edmentum™

WHITE PAPER

Test Packs Benchmark Assessments from Edmentum Can Bridge the Gap

Executive Summary

Transitioning to the Common Core State Standards

With the start of the 2014–2015 school year, current state tests will be replaced by exams that will evaluate students against a new set of higher, more rigorous standards. The Common Core State Standards Initiative (CCSSI) is a state-led effort to define the knowledge and skills students should acquire within their K–12 education in order to graduate from high school with the ability to succeed in college courses and workforce training programs.

Part of the CCSSI process includes an assessment component designed to document whether students are on track to become college- and career-ready by the time they graduate from high school. Not only will students be held to a more rigorous and relevant set of standards, they will also be introduced to a new type of assessment that is significantly different from current state tests.

So while many educators, schools, districts, and states are currently implementing plans to transition to the new standards, we are clearly in a time of transition. Until the new assessments are in place, most state assessments continue to be based on the old standards. This presents a problem. Many educators are hesitant to teach to the new standards until they are reflected in the assessments. Edmentum is keenly aware of the challenge facing educators as they attempt to bridge current and future assessments.

How Edmentum Can Jumpstart Your Common Core Transition

Gone are the days when states mapped test items to a single standard, and multiple-choice items were expected to be a true measure of mastery. The new Common Core assessment requirements measure critical thinking, problem solving, and 21st century skills. They are aligned not only to factual knowledge, but also to the requirements of college coursework and the demands of the workplace. Naturally, these new standards will require new types of assessments.

To address these challenges, Edmentum has developed an entirely new benchmark assessment solution that has been designed to specifically measure the Common Core State Standards (CCSS). Edmentum's Test Packs Common Core benchmark assessments can be used now to familiarize teachers and students with the new standards. Because the Test Packs provide valid and reliable information on how students are progressing toward the new standards, students will have an advantage when the new tests are operationally administered in 2015.

Getting Ahead of the Challenges with Edmentum Benchmark Assessments

Early indicators reveal that the transition to the new standards and assessments will be particularly challenging for students, educators, and administrators. During the 2011–12 academic year, Kentucky—the first state to adopt the CCSS—administered statewide tests that were developed explicitly to measure proficiency in the Common Core State Standards. While the new Kentucky tests are not the actual tests from either assessment consortium, they are closely aligned and thus seen as a harbinger of things to come. Compared to the previous academic year (2010–11), the percentage of students scoring "proficient" or better dropped by more than 33 percent in math and reading.

Kentucky's steep drop in proficiency rates will likely be repeated in dozens of other states as Common Core assessments are implemented. The results confirmed what most stakeholders expected: Assessments measuring the Common Core will be a significant challenge to educators helping students prepare for college and the workplace. While it's common to see performance declines when new assessment systems are introduced, a lack of available resources compounds the issue for teachers, schools, districts, and states transitioning to the Common Core. The move from No Child Left Behind (NCLB) to the Common Core Standards will require major curriculum changes and new types of assessments. Edmentum has responded to this need and is heavily investing in the development of resources that support classroom learning and provide a full spectrum of year-round assessments.

During the same period in which Kentucky experienced a precipitous decline in test scores, students in schools using Edmentum benchmark assessment solutions outperformed the state average on the new assessments by approximately 25 percent. Figure 1 compares the performance of students who used Edmentum benchmark assessment solutions during the 2011–12 academic year to the state average on Kentucky's Common Core assessments.

Figure 1. Proficiency Rates on Kentucky's Common Core Assessments
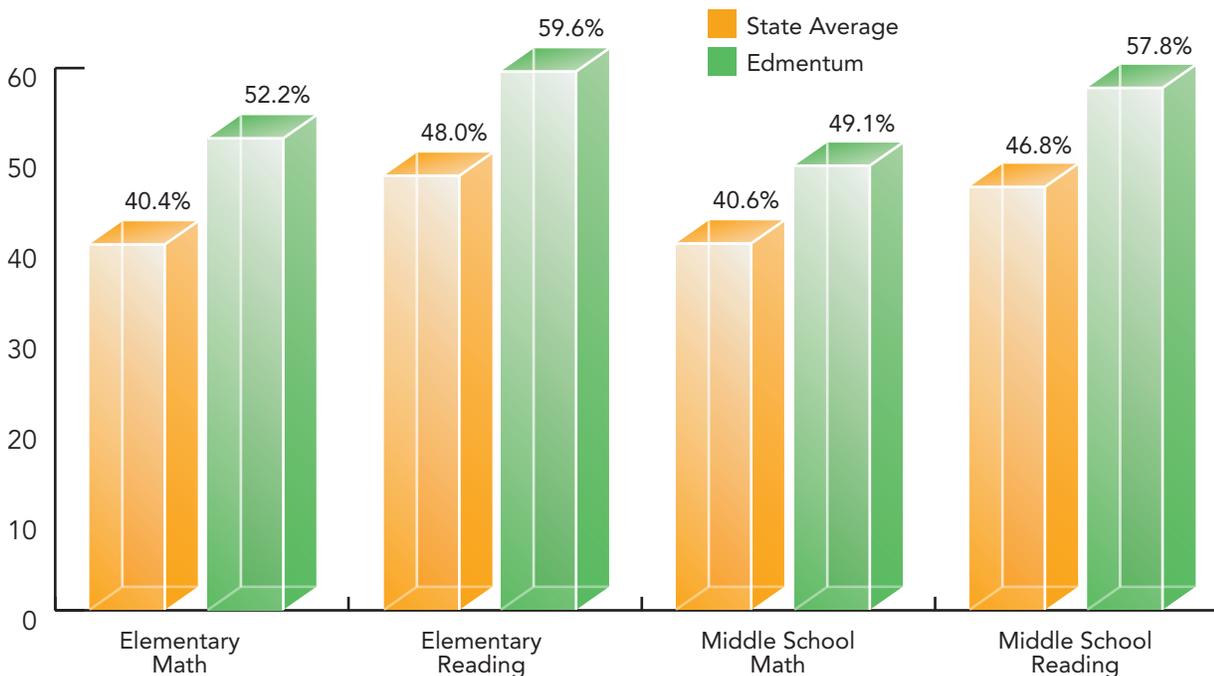
Teachers will bear the brunt of the transition to the Common Core. Ideally, the new standards would be implemented with extensive curricular resources, instructionally supportive assessments, and professional development for teachers, but funding has been scarce in many states still recovering from the recession. Edmentum's benchmark assessments help bridge the gap to help students and educators prepare for the Common Core.

Edmentum's Common Core Benchmark Assessments

This report has been prepared to help educators, schools, districts, and states prepare for the transition to the Common Core assessments that will be administered beginning in the 2014–15 academic year. Edmentum is committed to partnering with educators and administrators as they transition from their current standards to the Common Core State Standards.

To ensure a successful transition, Edmentum has significantly invested in a new series of assessments, developed to measure readiness for the CCSS. Edmentum will continue to invest in these development initiatives, and as specific details emerge from the two assessment consortia, we will add new item types to our Common Core benchmark assessment solutions in order to better measure some of the more complex Common Core State Standards.

Bridging the Gap: The Solution for Educators

Edmentum specifically developed the Test Packs benchmark assessments to measure these new, more challenging standards. Preliminary item- and test-level analyses of the Test Packs detailed in this report indicate that Edmentum's benchmark assessments capture the standards' content focus and higher-order cognitive challenges.

Because students will be challenged by the new standards and corresponding assessments, educational experts have warned of likely precipitous drops in student performance. The 2012 assessment results from Kentucky corroborate the challenges that students, educators, and administrators will encounter once the new assessments are administered operationally. Edmentum's new Common Core benchmark assessments provide accurate details on student progress relative to the CCSS, and give educators and administrators a valid tool to begin the transition to the new standards and assessments.

Overview

Purpose of Report

The purpose of this report is to introduce Edmentum's newest benchmark assessment solution for grades 2–8 and high school. These solutions have been developed specifically to measure the Common Core State Standards in mathematics and English language arts, and will document technical aspects of these new tests.

The Standards for Educational and Psychological Testing (AERA, APA & NCME, 1999) identify professional standards, criteria, and recommendations for test developers and test publishers. One of those standards requires documentation that enables potential test users to evaluate the quality of a test, including evidence for the reliability and validity of test scores. This report provides educators, administrators, decision makers, and other stakeholders the technical information necessary to understand the theoretical framework and design of Edmentum's Common Core benchmark assessment solutions.

Additionally, this report has been prepared to help educators, schools, districts, and states prepare for the transition to the Common Core assessments that will be administered beginning in the 2014–15 academic year. Edmentum is committed to partnering with educators and administrators as they think through the transition from their current standards to the Common Core State Standards.

The State of State Standards Before the Common Core

TThe Common Core State Standards Initiative is the first effort aimed at bringing national consistency and rigor to content standards. The new standards align a wide range of K–12 education content standards into a single, unified platform. Although the CCSS is a national effort, it is not a federal project. Rather, it was led by states through the associations representing governors and chief state school officers.

The CCSS were publicly released on June 2, 2010. States have been encouraged to adopt these standards, with a commitment that the Common Core will constitute a minimum of 85 percent of the state standards in covered content domains. The District of Columbia, 48 states, and two U.S. territories helped develop the standards.

The Common Core Standards are not federally mandated, and some states may choose not to adopt them. The federal government can provide support to states that adopt the standards through a range of tiered incentives, such as providing states with greater flexibility in the use of existing federal funds, supporting a revised state accountability structure, and offering financial support for state implementation.

The intent of the Common Core State Standards is to bring greater rigor to standards driving instruction and expectations, to focus curriculum and instruction on deeper student understanding of a few critically important areas, and to provide clear direction to educators on what American students should know before they leave high school.

State Standards and Assessments Prior to the CCSS: The No Child Left Behind Era

The authority to operate school systems is constitutionally vested in states. Prior to the Common Core State Standards Initiative, all 50 states have independently operated and administered 50 different curricular systems, assessment systems, and performance classifications. While the No Child Left Behind (NCLB) era was controversial, its initiatives attempted to address educational systems challenged by the lack of clear expectations for achievement levels.

NCLB required each state to establish its own standards of what students should know and be able to do—guideposts for academic achievement—in the core content areas. The intent of NCLB was to require states to adopt standards to help direct schools toward common academic goals and unite the education community behind reform and high achievement levels. Under NCLB, states have been required to establish their own annual tests, aligned with state standards that measure how well students are meeting expected learning standards.

While well-intentioned, NCLB left many students unprepared to succeed at entry-level college courses and/or lacking the skills to successfully pursue a career. Every state set its own quality standards for achievement, resulting in wide proficiency variations as measured by National Assessment of Education Progress (NAEP) scores. Fordham Institute's 2007 report, The Proficiency Illusion, made a similar point, questioning the wisdom of using 50 different assessments and 50 different definitions of acceptable performance. How can a school in one state be labeled a failure while a school in another state -- with nearly identical test scores – be considered a success?

In the NCLB era, there was no shortage of criticism or accounts of poorly written state tests that focused on lower-order thinking skills, the manipulation of cut scores to artificially boost the number of students in higher performance levels, and assessments in which students could earn a proficient score with less than 50 percent of items correct.

A Study in Contrasts: The Common Core State Standards Initiative

In 2004, the American Diploma Project (ADP) released the report, Ready or Not: Creating a High School Diploma That Counts. The authors found that both employers and colleges are demanding more of high school graduates than in the past. At the same time, high school graduates are not gaining the skills and knowledge they needed to succeed in a college or workplace setting. The authors used data-based evidence to argue that high school diplomas have lost their value because graduates cannot successfully compete beyond high school. The ADP suggested a series of rigorous benchmarks as a solution to this gap in knowledge and standards.

The Common Core State Standards Initiative was announced on June 1, 2009 with the stated purpose to "provide a consistent, clear understanding of what students are expected to learn, so teachers and parents know what they need to do to help them." The announcement went on to describe how the design strategy would help position American students to compete in a global economy: "The standards are designed to be robust and relevant to the real world, reflecting the knowledge and skills that our young people need for success in college and careers." A year later, in June 2010, the Common Core State Standards were publically released.

The CCSS are internationally benchmarked and backed by evidence showing that students' mastery of the benchmarks prepares them for higher education and the workforce. The initiative defines college and career readiness as the ability "to succeed in entry-level, credit-bearing academic college courses and in workforce-training programs." The CCSS are designed to improve American education in three ways. First, proponents argue, the Common Core is superior to most current state standards. In a recent study, The Fordham Institute concluded that Common Core standards are better than 37 states' standards in English/language arts and 39 states' standards in mathematics. Therefore, proponents believe that the Common Core will raise the quality of education nationally by defining a higher-quality curriculum in English/language arts and mathematics than is currently taught.

> **Myths v. Facts About the Common Core Standards:**
> www.corestandards.org/assets/CoreFacts.pdf
> www.ascd.org/ASCD/pdf/siteASCD/policy/CommonCoreStds.pdf

Proficiency Standards

Supporters also point out that the Common Core sets higher expectations than current state standards. Comparisons with the NAEP lead many analysts to conclude that states set proficiency standards far too low. States routinely report students attaining higher proficiency levels than the NAEP assessments indicate (often 30–40 percentage points more). Even prior to NCLB, states have reported a high number of students as proficient in English and mathematics. Proponents believe that new Common Core tests will presumably end such discrepancies by evaluating proficiency using the same rigorous standards for every state.

Finally, CCSS proponents argue that standard curricular content will yield greater efficiencies. Their expectation is that all same-grade students using high-quality instructional materials aligned to common content standards will be an improvement over the status quo. Additionally, supporters point to the potential gaps in learning that may occur as students move from state to state.

While there are still many questions about the Common Core Standards, one thing is certain: In the 2014–2015 school year, states that have voluntarily chosen to use the CCSS will replace their current tests with one of two exams. These exams will evaluate students against a new set of standards that have been designed to be more rigorous than previous standards. Schools in states that choose to adopt the new standards need a plan to transition from their current testing programs to a new, national achievement test that measures the Common Core. Schools that obtain accurate, reliable, and timely results of student progress on the Common Core now will be better prepared for a successful transition.

Common Core Assessment Consortia

The creation and widespread adoption of the Common Core State Standards presents the single most significant development in student assessment since NCLB. The U.S. Department of Education has funded two consortia to develop a set of common assessments for implementation in spring 2015: the Partnership for Assessment of Readiness for College and Careers (PARCC) and the Smarter Balanced Assessment Consortium (SBAC). These groups are currently developing their respective assessment systems, and Edmentum is closely following their efforts.

The Common Core approach and organizational structure will require different teaching methods than in the past. Rather than teaching a little about a lot of subjects (the so-called "inch deep, mile-wide" curriculum), the Common Core encourages deeper learning on fewer subjects, giving students more complete understanding of the principles behind concepts and a solid foundation for future learning. Testing against these standards will require students to demonstrate greater depth of knowledge, use technology in their answers, and justify their conclusions, rather than picking a letter from a list of multiple-choice questions.

> **Learn more about the Common Core Assessments:**
>
> **Partnership for Assessment of Readiness for College and Careers (PARCC):**
> www.parcconline.org
>
> **Smarter Balanced Assessment Consortium (SBAC):**
> www.smarterbalanced.org

How the Common Core Assessments will be Different

While details about the assessment design and test specifications have not yet been clearly articulated by the two consortia, we know more about how Common Core assessments will differ from current state assessment systems. Assessments that measure student achievement and progress on the Common Core will require new thinking, new technologies, new approaches to content, and new ways to analyze data and report assessment results. One significant change is the shift from basic factual knowledge to an increased emphasis on performance. These next-generation assessments, as described by the two assessment consortia, will address a new set of design considerations and criteria.

Common Core assessments will include multiple-choice items as well as new types of assessments that reflect the depth and rigor of the CCSS, including:

- Performance tasks that help uncover deeper levels of student understanding. Tasks will measure a student's ability to think through a complex problem that may have more than one correct answer. They call for students to apply their knowledge within an authentic learning experience, which may take anywhere from minutes to days to complete. These items will be scored using rubrics based on the cognitive skills being assessed.

- Technology-enhanced items and assessments that provide instant feedback to students and teachers, helping to quickly fill learning gaps and support students in moving steadily toward achievement.

- Extended constructed-response items that measure difficult-to-assess skills. This could include writing an essay or answering an open-ended question in English language arts or mathematics.

- Writing items that require students to write logical arguments based on substantive claims, sound reasoning, and relevant evidence.

Preparing for the Common Core Assessments

Many educators, schools, and districts in states that have adopted the Common Core State Standards are beginning to implement instructional and professional development efforts. One of the significant unknowns, however, is how to effectively monitor and provide feedback on these efforts to students and teachers. Quite simply, we are in a period of transition. Educators are beginning to teach to the Common Core State Standards, but they are still being held accountable to old systems of student assessment.

With all of the focus on the CCSS, it is tempting to think that educators and students are ready for the change that will accompany these new, more rigorous standards. Yet compelling evidence – specifically the early implementations assessments in Kentucky based on the CCSS – suggests there will be a great discrepancy between the theorized effects of the Common Core and reality once the new assessments are administered.

Table 1. Comparison of 2010–11 and 2011–12 Proficiency Rates

	Elementary		Middle	
Kentucky	Math	Reading	Math	Reading
Proficiency rate 2010–11	73.0	76.0	65.0	70.0
Proficiency rate 2011–12*	40.4	48.0	40.6	46.8
Decrease	-32.6	-28.0	-24.4	-23.2
Percentage decrease	-45%	-37%	-38%	-33%

*First year of state tests aligned to the CCSS.

While these results are startling, they are consistent with what most educators and policymakers expected: The new assessments will challenge teachers and students to meet more rigorous demands for proficiency.

Using the same data and methodology that the Kentucky Department of Education used in reporting these results, Edmentum examined performance data from 2010–11 to 2011–12 for Edmentum assessment users within the state of Kentucky. On average, Edmentum assessment users performed significantly better on the state's Common Core-developed assessments than the state average.

Table 2 details the performance of Edmentum assessment users tested on Kentucky's Common Core-aligned assessments compared to the state averages reported by the Department of Education. Figure 1 that follows displays the same information in graphical form.

Table 2. Comparison of 2011–12 Proficiency Rates on Common Core Assessments

	Elementary		Middle	
Kentucky	Math	Reading	Math	Reading
Kentucky state average 2011–12	40.4	48.0	40.6	46.8
Edmentum user average 2011–12	52.2	59.6	49.1	57.8
Increase	+11.8	+11.6	+8.5	+11.0
Percentage increase	+29%	+24%	+21%	+24%

*First year of state tests aligned to the CCSS.

Figure 1. Proficiency Rates on Kentucky's Common Core Assessments
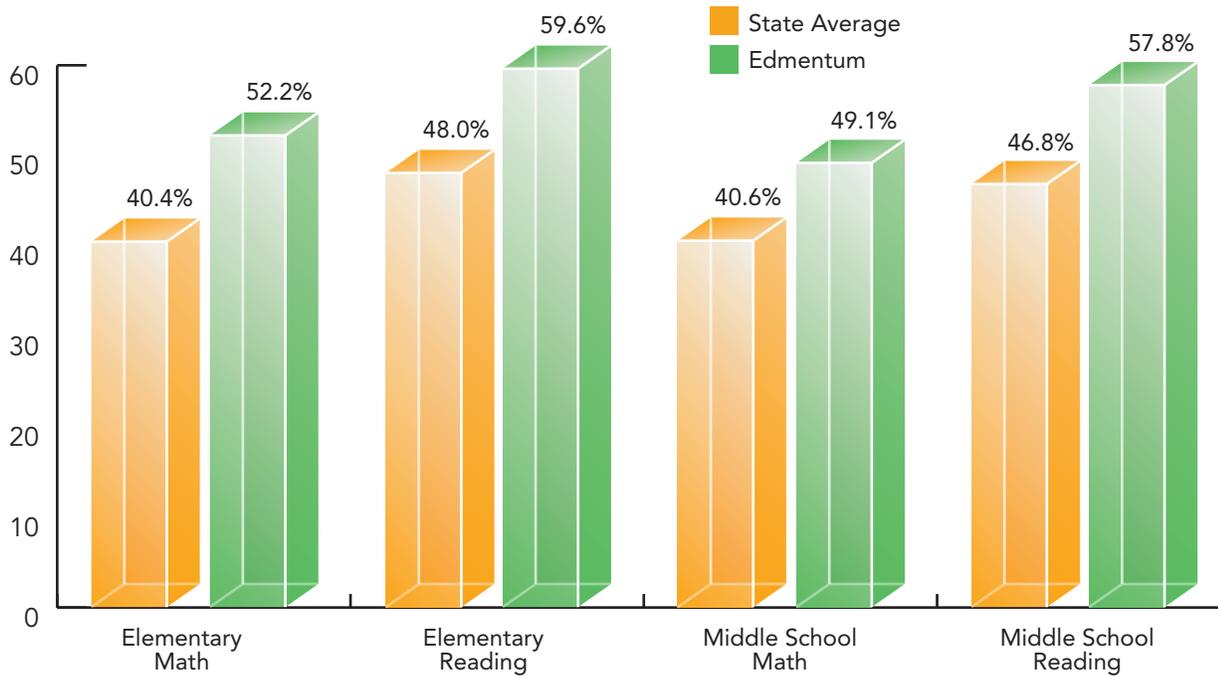
On average, students using Edmentum's benchmark assessment solution performed approximately 25 percent better than the state average on Kentucky's new, more rigorous assessments based on the Common Core. It's highly likely that other states will have a similar experience to Kentucky's as they move toward testing the Common Core, which requires higher levels of rigor in English language arts and math than ever before. Edmentum's benchmark assessments are not only valid and reliable measures of progress towards college and career readiness, but they provide rich information and resources that can be used during this time of transition and uncertainty to help students and educators prepare for the Common Core.

Edmentum has a long history of developing, evaluating, documenting, and refining our assessment solutions to ensure they exceed the expectations of educators and comply with the required standards of professional organizations such as the American Educational Research Association, the National Council on Measurement in Education, and the American Psychological Association.

As we venture together into this new era of assessment, Edmentum is renewing its commitment to documenting the technical strengths of our assessment solutions, including their validity, reliability, and equity. The remainder of this report will document the technical characteristics of Edmentum's Common Core benchmark assessments.

 uce a technical manual that provides information documenting the technical quality of an assessment and its items. This report contains the technical information for Edmentum's Test Packs Common Core benchmark assessments. The information included is intended for use by those who evaluate tests, interpret scores, or use test results in making educational decisions. It is assumed that the reader has some technical knowledge of test construction and measurement procedures.

Overview of Edmentum's Test Packs Common Core Benchmark Assessments

Edmentum's Test Packs Common Core benchmark assessments are designed to measure the academic performance of students in the content areas of mathematics and English language arts in grades 2–8 and high school. Developed specifically to measure the Common Core State Standards, the Common Core benchmark assessments have been developed to provide information about the strengths and needs of individual students, classrooms, schools, and districts.

For more than a decade, Edmentum has been an industry leader in developing assessments that provide valid and reliable data to school- and district-level educators and administrators that can be aggregated or disaggregated at any level. The data reveals trends and patterns in student performance, identifies best practices, or signals the need for additional resources in specific content areas, grade levels, or student groups.

Principles Guiding Edmentum Assessments

Educators use detailed information about individual student achievement to place students in appropriate classes and form instructional groups within a class. They also use it to help parents and others understand the strengths of a particular student's educational development as well as the areas that require further instruction. When constructed and used properly, educational assessments are efficient tools that yield consistent, precise information concerning student achievement and growth.

The purpose of this report is to address how Edmentum's Test Packs Common Core benchmark assessment solutions effectively address these aims through the technical documentation of item- and test-level properties and characteristics. First, it's important to consider the critical role of reliability, validity, and precision to the measurement process. In much the same way as a scale is used to measure the weight of an object, similar principles and qualities apply to the measurement of education.

Reliability is a primary requirement of measurement. Reliability is defined as the consistency of measures obtained from a measurement tool. For example, when an object is weighed on a scale, one expects that additional measurements of the same object with the same scale will have identical results. With repeated observations, it can be determined that the scale measurements are consistent and reliable.

Educational assessments that yield consistent results with repeated testing or within different portions of the same test are considered reliable. Those that provide inconsistent results are unreliable and less valuable to educators. Therefore, one of the primary goals in creating an educational assessment is to create an assessment that has reliable scores.

Validity is another quality that is important in designing and evaluating assessments. Validity can be defined as the degree to which an educational assessment measures what it purports to measure. This attribute of a test and its scores is fundamental to the quality of the measurement.

There are two critical aspects of validity:

• The degree to which an assessment measures the content that will be taught. The degree to which content assessment scores correspond to other indicators of student achievement in the same content area.

Measurements that are accurate and complete indicators of the quality to be measured defines its validity as an expression of the content being assessed. In educational assessment, this is highly dependent upon the curriculum. For example, a generic mathematics assessment may contain content that differs greatly from that which is expected to be taught in the classroom. On the other hand, a mathematics assessment that contains content included in the curriculum fits the content needs of the classroom, school, district, or state more completely. The latter, more valid assessment is clearly more valuable to educators in understanding student achievement. Edmentum's Test Packs Common Core benchmark assessments have been developed to specifically measure the Common Core State Standards.

Precision is a third important aspect of any assessment tool. Precision is the level of detail that the assessment can render. A scale capable of measuring weight to the nearest ounce is more precise than a scale that measures to the nearest pound. With a more precise scale, the variation in any single measurement should be less than one ounce.

In educational assessment, precision (commonly referred to as error) can be thought of in the same manner. The amount of error to be expected in a test score is commonly referred to as the Standard Error of Measurement (SEM). A more precise assessment instrument allows educators to see the difference between two similar students better than a less precise instrument. This explains why measurement precision is valuable to educators – and why it is one of the primary goals in designing an educational assessment. Precision is affected by the amount of information obtained from each test question and the number of questions on a test. Test developers can optimize precision by developing a test that yields a high amount of information from each item.

Edmentum has embraced these principles of sound assessment, and they are reflected in the design of every assessment solution we develop, including our Test Packs Common Core benchmark assessments. The structure of Edmentum assessments, as well as the technology-based testing modality, ensures that Test Packs provide useful information at the student, classroom, school, district, and state levels relative to the Common Core State Standards.

Test Design and Development

I. Test Specifications

A. Criterion-Referenced Test

Edmentum's Test Packs Common Core benchmark assessments are criterion-referenced assessments that reference the Common Core State Standards as their criterion. Items on Edmentum's benchmark assessments have been developed specifically to measure the CCSS. Although items are designed to measure a specific grade level or content area, an item may address several standards within a strand.

The purpose of an achievement test is to determine how well a student has learned important concepts and skills. Test scores are used to make inferences about students' overall performance in a particular domain. Comparing a student's actual performance to a performance target is considered a criterion-referenced interpretation. Comparing a student's performance relative to the performance of other students is considered a norm-referenced interpretation.

Criterion-referenced tests measure students' level of achievement relative to a desired set of learning targets, conceptual understanding, and grade-level or developmentally appropriate skills. It requires care and attention to ensure that the items on the test represent only the desired learning targets, and that there are sufficient items for each learning target to make reliable statements about students' degree of achievement. Criterion-referenced test scores, such as those obtained from Test Packs, are used to make inferences about whether students have attained the desired level of achievement.

B. Item Types

To assist educators in the transition to the new standards, Edmentum's Test Packs Common Core benchmark assessments have been released prior to the consortia's finalization of test specifications and sample item formats. Edmentum's Test Packs currently use the multiple-choice item format across all grade level and content area forms. However, now that the consortia are providing samples of innovative item types to the public, Edmentum is working in parallel to create new items that will mirror those being developed by PARCC and SBAC in terms of cognitive challenge and format. Additionally, as test blueprints and specifications of the new assessments are published, Edmentum will update our benchmark assessments to reflect the areas of emphasis related to content specifications and item types.

The two assessment consortia have indicated that the operational Common Core assessments currently in development will employ additional item types, such as performance tasks, extended constructed response, writing prompts, and additional technology-enhanced items. While the consortia have provided examples of these item types, details about the number of item types have not been released. Edmentum is currently developing and field testing next-generation assessment items and is following the work of the two consortia closely. New item formats will be incorporated into the Test Packs Common Core benchmark assessments as the consortia reveal their plans.

Test Packs Common Core Benchmark Mathematics Test Specifications

I. Standards

The test framework for mathematics for grades 2–8 and high school (Algebra I, Geometry, and Algebra II) is based on the Common Core State Standards for Math, and all items on the grades 2–8 and high school Test Packs tests were designed to measure a specific CCSS objective. Items appearing on the grade level forms are organized according to content categories defined by the CCSS. Additionally, the Standards for Mathematical Practice are embedded throughout the Test Packs grade level assessment forms.

II. Test Design

The table below summarizes the number of items contained within each grade-level mathematics form of Edmentum's Test Packs Common Core benchmark assessment in math. At each grade level, two forms are available for administration. Each form has been developed to the same content specifications.

Table 3. Mathematics Test Items

Grade/Content Area	Number of Items per Form	Number of Grade-level Forms
2	40	2
3	45	2
4	50	2
5	60	2
6	50	2
7	50	2
8	50	2
Algebra I	49	2
Geometry	52	2
Algebra II	39	2

III. Test Blueprints

The distribution of emphasis for Test Packs Common Core content strands for mathematics is shown in the tables below.

Table 4. Grades 2–5 Mathematics Test Blueprint

	Grade 2		Grade 3		Grade 4		Grade 5	
CCSS Reporting Category	Items	%	Items	%	Items	%	Items	%
Operations and Algebraic Thinking	6	15%	17	38%	8	16%	7	12%
Measurement and Data	13	33%	12	27%	12	24%	13	22%
Number and Operation-Fractions	N/A	N/A	6	13%	13	26%	17	28%
Number and Operations in Base Ten	15	38%	5	11%	10	20%	15	25%
Geometry	6	15%	5	11%	7	14%	8	13%
Total	40	100%	45	100%	50	100%	60	100%

Table 5. Grades 6–8 Mathematics Test Blueprint

	Grade 6		Grade 7		Grade 8	
CCSS Reporting Category	Items	%	Items	%	Items	%
Expressions and Equations	11	22%	9	18%	10	20%
The Number System	10	20%	9	18%	N/A	N/A
Ratios and Proportional Relationships	10	20%	12	24%	6	12%
Statistics and Probability	8	16%	12	24%	7	14%
Geometry	6	12%	8	16%	14	28%
The Number System—Absolute Value	5	10%	N/A	N/A	N/A	N/A
Expressions and Equations—Linear Equations	N/A	N/A	N/A	N/A	7	14%
Functions	N/A	N/A	N/A	N/A	6	12%
Total	50	100%	50	100%	50	100%

Table 6. Algebra I Mathematics Test Blueprint

CCSS Reporting Category	Items	%
Interpreting Categorical and Quantitative Data	9	18%
Interpreting Functions	8	16%
Seeing Structure in Expressions	6	12%
Reasoning with Equations and Inequalities	6	12%
Building Functions	4	8%
Creating Equations	4	8%
Functions: Linear, Quadratic, and Exponential Models	3	6%
The Real Number System	3	6%
Quantities	3	6%
Arithmetic with Polynomials and Rational Expressions	3	6%
Total	49	100%

Table 7. Geometry Mathematics Test Blueprint

CCSS Reporting Category	Items	%
Congruence	9	17%
Expressing Geometric Properties with Equations	8	15%
Conditional Probability and the Rules of Probability	7	14%
Similarity, Right Triangles	7	14%
Trigonometry	7	14%
Modeling with Geometry	5	10%
Circles	5	10%
Geometric Measurement and Dimension	4	8%
Total	52	100%

Table 8. Algebra II Mathematics Test Blueprint

CCSS Reporting Category	Items	%
Trigonometric Functions	5	13%
Interpreting Functions	7	18%
Seeing Structure in Expressions	3	8%
Statistics and Probability	5	13%
Reasoning with Equations and Inequalities	3	8%
Building Functions	3	8%
Creating Equations	4	10%
The Complex Number System	4	10%
Arithmetic with Polynomials and Rational Expressions	5	15%
Total	39	100%

Test Packs Common Core Benchmark English Language Arts Test Specifications

I. Standards

The test framework for English language arts for grades 2–11 is based on the Common Core State Standards for English language arts, and all items on the grades 2–11 Test Packs tests were designed to measure a specific CCSS objective. Items appearing on the grade level forms are organized into the content categories defined by the CCSS.

II. Test Design

The table below summarizes the number of items contained within each grade level English language arts form of Edmentum's Test Packs Common Core benchmark assessment. At each grade level, two forms are available for administration. Each form was developed to the same content specifications.

Table 9. English Language Arts Test Items

Grade/Content Area	Number of Items per Form	Number of Grade-level Forms
2	30	2
3	45	2
4	45	2
5	50	2
6	51	2
7	50	2
8	49	2
9	51	2
10	52	2
11	49	2

III. Test Blueprints

The distribution of emphasis for Test Packs Common Core content strands for English language arts is shown in the table

Table 10. Grade 2 English Language Arts Test Blueprint

CCSS Reporting Category	Items	%
Conventions of Standard English	13	43%
Reading Informational Text	7	23%
Reading Literature	6	20%
Writing	4	13%
Total	30	100%

Table 11. Grades 3–6 English Language Arts Test Blueprint

	Grade 3		Grade 4		Grade 5		Grade 6	
CCSS Reporting Category	Items	%	Items	%	Items	%	Items	%
Conventions of Standard English	12	27%	7	16%	8	16%	7	14%
Text Type and Purposes	8	18%	11	24%	14	28%	7	14%
Vocabulary Acquisition and Use	7	16%	5	11%	5	10%	6	12%
Informational Text Craft and Structure	3	7%	3	7%	3	6%	5	10%
Informational Text Key Ideas and Details	3	7%	4	9%	4	8%	5	10%
Informational Integration of Knowledge and Ideas	3	7%	4	9%	5	10%	4	8%
Literature Key Ideas and Details	3	7%	4	9%	4	8%	5	10%
Literature Integration of Knowledge and Ideas	3	7%	N/A	N/A	N/A	N/A	3	6%
Literature Craft and Structure	3	7%	3	7%	4	8%	6	12%
Research to Build and Present Knowledge	N/A	N/A	4	9%	3	6%	3	6%
Total	45	100%	45	100%	50	100%	51	100%

Table 12. Grades 7–9 English Language Arts Test Blueprint

CCSS Reporting Category	Grade 7 Items	Grade 7 %	Grade 8 Items	Grade 8 %	Grade 9 Items	Grade 9 %
Conventions of Standard English	4	8%	6	12%	7	14%
Text Type and Purposes	6	12%	11	22%	11	22%
Vocabulary Acquisition and Use	6	12%	6	12%	7	14%
Informational Text Craft and Structure	5	10%	5	10%	4	8%
Informational Text Key Ideas and Details	6	12%	5	10%	4	8%
Informational Integration of Knowledge and Ideas	5	10%	4	8%	4	8%
Literature Key Ideas and Details	5	10%	5	10%	5	10%
Literature Integration of Knowledge and Ideas	3	6%	N/A	N/A	N/A	N/A
Literature Craft and Structure	6	12%	4	8%	3	6%
Research to Build and Present Knowledge	4	8%	3	6%	6	12%
Total	50	100%	49	100%	51	100%

Table 13. Grades 10 and 11 English Language Arts Test Blueprint

CCSS Reporting Category	Grade 10 Items	Grade 10 %	Grade 11 Items	Grade 11 %
Conventions of Standard English	7	14%	4	8%
Text Type and Purposes	10	19%	11	22%
Vocabulary Acquisition and Use	8	15%	N/A	N/A
Informational Text Craft and Structure	4	8%	5	10%
Informational Text Key Ideas and Details	4	8%	5	10%
Informational Integration of Knowledge and Ideas	4	8%	4	8%
Literature Key Ideas and Details	5	10%	5	10%
Literature Craft and Structure	4	8%	4	8%
Research to Build and Present Knowledge	6	12%	4	8%
Knowledge of Language	N/A	N/A	7	14%
Total	52	100%	49	100%

Test-Development Process

The Edmentum test-development process is an iterative, recursive process that has been implemented and refined for over a decade. Modeled on methods pioneered by experts in the fields of psychometrics and educational assessment, the process has since been refined to maximize the quality and efficiency of Edmentum assessments. The process is guided by the standards for educational and psychological testing as established by professional organizations in the field (AERA, APA, & NCME, 1999). Additional input for the test design process was taken from the Common Core State Standards as well as information available from PARCC and SBAC.

Edmentum's Test Packs Common Core benchmark assessments were developed using best-practice criteria emphasized by leading assessment expert Thomas Haladyna. Additional assessment best-practice guidelines were derived from Measurement and Assessment in Teaching (Linn & Miller, 2005).

The central goal in test development at Edmentum is to develop a robust pool of high-quality items. Edmentum has established definitive standards for high-quality items by adhering to the principles of assessment leaders, and by engaging a comprehensive design, development, and review process. High-quality items are unambiguous, closely measuring the intended learning objective. Barriers to meeting this objective include ambiguous statements, excessive wordiness, difficult vocabulary, unclear instructions, and bias (Haladyna, 2004).

I. Item Development

Edmentum uses the following process to develop new test items:

1. Select an appropriate item type
The first step is to select the most appropriate item type (multiple choice, true/false, matching, or gridded response) to meet the stated learning objective.

2. Follow quality guidelines
Following guidelines and principles for each item type helps minimize ambiguity and ensure that the item matches the intended learning outcome. Edmentum uses these principles as guidelines for writing items.

All item types adhere to the following guidelines:

- Model correct grammar, punctuation, usage, and spelling
- Use a clear, concise writing style
- Contain clear directions on what is required to attain a maximum score
- Use a reading level that allows the student to demonstrate his or her knowledge of the tested subject matter, regardless of reading ability
- Exhibit high technical quality in psychometric characteristics
- Provide appropriate answer options
- Avoid potentially sensitive content
- Use novel material to test higher-level learning
- Avoid trivial content

3. Provide targeted trainings
Edmentum training is developed around each content area to guide writers in creating new items. Curriculum experts, assessment specialists, and instructional designers write and conduct peer reviews of sample item sets. Using the results of the item reviews as a needs assessment, subject area leaders meet to identify which training sessions to develop. This targeted training includes topics such as writing grade-level appropriate items, using best practices in assessment writing, incorporating Edmentum's acceptance criteria for quality assurance, and avoiding bias in item writing.

4. Evaluate quality
Next, Edmentum uses the item-writing guidelines to define a quality measure or acceptance criteria. We use these criteria to evaluate every written item, ensuring that it meets Edmentum's expected quality level. Subject-area teams review every item in committee, using the acceptance criteria and item-writing guidelines to evaluate item quality. If an item does not meet the acceptance criteria, it is not used on an operational Edmentum test form. In addition, the acceptance criteria forms are tracked using metrics that help Edmentum developers make decisions about future training.

II. Item Reviews

Edmentum engages in rigorous, thorough reviews of every item developed as part of a specified test construction effort. For these internal item reviews, the lead Edmentum test developer within the content area performs the following activities:

- Reviews the formatted item, scoring guide, and any reading selections and graphics
- Evaluates item integrity, content, and structure; appropriateness to designated content area; format; clarity; possible ambiguity; answer cueing; appropriateness and quality of reading selections and graphics; and appropriateness of scoring guide descriptions and distinctions (in relation to each item and across all items within the guide)
- Verifies that each item has only one correct answer
- Determines whether the scoring guide adequately addresses performance on the item

The lead Edmentum test developer considers the following questions when evaluating an item:
- What is the item asking?
- Is the item appropriate for the designated grade level?
- Is the key the only possible key? (Is there only one correct answer?)
- Can the item be scored as written? (Were the correct words used to elicit the response defined by the guide?)
- Is the wording of the scoring guide appropriate and parallel to the item wording?
- Is the item complete? (Does it include a scoring guide, content codes, key, and grade level?)

III. Bias and Sensitivity Review

Bias review is an essential part of the development process. During the bias review process, Edmentum Common Core benchmark assessment items and passages were reviewed by committees of curriculum and assessment experts representing the interests of legally protected and/or educationally disadvantaged groups. Items and passages were examined for content that could be deemed offensive or controversial by students, teachers, or parents. Including such groups in test item and material development can prevent unintended consequences and distractions before the test forms are produced.

IV. Item Selection and Operational Test Form Assembly

At Edmentum, test assembly involves the sorting and laying out of item sets into test forms. During the assembly process, criteria considered for the Test Packs Common Core benchmark assessments includes the following:

- **Content coverage/match to test design**. Edmentum test developers complete an initial sorting of items into sets, based on a balance of CCSS-reporting categories across forms, as well as a match to the test design.

- **Item difficulty and complexity**. Items are used to ensure similar levels of difficulty and complexity across forms.

- **Visual balance**. Item sets are reviewed to ensure comparable presentation (e.g., length/complexity of reading selections, number of graphics).

- **Response option balance**. Each item set is checked to ensure an equivalent number of key options (As, Bs, Cs, and Ds).

- **Bias**. Each item set is reviewed to ensure fairness and balance based on gender, ethnicity, religion, socioeconomic status, and other factors.

Classical Item Analysis

As noted by F. G. Brown (1983), "A test is only as good as the items it contains." A complete evaluation of a test's quality must include an evaluation of each item. Both Standards for Educational and Psychological Testing (AERA, APA & NCME, 1999) and the Code of Fair Testing Practices in Education (2004) include standards for identifying quality items. Items should assess only knowledge or skills that are identified as part of the tested domain, and should avoid assessing irrelevant factors. Items should also be unambiguous and free of grammatical errors, potentially insensitive content or language, and other confounding characteristics. In addition, items must not pose unfair disadvantages to students in particular racial, ethnic, or gender groups.

Both qualitative and quantitative analyses are conducted to ensure that Edmentum's Test Packs Common Core benchmark assessment items meet these standards. Qualitative analyses are described in earlier sections of this report; this section focuses on quantitative evaluations. Statistical evaluations in this section are presented in terms of difficulty indices. A future version of this report will also include item-test correlations (discrimination), differential item functioning statistics, and dimensionality analyses. The item analyses presented here are based on administrations of items contained in Edmentum's Test Packs Common Core benchmark assessments through February 2013.

I. Classical Difficulty Indices

When sufficient student response data have been obtained (n>50),all items in Edmentum's Test Packs Common Core benchmark assessments are evaluated for difficulty according to standard classical test theory practices. Difficulty is defined as the average proportion of points achieved on an item; it is measured by obtaining the average score on an item and dividing it by the maximum possible score for the item.

Multiple-choice items are scored dichotomously (correct versus incorrect); and for these items, the difficulty index is simply the proportion of students who correctly answered the item. Computing the difficulty index as the average proportion of points achieved puts all items on a similar scale, ranging from 0.0 to 1.0. Although this index is traditionally described as a measure of difficulty, it is properly interpreted as an easiness index because larger values indicate easier items. An index of 0.0 indicates that all students received no credit for the item, and an index of 1.0 indicates that all students received full credit for the item.

Items that are correctly answered by almost all students provide little information about differences in student abilities, but they do indicate knowledge or skills that have been mastered by most students. Similarly, items that are correctly answered by very few students provide little information about differences in student abilities, but may indicate knowledge or skills that have not yet been mastered by most students. In general, to provide the best measurement, difficulty indices should range from near-chance performance (0.25 for four-option multiple-choice items) to 0.90, with the majority of items generally falling between 0.4 and 0.7. However, on a standards-referenced assessment such as Edmentum's Test Packs Common Core benchmark assessment, it may be appropriate to include some items with very low or very high item difficulty values to ensure sufficient content coverage.

It is desirable to include an item that allows higher-ability students to perform better than lower-ability students. On a single item, the correlation between student performance and total test score is a commonly used measure of this item characteristic. Within classical test theory, the item-test correlation is referred to as the item's discrimination because it indicates the extent to which successful performance on an item discriminates between high and low scores on the test.

For multiple-choice items, the corresponding statistic is commonly referred to as a point-biserial correlation. The theoretical range of these statistics is from –1.0 to 1.0, with a typical observed range from 0.2 to 0.6. For the current report, item discrimination indices were not obtained, but they will be included in the second phase of this project (estimated 2013–14 academic year).

The tables below summarize, by content area, the classical item-level statistics for each grade-level form in which sufficient data was obtained. The following indices were obtained for each grade-level form.

- **Items on Form:** The number of items on the grade level assessment form

- **# Analyzed:** The number of items analyzed on the grade level form for which sufficient student response data were available

- **Mean P:** The average (mean) item difficulty statistic for the test form; also the average proportion-correct score

- **Median P:** The average (median) item difficulty statistic for the test form; also the average proportion-correct score

- **SD:** The standard deviation, which is a measure of dispersion (a range of ± two SDs from the mean includes approximately 95 percent of the examinees)

- **Min:** The minimum item difficulty statistic for the test form

- **Max:** The maximum item difficulty statistic for the test form

Table 14. Summary of Mathematics Item Difficulty Statistics by Grade Level Form

Grade	Form	Items on Form	# Analyzed	Mean P	Median P	SD	Min	Max
3	1	45	15	.53	.49	.21	.26	.98
3	2	45	16	.55	.51	.18	.28	.93
4	1	50	11	.59	.58	.18	.25	.96
4	2	50	18	.61	.61	.19	.29	.90
5	1	60	11	.46	.43	.13	.28	.66
5	2	60	8	.52	.50	.18	.35	.88
6	1	50	43	.49	.46	.21	.11	.87
6	2	50	45	.51	.52	.19	.13	.87
7	1	50	48	.57	.57	.22	.16	.97
7	2	50	13	.47	.47	.21	.15	.90
8	1	50	50	.40	.39	.17	.07	.74
8	2	50	10	.46	.47	.18	.21	.74
ALG	1	49	49	.31	.28	.13	.10	.78
ALG	2	49	12	.36	.31	.18	.17	.78
GEO	1	52	52	.38	.35	.19	.10	.92
GEO	2	52	12	.31	.29	.19	.09	.87
ALG2	1	39	39	.31	.31	.11	.11	.62

Table 15. Summary of English/Language Arts Item Difficulty Statistics by Grade Level Form

Grade	Form	Items on Form	# Analyzed	Mean P	Median P	SD	Min	Max
2	1	30	28	.48	.46	.14	.22	.85
3	1	45	45	.52	.51	.15	.27	.89
3	2	45	9	.43	.37	.19	.22	.76
4	1	45	45	.54	.53	.17	.17	.84
5	1	50	49	.55	.55	.17	.23	.98
5	2	50	13	.60	.56	.19	.26	.83
6	1	51	19	.48	.46	.18	.12	.73
6	2	51	9	.50	.47	.14	.33	.77
7	1	50	16	.51	.46	.18	.23	.74
7	2	50	9	.44	.46	.17	.24	.68
8	1	49	13	.46	.37	.23	.13	.84
8	2	49	12	.44	.45	.09	.26	.61
9	1	51	50	.48	.50	.21	.12	.94
9	2	51	11	.51	.46	.16	.30	.78
10	1	52	52	.51	.51	.20	.11	.86
10	2	52	13	.50	.51	.16	.25	.84
11	1	49	11	.45	.46	.11	.25	.63
11	2	49	9	.37	.38	.13	.19	.55

A comparison of indices across grade levels is complicated because these indices are population- dependent. Direct comparisons would require commonality in either items or students across groups. Since that is not the case, it cannot be determined whether differences in performance across grade levels reflect differences in student abilities, differences in item difficulties, or both. With this caveat in mind, it generally appears that mathematics students in higher grade levels found their items more difficult than students in lower grades. In the area of reading, difficulty indices were more consistent across grades.

As noted previously in this report, the CCSS are widely considered more rigorous than existing standards. Edmentum has developed the Test Packs benchmark assessments to specifically measure these new, more challenging standards rather than repackaging existing assessments as "Common-Core aligned."

The preliminary item and test-level analyses of Test Packs as detailed above indicate that Edmentum's benchmark assessments capture the focus and increased cognitive challenge of the standards. Students will be challenged by the new standards and corresponding assessments, and educational experts have warned of likely precipitous drops in student performance.

Kentucky's 2012 assessment results corroborate the challenges that students, educators, and administrators will encounter once the new assessments are operationally administered. As evidenced by the data above, Edmentum's Test Packs Common Core benchmark assessments are challenging. Edmentum's new assessment solution provides highly accurate information on student progress relative to the CCSS, and gives educators and administrators a valid tool to ease the transition.

Validity

Because interpretations of test scores, and not the test itself, are evaluated for validity, the purpose of the Test Packs Common Core Benchmark Assessment technical report is to describe several technical aspects of the tests in support of score interpretations (AERA, APA & NCME, 1999). Each section of this report contributes an important component in the investigation of score validation, including test development and design, test administration, and item analyses.

The Standards for Educational and Psychological Testing (AERA, APA & NCME, 1999) provides a framework for describing sources of evidence that should be considered when constructing a validity argument. The evidence around test content, response processes, internal structure, relationship to other variables, and consequences of testing all speak to different aspects of validity – but they are not distinct types of validity. Instead, each contributes to a body of evidence about the comprehensive validity of score interpretations.

Evidence on test content validity is used to determine how well the assessment tasks represent the curriculum and standards for each grade level form and content area. Content validation is informed by the item development process, including how the test blueprints and test items align to the curriculum and standards the assessment is intended to measure.

Viewed through the CCSS lens, evidence based on test content was extensively described in the Test Design and Development and Test Specifications sections of this report. Based on test content, components of validity evidence include item alignment with the Common Core State Standards; item bias, sensitivity, and content appropriateness review processes; adherence to the test blueprint; use of standardized administration procedures with accommodated options for participation; and appropriate test administration training.

As discussed earlier, all Test Packs items are aligned by experts trained in curriculum and assessment, and items are extensively reviewed for content fidelity and appropriateness. Finally, tests in a computer-based format are administered in a standardized manner, with allowable accommodations. All test coordinators and administrators are required to become familiar with and adhere to all of the procedures outlined in the Test Packs Common Core Benchmark Assessment Test Administrator Manual.

Detailed evidence based on internal structure is presented in the classical item analyses. Technical characteristics of the assessment's internal structure are presented using classical item-difficulty statistics.

In general, item difficulty and indices were within acceptable and expected ranges. Very few items were correctly answered at near-chance or near-perfect rates. Edmentum's Test Packs Common Core benchmark assessment is an entirely new solution in its first year of operational administration; a second phase of this report will be completed during the 2013–14 academic year. This report will detail the technical characteristics of the assessment's internal structure, including comprehensive item difficulty, discrimination (item-test correlation), differential item functioning analyses, dimensionality analyses, reliability, and standard errors of measurement.

Future validation studies may be undertaken to provide evidence on the relationship of Test Packs results to other variables, including Test Packs score convergence with measures of similar constructs, and the extent to which they diverge from measures of differing constructs. Refining the construct definition can reveal relationships among similar construct measures, informing scores and score interpretations.

Future Phases

When the CCSS were initially released, Edmentum's team of curriculum and assessment experts carefully analyzed the differences between existing standards and the Common Core. Because the Common Core requires students to master skills that have rarely been assessed by state or benchmark assessments, our content experts found that existing assessments could not adequately prepare students for the new standards. Simply reconfiguring our existing tests without regard for the instructional intent of the Common Core would have run contrary to the high level of quality that educators expect -- and that we demand of ourselves.

Edmentum has chosen a different strategy for fully assessing the Common Core. The Test Packs Common Core benchmark assessment solution is one way Edmentum is partnering with educators to bridge the gap during this time of transition.

Not only is Edmentum building new items to match the innovative item types in the new assessments, but it is also evaluating how the PARCC and SBAC tests will be built. As the consortia finalize their assessment blueprints, Edmentum will modify Test Packs Common Core benchmark to reflect these respective specifications. Our goals are to create high-quality items and build upon Edmentum's integrated suite of assessment solutions for formative, benchmark, and summative use in measuring the CCSS.

These flexible and customizable assessments will give classroom teachers valid and reliable information on the progress of their students relative to the CCSS. Edmentum is building a solution that is responsive enough to meet the needs of the classroom teacher and rigorous enough to meet the new demands of assessing the Common Core State Standards.

At Edmentum, our work has just begun. We will continue to make significant investments in products that help educators prepare students for college and the workplace.

References

Achieve, Inc., The Education Trust, & The Thomas B. Fordham Foundation. (2004). R*eady or not: Creating a high school diploma that counts.* Washington, DC: American Diploma Project.

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

Brown, F. G. (1983). *Principles of educational and psychological testing* (3rd ed.). Fort Worth: Holt, Rinehart and Winston.

Haladyna, T. M. (2004). *Developing and validating multiple-choice test items.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Joint Committee on Testing Practices (2004). *Code of fair testing practices in education.* Washington, DC: Joint Committee on Testing Practices.

Linn, R.L. & Miller, M.D. (2005). *Measurement and assessment in teaching* (9th ed.). Upper Saddle River, NJ: Merrill.

Loveless, T. (2012). *The 2006 Brown Center report on American education: How well are American students learning?* Washington, DC: The Brookings Institution.

Porter, A., McMaken, J., Hwang, J., & Yang, R. (2011). Common core standards: The new US intended curriculum. Educational Researcher, 40(3).