virginia Standards of Learning (SOL)	3-8, A1, G, and A2	9,519	0.86 to 0.89*
Kentucky Performance Rating for Educational Progress (K-PREP)	3 – 8	6,859	0.81 to 0.85*
North Carolina ACT	11	2,707	0.90
North Carolina READY End-of-Grade/End-of-Course Tests (NC EOG/NC EOC)	3, 4, 6, 8, and A1/I1	8,720	0.87 to 0.90*
aimsweb – Math Concepts and Applications (Pearson)	2 – 8	2,547	0.87
ACT Aspire Math ACT Math	4, 6, 8, and EHS 11 – 12	1,269 650	0.81 0.82
West Virginia SAT Math	11	4,947	0.84
South Carolina READY Mathematics	3 – 8	11,104	0.88
ISIP Early Math ISIP Math	K, 1 2 - 8	1,155 4,332	0.57, 0.67 0.63 – 0.76*
State of Texas Assessments of Academic Readiness (STAAR)	3 – 8, Alg. I	6,350 909	0.86 0.84

*Notes: * Tests were not vertically scaled; separate linking equations were derived for each grade/course.*

Multidimensionality of Quantile Framework Items

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).

The complex nature of mathematics and the curriculum standards most states have adopted also contribute to unintended dimensionality. Application and process skills, the reading demand of items, and the use of calculators could possibly add features to an assessment beyond what the developers intended. These standards, or sub-domains of mathematics, are useful in organizing mathematics instruction in the classroom. These standards could represent different constructs and thereby introduce more sources of dimensionality to tests designed to assess these standards. The following studies were conducted to examine the dimensionality of the Quantile scale.

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) (MetaMetrics, 2014). 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. 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.

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 and the items measured a general construct of mathematics. The results showed that items from all strands loaded most highly on the first (general factor) and no set of items from a particular strand loaded on a specific factor. As a sub-analysis, items from the Geometry and Algebra and Algebraic Thinking strands were analyzed. It was hypothesized that if multidimensionality were to be evidenced in the data, these strands would be the most likely contrast. The Rasch model explained 54.1% of the variance in the Geometry and Algebra and Algebraic Thinking items. The results from the study are consistent with the interpretation of a single construct for each of the analyses (mathematics).

The third step was to examine the results of reading (considered a unidimensional construct) with the mathematics results. The Rasch model explained 60.6% of the variance in the reading comprehension items. Along with the results presented in the first two steps of the process, these data are consistent with the use of a unidimensional item response theory model 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 not a consensus within the measurement community on a single method to determine dimensionality, Burg employed four different methods for assessing dimensionality:

- exploring the conditional covariances (DETECT),
- assessment of essential unidimensionality (DIMTEST),
- item factor analysis (NOHARM), and
- 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. 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.

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.

Furthermore, these findings support the goals of college- and career-readiness standards for mathematics by providing the foundations of a growth model by which a single measure can inform progress toward college and career readiness.

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 were 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 8* shows the number of students and the number of items, combined and by test, for each grade.

Table 8. Number of items included in analyses (Hennings and Simpson, 2012).

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 (2015) 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 9* lists the eigenvalues for the first five principal components by grade, *Table 10* shows the ratios of first-to-second eigenvalues, and *Table 11* shows the proportion of variance accounted for by the first five principal components for each grade.

Table 9. Eigenvalues for the first five principal components, by grade.

	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 10. 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 11. 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 Quantile Framework and Instruction

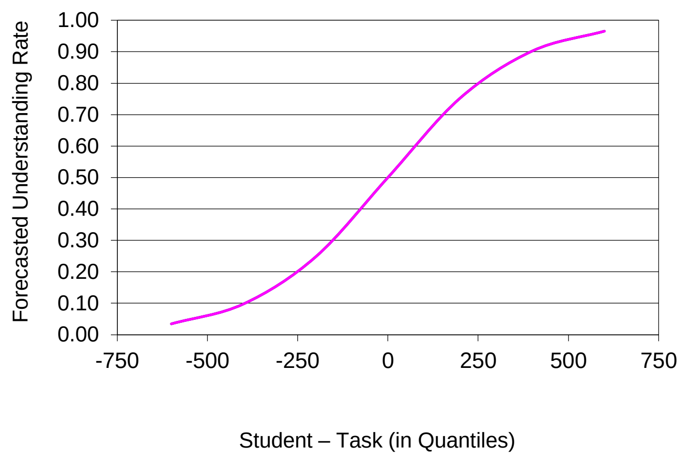
Quantile measures are available from many norm-referenced and criterion-referenced assessments, in addition to state tests and instructional products. Students who take a mathematics achievement test that is linked with the Quantile Framework or one that reports directly in the Quantile metric will receive 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 below EM400Q 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 Precalculus. The measure can be used by a teacher to determine what mathematical instruction the student is likely to be ready for next.

Figure 7 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 particular skill or concept.

Figure 7. *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 to 50Q below the Quantile measure of the student (44% - 56% competency). This range identifies the 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 or 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 7*). This flexibility allows the teacher, parent, or student the ultimate control to modulate the fit between person and task.

Table 12 gives an example of the forecasted understanding (or likely success rates) for specific skills for a specific student. *Table 13* shows forecasted understanding for one specific skill calculated for different student achievement measures.

Table 12. Success rates for a student with a Quantile measure of 750Q and skills of varying difficulty (demand).

Student Mathematics Achievement	Skill Demand	Skill Description	Forecasted Understanding
750Q	350Q	Locate points on a number line.	90%
750Q	550Q	Use order of operations, including parentheses, to simplify numerical expressions.	75%
750Q	750Q	Translate between models or verbal phrases and algebraic expressions.	50%
750Q	950Q	Estimate and calculate areas with scale drawings and maps.	25%
750Q	1150Q	Recognize and apply definitions and theorems of angles formed when a transversal intersects parallel lines.	10%

Table 13. Success rates for students with different Quantile measures of achievement for a task with a Quantile measure of 850Q.

Student Mathematics Achievement	Problems Related to "Locate points in all quadrants of the coordinate plane using ordered pairs."	Forecasted Understanding
450Q	850Q	10%
650Q	850Q	25%
850Q	850Q	50%
1050Q	850Q	75%
1250Q	850Q	90%

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 Quantile Framework's accuracy. The Quantile 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 Quantile Skill and Concept (QSC) 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 student with a Quantile measure of 1000Q being instructed on materials that measure 1000Q (50% understanding) has a successful instructional experience—he or she 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 student with a Quantile measure of 1000Q being instructed on materials that measure 1200Q (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 student with a Quantile measure of 1000Q being instructed on materials that measure 800Q (75% understanding) reports that he or she 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.

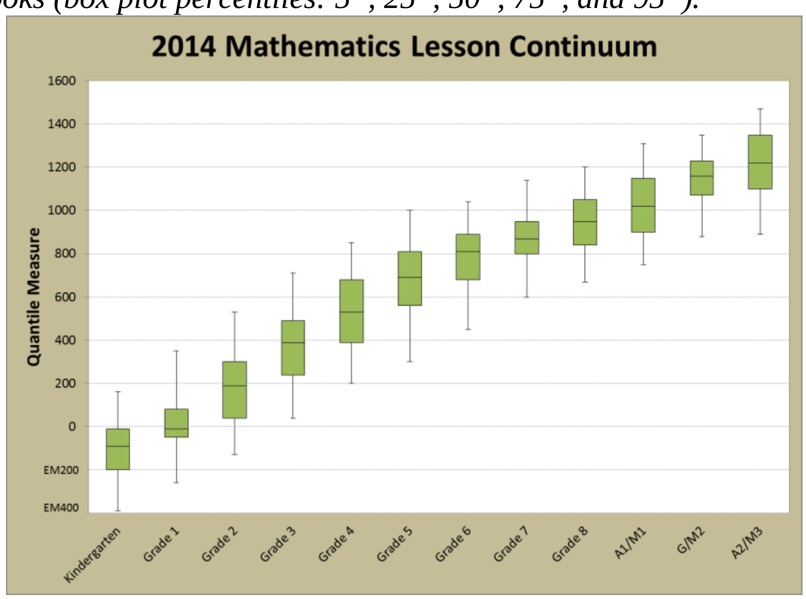
Quantile Framework Mathematical Demands in Education and Careers

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. Many state curricular frameworks developed over the past decade 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. These college- and career-readiness standards for mathematics suggest that "college and career ready" means completing a sequence that covers Algebra 1, 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 throughout K-12.

The *mathematical demand* of a mathematical textbook (in the Quantile metric) 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 QSCs 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 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 by the Quantile Analyzer 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 80,000 instructional materials (e.g., textbook lessons, instructional resources) across the K-12 mathematics curriculum (Smith & Turner, 2012). *Figure 8* shows the continuum of calibrated textbook lessons from Kindergarten through Algebra II/Math 3 from 27,630 lessons (370 test books) from materials published between 2005 and 2013 (Sanford-Moore, Williamson, Bickel, Koons, Baker, and Price, 2014).

Figure 8. *A continuum of mathematical demand for Kindergarten through Precalculus textbooks (box plot percentiles: 5th, 25th, 50th, 75th, and 95th).*



In 2016, Williamson, Sanford-Moore, and Bickel began the examination of the mathematics demands of college and careers to answer the question, "What mathematics will a student likely encounter when entering college or a career?" To address this question, the mathematical concepts and skills that students are likely to encounter as they begin their postsecondary education and/or enter the workplace were examined. For college, being ready for instruction in the types of courses typical of those beyond high school graduation requirements and of first year college were examined (e.g., Precalculus, Trigonometry). For careers, competently performing the mathematics content required for a high school diploma (e.g., Algebra I content, Algebra II content) was examined. In this research, "competently perform" was defined as 75% understanding of the mathematics skills and concepts. The range (interquartile range) of mathematical demands students are likely to encounter as they enter college and careers is 1220Q to 1440Q, with a median of 1350Q.

Recommendations for Using The Quantile Framework for Mathematics

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 a student with textbook lessons, worksheets, games, websites, and trade books that have appropriate Quantile measures (Smith, 2010; Smith and Turner, 2012). By incorporating Quantile measures into the planning of mathematics instruction, it becomes possible to forecast with greater probability how successful 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.

Three resources are available on the Quantile Framework website – the Quantile Math Skills Database, the Quantile Teacher Assistant, and Quantile Math@Home (Smith, 2010; Smith and Turner, 2012, no date). The Math Skills Database (https://hub.lexile.com/math-skills-database) allows teachers and parents to search for Quantile Skills and Concepts (QSCs) by their state standards, by keywords (e.g., adding fractions), and by Quantile measure. The database contains targeted, free resources appropriately matched to students by Quantile measure and math content. 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 (https://hub.lexile.com/quantile-teacher-assistant), teachers can identify a specific state objective or a college and career readiness standard and determine the knowledge base. In addition, teachers can differentiate instruction by indicating the range of Quantile measures for their students in their classrooms. Quantile Math@Home (https://hub.lexile.com/math-at-home) 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 has conducted extensive research to describe the mathematics demands students will likely encounter as they enter college. This research is being extended to describe the mathematics demands of careers student may enter after high school or after additional postsecondary education. Currently, the mathematics demands of more than 450 careers have been examined and the results are available in the Quantile Career Database (hub.lexile.com/career-database).

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 "Summer Math Challenge" is a six-week, e-mail-based initiative designed to help students on summer vacation fight "summer slide" in mathematics skills. The initiative combats summer math slide by helping students retain mathematics skills acquired during the previous school year. The initiative, started in the summer of 2013, targets Grades 1 through 8 by reinforcing mathematics concepts presented from Kindergarten through Grade 8 aligned with college- and career-readiness standards for mathematics. Participants receive targeted instructional materials for a weekly concept along with personalized e-mail activity suggestions and resources that support each concept. Features include activities grounded in everyday life on "Real World Wednesdays," and online math fact fluency practice on "Fluency Fridays." Thirty SEA chiefs requested assistance in launching a 2019 Summer Math initiative in conjunction with the CCSSO Chief's Summer Reading Challenge. Support materials for states and schools are available on the Quantile web site at https://www.quantiles.com/parents-students/find-math-resources-to-support-classroom-learning/summer-math-challenge/. Students from 30 U.S. states participated in the 2019 Summer Math Challenge.

The following list suggests ways 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 Quantile 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 Quantile Math Skills Database on the Lexile & Quantile Hub (hub.lexile.com).
- 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 website 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 a certain amount of 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 websites 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. (2014). *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.

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.

Gorsuch, R. L. (1983). *Factor analysis*. (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

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.

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.

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

Larson. M. R. (2011). *Administrator's guide: Interpreting the Common Core State Standards to improve mathematics education*. Reston, VA: National Council of Teachers of Mathematics.

McGraw-Hill Book Company. (1983). *Guidelines for bias-free publishing*. Monterey, CA: Author.

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

MetaMetrics, Inc. (2014). "Quantile Framework -- Overview." Unpublished manuscript.

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 Governors Association Center for Best Practices (NGA Center) & the Council of Chief State School Officers (CCSSO). (2014). *Key Shifts in Mathematics.* Retrieved from Common Core State Standards Initiative website: http://www.corestandards.org/other-resources/key-shifts-in-mathematics/

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.

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.

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.

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

Sanford-Moore, E.E., Williamson, G.L., Bickel, L., Koons, H., Baker, R.F., & Price, R. (2014). *A quantitative task continuum for K-12 mathematics* (MetaMetrics Research Brief). Durham, NC: MetaMetrics, Inc.

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

SAS Institute. (2015). *SAS/STAT ® 14.1 User's Guide The TRANSREG Procedure*. Cary, NC. Retrieved from https://support.sas.com/documentation/onlinedoc/stat/141/transreg.pdf

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

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

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

Stenner, A.J., Burdick, H., Sanford, E.E., & Burdick, D.S. (2006). How accurate are Lexile text measures? *Journal of Applied Measurement*, *7*(3), 307–322.

Stenner, A.J., Smith, M., & Burdick, D.S. (1983). Toward a theory of construct definition. *Journal of Educational Measurement*, *20*(4), 305-315.

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

Williamson, G.L., Sanford-Moore, E.E., & Bickel, L. (2016). *The Quantile® Framework for Mathematics quantifies the mathematics ability needed for college and career readiness*. (MetaMetrics Research Brief). Durham, NC: MetaMetrics, Inc.

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.

Appendix A

The Quantile® Framework for Mathematics Map

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!

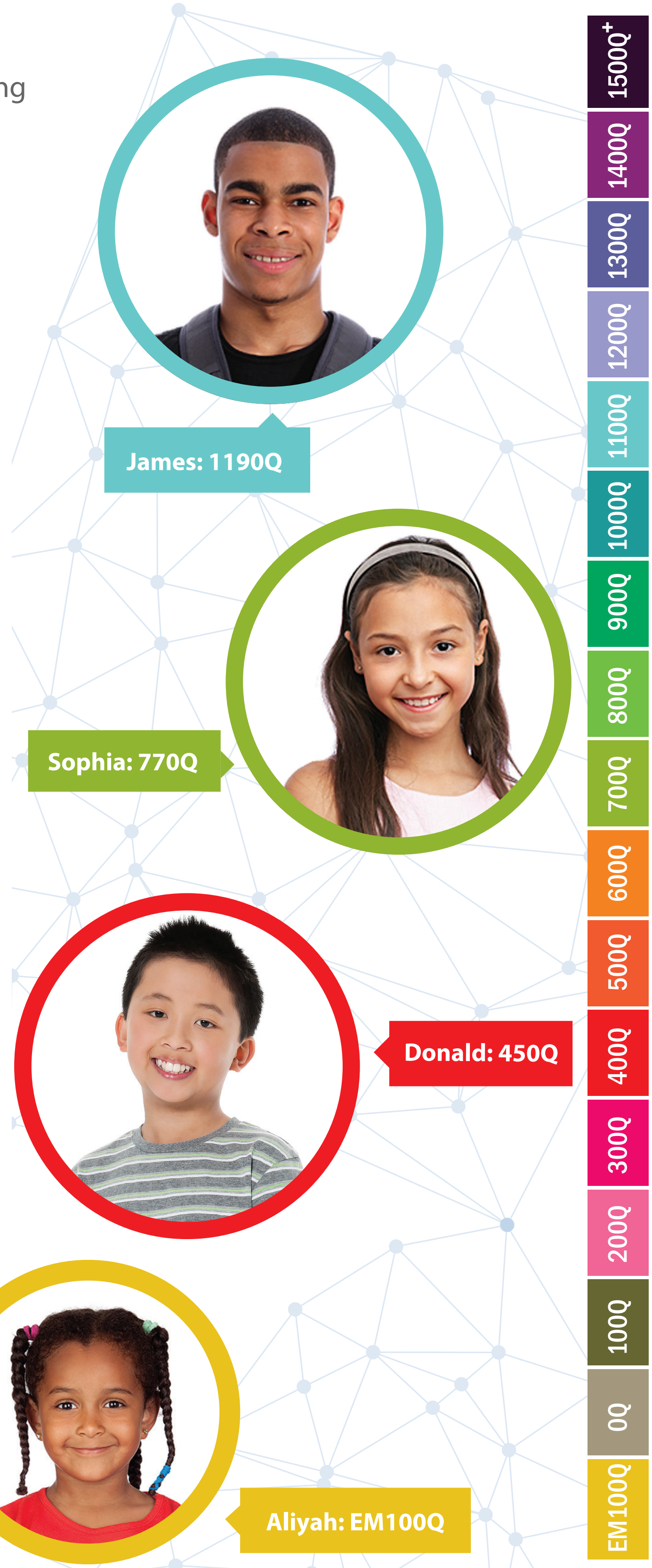
HOW IT WORKS

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.



MetaMetrics

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.

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.

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

MetaMetrics.

For more information, visit Quantiles.com.

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.OA.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