



James G. Nicholson
President
Riverside

December 5, 2013

Arizona Department of Education
c/o Ms. Lupita Gomez, Chief Procurement Officer
1535 West Jefferson Street
Phoenix, Arizona 85007

Via e-mail: Procurementinbox@azed.gov

Dear Ms. Gomez:

On behalf of Riverside, the assessment arm of Houghton Mifflin Harcourt (HMH), I am pleased to submit this response to the Arizona Department of Education's (ADE) Request for Information (RFI) #AED14-0003, Standards-Based Competency Assessments, Grades 3–11. Through this RFI, the ADE is seeking information from vendors about criterion-referenced assessments that measure student achievement in reading, writing, and mathematics in Grades 3–11. The assessments must be available for use as the State's standards-based competency assessments in school year 2014–2015. The selected assessment would be used for the State's mandated accountability measures, including Arizona's school accountability system, teacher and principal evaluations, and third grade retention decisions.

Riverside is pleased to present information about the new *HMH-Riverside Next Generation Assessment Suite*. These new assessments are a collaborative project between Riverside and the authors of the *Iowa Assessments™* at The University of Iowa. Currently under development, this assessment suite will be ready to administer in the 2014–2015 school year. The authors from The University of Iowa and Riverside's Test Development team are following our proven and established development processes to ensure that all test items are clearly written and that the tests align with the Common Core State Standards (CCSS). The ADE and Arizona educators can be confident that the *HMH-Riverside Next Generation Assessment Suite* will be technically strong and supported by empirical research data.

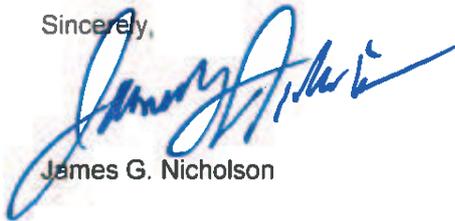
The *HMH-Riverside Next Generation Assessment Suite* is being designed to measure students' understanding of the CCSS in Reading, Writing, and Mathematics. In addition, these new assessments will give educators, students, policymakers, and the public the data they need to identify whether students are on track for postsecondary success, where learning gaps may exist, and how these gaps can be addressed.

Ms. Lupita Gomez
December 5, 2013
Page 2

Educators have used and relied on Riverside's products since 1916; we are dedicated to providing the finest professional testing products and services available. We have distinguished ourselves as a leader in the assessment publishing industry by offering a broad range of high-quality, time-tested products and services supported by a team of highly qualified and creative professionals who are committed to excellence in all areas. Riverside offers a broad portfolio of research-based assessment solutions, serving both the K-12 educational market and the clinical assessment market with print and online assessment tools, including criterion-referenced tests of student performance that have been designed to assess state and national content standards and provide students, teachers, parents, administrators, and all other constituents with useful information about student learning.

We look forward to responding to the ADE's expected Request for Proposal that will result from the RFI. Should you have any questions about this response or need additional information, please contact Robert Lauron, Account Executive, at 949.370.0105 or via e-mail at robert.lauron@hnhco.com; or Bill McCauslen, Regional Vice President, at 512.426.0092 or via e-mail at bill.mcauslen@hnhco.com.

Sincerely,

A handwritten signature in blue ink, appearing to read "James G. Nicholson", with a long, sweeping flourish extending to the right.

James G. Nicholson

I/We The Riverside Publishing Company the undersigned, do hereby submit this response for information with regard to the Standards-Based Competency Assessments in accordance with ARS §41-2555 Request for Information specification contained herein.



James G. Nicholson

Name

December 2, 2013

Date

The Riverside Publishing Company

President

Company Name

Title

3800 Golf Road, Suite 200
Rolling Meadows, Illinois 60008

800.767.8420

Address

Telephone Number

Information Being Requested

1. Background

a. Provide a brief history of the organization and its governance structure.

Riverside is dedicated to providing the finest professional testing products and services available through the delivery of meaningful information for the sole purpose of enhancing the lives of children and adults. We have distinguished ourselves as a leader in the assessment publishing industry by offering a broad range of high-quality, time-tested products and services supported by a team of highly qualified and creative professionals who are committed to excellence in all areas. A charter member of the Association of Test Publishers, Riverside serves both the K–12 educational market and the clinical testing market with a broad range of assessment tools.

A Full Service Assessment Partner

Building on a 95-year tradition of growth, excellence, and service, Riverside offers customers a full array of products and services, including:

- Custom assessments
- Research-based item and test development
- Program management
- Web-based tools
- Scoring and reporting services
- Responsive customer support
- Professional development workshops

Leveraging our capabilities and technical expertise in the areas of project management, test development and publishing, test administration, scoring and reporting, and training and support, Riverside offers customers the following capabilities:

- Extensive experience in developing and implementing large-scale assessment programs
- Demonstrated ability to deliver high-quality, psychometrically sound assessment items and test forms on time to successfully meet customer needs
- Personnel resources, including approximately 330 professionals who are highly trained and experienced in the fields of education, customer support, psychometrics, and management

- Technological resources, including the capability to develop custom assessment software, ranging from item banking technology to content management and test construction systems to benchmark and interim test delivery and reporting systems
- Leadership and innovation in test development and implementation
- Consistent on-time delivery of quality products and responsive customer service for large-scale assessment projects
- State-of-the-art scoring and reporting services

Riverside is part of Houghton Mifflin Harcourt Publishing Company (HMH), a global education company that is committed to addressing education’s dynamic challenges through individualized content, innovative technology, and a holistic approach that inspires achievement and a lifelong love of learning. HMH is the preeminent educational publisher in the United States and the world’s largest publisher of educational materials for pre-K–12 schools. HMH also publishes an extensive line of reference works and award-winning literature for adults and young readers. With origins dating back to 1832, HMH combines its tradition of excellence with a commitment to innovation. For more than a century and a half, the company has shaped ideas, information, and instructional methods into a variety of forms to satisfy people’s lifelong learning needs.

We offer a broad portfolio of research-based assessment solutions, serving both the K–12 educational market and the clinical assessment market with print, CD-ROM, and online assessment tools, including:

- Custom criterion-referenced tests (CRT) that measure student performance using multi-format test items that have been designed to assess state and local content standards and provide students, teachers, parents, administrators, and all other constituents with useful information about student learning;

- Norm-referenced tests (NRT) that enable school personnel to assess the achievement and educational progress of their students compared to students nationally; and
- Benchmark and interim assessments that provide educators with tools to assess student progress toward instructional objectives, allowing for targeted intervention and remediation;
- Diagnostic special-needs assessments for children and adults that allow clinicians and specialists to measure attributes such as general intellectual ability, specific cognitive abilities, scholastic aptitude, and oral language development.

Riverside testing products have been in use since 1916. By employing many talented test

Riverside, an industry pioneer, was the first publisher to provide:

- Narrative reports
- Local norms
- Frequency distributions
- Combined achievement/ability reports
- Group item analyses
- Machine-scorable booklets

development professionals, as well as establishing relationships with many eminent authors in the educational and psychological testing fields, we currently publish more than 100 different assessments and related support materials.

In addition to our longstanding reputation for high-quality assessments, Riverside has a tradition of

innovation directed at enabling our customers to use and interpret tests more effectively. Our company was the first test publisher to provide narrative reports, local norms, frequency distributions, combined achievement/ability reports, group item analyses, and machine-scorable booklets.

In keeping with this tradition of innovation, our company has diversified over the years to become a major provider of custom assessments, web-enabled assessment tools, and software systems for item banking, local scoring of tests, paper-to-web scanning and scoring, data management, and reporting. Riverside effectively responds to the focus on standards and assessment in today's classrooms by working with school systems and state

education departments to provide solutions that allow educators to better understand the progress of their students.

To strengthen this focus, beginning in 2005 our company acquired two innovative assessment systems, which have allowed Riverside to provide states and school districts/systems nationwide with the most comprehensive standards-based benchmark and interim assessment solutions available. Both the *Edusoft*[®] *Assessment Management System* and *DataDirector*[™] are powerful but easy-to-use web-based, integrated assessment solutions that schools can use to create and score assessments and view and act on the assessment results quickly to improve student performance.

As a result of these acquisitions, in addition to the full range of quality summative testing products and services we have offered for many years, we now offer our customers, through *DataDirector* and the *Edusoft* system, high-quality benchmark and interim assessment solutions with:

- The ability to gather and manage student assessment information electronically by either administering tests online or through a paper-to-web test delivery system;
- An effective scoring engine with flexible data analysis features;
- Reports that connect assessment results to instruction; and
- Enhanced professional development.

This service expansion illustrates our dedication to maintaining the standard of excellence in educational and psychological measurement that we have established during the past 95 years and to continually seek to improve the usefulness of our products.

In addition to the benchmark and interim products discussed above and the custom solutions Riverside develops to meet the unique needs of states and school districts across the nation, we offer a wide range of highly regarded educational, cognitive, and developmental testing products. Among our flagship products are the *Iowa*

Assessments[™]; the *Gates-MacGinitie Reading Tests*[®] (*GMRT*[®]); the *Cognitive Abilities Test*[™] (*CogAT*[®]); the *Assess2Know*[®] item banks; the *Riverside Early Assessments of Learning*[™] (*REAL*[™]); the *Battelle Developmental Inventory*[™] 2nd Edition (*BDI-2*[™]); and the Woodcock family of clinical assessments.

Riverside Products and Services

Riverside’s reputation as a full-service provider of assessment and informational products and services is based on three elements:

1. A superb content-area development staff that produces the wide array of Riverside’s custom and published assessments;
2. State-of-the-art, user-friendly, technological solutions that help educators use assessment information in their classrooms; and
3. The ability to innovate and quickly adapt to new situations with the flexibility that our customers need for successful assessment programs.

As a publisher, Riverside’s products and services fall into two principal product lines: (1) educational assessments, and (2) clinical/special needs assessments. While the company has established distinct product lines primarily for developmental purposes, all of its products are interrelated, and Riverside supports both product lines with the resources of the entire organization.

Educational Assessments

Riverside’s largest product line is our educational and group-administered assessments. We publish three types of tests: (1) group-administered assessments for large-scale measurement of educational progress that are written by authors with whom we have contracts; (2) benchmark and interim assessments, and (3) internally developed custom summative assessments aligned to customer standards.

Published Group-Administered Assessments

Riverside produces group-administered assessments that provide a broad range of related measures of achievement, performance, ability, and reading to meet the changing needs of today’s schools. These instruments include multiple-choice and constructed-response items, tasks, and prompts that can be used—formally or informally—in a coordinated assessment program that is fully supported by a comprehensive scoring and information system with data management and local scoring software. Foremost among Riverside’s published educational/group products and services are the *Iowa Assessments*, *CogAT*, and *GMRT*. The *HMH-Riverside Next Generation Assessment Suite*, which we describe in this proposal, is our newest group-administered assessments. Additionally, we publish a wide range of assessments and provide related services in the following categories:

- Achievement tests
- Abilities tests
- Early learner tests
- Reading tests
- Spanish language tests

Benchmark and Interim Assessment Solutions

Riverside offers high-quality items that are aligned to state and local standards and that are designed to help customers create powerful, benchmark and interim assessments that can make a positive instructional difference with our *Assess2Know*[®] product. Our *Assess2Know* item banks include tens of thousands of high-quality items in Reading/Language Arts, Mathematics, Science, and Social Studies. These items were specifically built for benchmark and interim assessment and were created using a comprehensive, multi-step development and quality assurance process. *Assess2Know* provides teachers with resources they can use to create valid, reliable assessments that are aligned to local curriculum standards.

As noted earlier, Riverside also offers two innovative test delivery systems, either of which can be used independently or in conjunction with the *Assess2Know* item banks as part of a benchmark and interim assessment system. Both *DataDirector* and the *Edusoft Assessment Management System* let us provide integrated assessment solutions so that educators can create and score assessments and view and act on the assessment results quickly in order to better improve student performance. Both systems are assessment development and management systems—delivered via an Application Service Provider (ASP) model—that allow educators to securely access the systems from any computer with an Internet connection.

While both *DataDirector* and the *Edusoft* system were designed as open content assessment systems that allow customers to continue to use existing or locally developed assessment items and tests, they are also fully integrated with *Assess2Know*. When *Assess2Know* items are used together with one of these two products, the resulting system provides seamless test development and administration, and rapid scoring and reporting.

Riverside also offers benchmark and interim assessment programs on a statewide level. Currently, we provide the online test administration engine for the Georgia Online Assessment System (OAS). For this custom program, we provide the Georgia Department of Education with a complete, ready-to-implement assessment solution through secure web-based test delivery and state-of-the-art technology for automated reporting and analysis of tests. We also supply a full suite of support services, including comprehensive program management, regional training workshops, and help desk services for all user levels.

Custom Summative Assessments

Another component of Riverside’s product and services offerings is the development of summative tests tied to specific customer standards. These custom assessments are criterion-referenced tests that are used to measure students’ learning in terms of state, district, or other content objectives. The primary purpose of these tests is twofold: to provide results that can be used to improve student achievement, and to meet the

requirements of state and federal accountability reporting. Although Riverside develops many of these tests for large-scale programs of state departments of education or school districts, we also create tests for national or targeted state-level use. Examples of recent criterion-referenced tests specifically developed for large-scale assessment use include those in Louisiana, Mississippi, Missouri, New York, and Oklahoma.

Clinical/Special Needs Assessments

Riverside offers a wide variety of individually administered clinical and special needs assessments and associated computer scoring and reporting programs. These measures and services are specially designed for use in psychology, special education, early childhood, speech-language, and ESL/bilingual testing. Of particular note among the many published and distributed clinical and special needs assessments available are the *REAL*, *BDI-2*, and the *Woodcock-Johnson® III (WJ III®)*. Riverside publishes a complete line of related products developed by Dr. Richard Woodcock and his associates, including the innovative *Bilingual Verbal Ability Tests™ (BVAT™)*, available in 18 languages. Other products include the *Woodcock-Muñoz Language Survey®-Revised (WMLS®-R)*, *Woodcock Interpretation and Instructional Interventions Program™ (WIIP™)*, *Bateria III Woodcock-Muñoz®*, and the *Dean-Woodcock™ Neuropsychological Battery (DW™)*.

Comprehensive Assessments Supported by Technology

Customer Friendly Technology

Riverside's technology-based products and services—delivered via the Internet or on CD-ROM—provide educators with user-friendly robust tools for:

- Test administration
- Scoring and reporting
- Data management
- Linking to instructional resources

Riverside uses state-of-the-art technology to support and enhance our products and services. Our ongoing investment in technology is evident in every phase of test development, production, scoring, and report delivery for both internal processes and customer-facing systems.

To manage item and test development, Riverside has designed and implemented a state-of-the-art *Content Management System (CMS)* that substantially increases

the accuracy and efficiency of item production, improves workflow management, and

provides quality control and version control. *CMS* is a fully web-based tool, aimed at facilitating the development process by allowing item writers and editors the ability to submit and review assessment materials electronically in real time. Once items are approved, we use Smart Connection Enterprise technology, combining composition and text editing software programs, to create, manage, and deliver content and to enable interactive material review sessions with customers. At Riverside, we continually review our internal development processes and procedures to ensure that, wherever applicable, new technologies are embraced.

In addition to our longstanding reputation for high-quality assessments, Riverside has a tradition of innovation in scoring and other test support services directed at enabling our customers to more effectively use and interpret tests. Here, too, state-of-the-art technology plays a critical role in our ability to deliver user-friendly assessment reports that are accurate and delivered on time. An innovator in implementing electronic scoring procedures, Riverside began using those services when they first became available in 1955. Since 1971, the Riverside Scoring Service[®] has used optical scanning, data processing, and score reporting procedures that have popularized large-group individualized testing. In 1978, Riverside pioneered the use of the IBM 3800 laser printer to provide students, parents, teachers, and school administrators with a variety of score reports (that had been unattainable with previous technology), including profile narrative reports targeted to teachers and parents, frequency distributions, combined achievement/ability reports, and group item analyses. In our efforts to continue to diversify our scoring and reporting solutions, Riverside now offers, in addition to paper reports, electronic delivery of reports and web-based reporting tools with our *Interactive Results Manager*[™] (*iRM*[™]). Interactive Results Manager, a dynamic reporting service, lets users access a wide array of reports.

The Technology Advantage

At Riverside, the latest technology is embedded in each stage of the assessment process to:

- Ensure highest quality
- Improve process
- Manage content development and workflow
- Improve the customer experience

Beyond improvement of internal processes and adherence to industry best practices, Riverside serves our customers by providing technology solutions and technologically-enhanced products that, in one form or another, enable educators to administer assessments, promptly and accurately score and report results, analyze data, and link results to instruction to maximize benefits derived from tests and improve learning. Our Technology Development group, comprised of more than 100 information technology professionals, has the experience and capacity to develop custom assessment software systems to facilitate item banking, test construction, test delivery, local scoring, paper-to-web scoring, data management, and reporting. With the proven capacity to deliver custom technology solutions, Riverside now has more than 50 technology products and services in our portfolio and can assure our customers that the utmost care and quality will be taken to ensure a fully functional and robust solution is delivered on time.

We first delivered tests via web- and computer-based testing services in 1998. Since that time, many school districts have adopted our online systems to efficiently deliver their large-scale testing programs. Technology also supports our published products. Both the *BDI-2* and the *Qualls Early Learning Inventory*[®] (*QELI*[®]) can be administered using personal digital assistant (PDA) technology, and the *Iowa Assessments*, *CogAT*, and *GMRT* can be delivered electronically with the results available to educators online. *DataDirector* and the *Edusoft Assessment Management System* are web-based platforms that offer the flexibility of administering tests online or through an innovative paper-to-web process that allows educators to administer tests on plain-paper answer sheets that are scored in the school using inexpensive off-the-shelf scanners. After scanning, test results can be viewed online within minutes of administration, permitting educators to create and implement next-day remediation strategies. *DataDirector* also offers school districts a powerful but easy-to-use web-based data warehousing system. Designed to help educators meet the needs of every student and improve student performance, *DataDirector* helps make sense of a huge range of data to provide longitudinal tracking and reporting of student assessment data, student demographic data, and program/event

attendance, as well as professional development activities, paraprofessional activities, student grades and teacher data.

Partnership for Success

School districts and state departments of education nationwide can confidently select Riverside as their partner for all aspects of assessment creation and delivery, including development, research, production, management, scoring, and reporting. We understand the assessment process from both the teacher’s and administrator’s perspective and can provide tests—norm-referenced, criterion-referenced, and benchmark and interim—on paper or electronically. Additionally, we deliver accurate assessment results in paper or electronic formats that are readily understood and that can be used to inform and improve instruction. Finally, we surround these high-quality products and services with dedicated and responsive management and customer support and training. Based on our 94-year tradition of excellence and innovation, Riverside stands ready to help customers meet their evolving assessment needs by bringing together experienced and highly-skilled assessment and education professionals with well-established procedures and leading technology to deliver high-quality assessment programs that produce reliable and valid results.

b. Identify the individuals from the organization that will be working with Arizona officials on all aspects of the assessments’ implementation.

Riverside is a full-service assessment company that has the organizational and technical capability to manage and coordinate a full range of assessment activities. These activities require in-house capabilities that range from program management and technical support to item and test development, research and measurement services, materials design and delivery, scoring and reporting, programming and software development, and conducting professional development workshops. As a result of this wide range of activities, we have assembled all the necessary resources to manage and support large-scale assessment programs. Our staff of approximately 330 professionals possesses a wealth of experience, specialized education, and/or academic training in such fields as customer support, education, psychometrics, information technology, and management. Staff throughout the

organization cooperates on projects to ensure the development of high-quality products that are delivered in the most efficient manner. As a result, Riverside can readily provide the specialized products and services that would be required for the successful implementation of Arizona’s standards-based competency assessments.

Riverside is committed to providing a highly skilled team of professionals to support Arizona in the implementation of its standards-based competency assessments program. Our management, professional, and technical staffs have broad experience in identifying and allocating all needed resources to deliver the products and services for such programs. Based on the final specifications for this program that will be outlined in the subsequent Request for Proposal, we would designate a highly qualified team that has the requisite knowledge, skills, and background to collaborate with the State in delivering a program of this scope and magnitude. The Arizona Department of Education can be assured that the staff members we would assign to this program will be consummate professionals who possess considerable experience in implementing comprehensive assessment programs that are similar to that described for this important new initiative.

Implementation of the *HMH-Riverside Next Generation Assessment Suite* in Arizona would be under the direction of one of our experienced Program Managers, who oversee and coordinate the implementation of all activities performed by each program team and act as the channel for all program communications within Riverside and with the customer and all other outside organizations and stakeholders. This individual is empowered to guide and direct the program team and ensures consistency, accuracy, and timeliness for all project work.

The Program Manager leads a team from across Riverside made up of experienced, effective professionals—including representatives from Technical Support, Customer Service, the Riverside Scoring Center, and training specialists—who will ensure Arizona educators know how to administer the tests and use the assessment administration and reporting systems. In addition, our Arizona Account Executive, Robert Lauron, will be available to provide on-site support to both the Department and Arizona districts.

Dr. Steve Dunbar and Dr. Cathy Welch, the authors of the assessments from The University of Iowa, will also be available to consult with the Department and provide information about the assessments, as needed, throughout any contract for use of the *HMH-Riverside Next Generation Assessment Suite*.

Riverside brings experience gained from serving other large-scale testing programs across the country to the implementation of our new assessment suite in Arizona. Our disciplined approach to managing complex testing programs ensures that all program tasks are accomplished with quality and accuracy, on schedule, and within budget.

2. Overview of assessment

- a. **Describe the assessments, the grades assessed, the subject areas included, and the formative, diagnostic and summative components. Also describe available end-of-course assessments for grades nine through eleven.**

The summative tests in the *HMH-Riverside Next Generation Assessment Suite* will assess English Language Arts (ELA) and Mathematics for students in Grades 3–11. While the fixed-form summative tests are designed as grade-level end-of-year assessments in ELA and for Grades 3–8 in Mathematics, the Suite will include end-of-course tests in Algebra I and II and Geometry.

The test specifications provide the blueprint for test construction, defining the necessary steps and procedures for development. Tables 1–2 below list the domains in the Common Core State Standards (CCSS) that will be assessed by the ELA and Mathematics tests of the *HMH-Riverside Next Generation Assessment Suite*.

Table 1. ELA/Literacy Domains Assessed

Domain	Grade								
	3	4	5	6	7	8	9	10	11
Foundational Skills	•								
Key Ideas and Details	•	•	•	•	•	•	•	•	•
Craft and Structure	•	•	•	•	•	•	•	•	•
Integration of Knowledge and Ideas	•	•	•	•	•	•	•	•	•
Conventions of Standard English/Knowledge of Language	•	•	•	•	•	•	•	•	•
Vocabulary Acquisition and Usage	•	•	•	•	•	•	•	•	•

Note that the passages at Grades 3–8 include informational texts and literature; at Grades 6–11 they include informational texts, literature, and science and social studies texts.

Table 2. Math Domains Assessed

Domain	Grade									
	3	4	5	6	7	8	9	10	11	
Operations and Algebraic Thinking	•	•	•							
Number and Operations in Base Ten	•	•	•							
Number and Operations – Fractions	•	•	•							
Measurement and Data	•	•	•							
Geometry	•	•	•	•	•	•	•	•	•	•
Ratios and Proportional Relationships				•	•					
The Number System				•	•	•				
Expressions and Equations				•	•	•				
Statistics and Probability				•	•	•	•	•	•	•
Functions						•	•	•	•	
Algebra							•	•	•	
Numbers and Quantity							•	•	•	
Modeling							•	•	•	

Item Types

Measuring the depth and breadth of the CCSS requires a balanced and layered approach that incorporates a range of tasks and stimulus materials. Multiple-choice items, such as those used in our current *Iowa Assessments*, are excellent for evaluating student knowledge and understanding of a variety of concepts and content included within the CCSS. However, additional assessment formats are needed to measure those skills that are not easily assessed by these more traditional formats. The intent of increasing the types of item formats in the assessments is to expand and improve the measurement of student understanding and proficiency overall.

The *HMH-Riverside Next Generation Assessment Suite* is designed to mirror the rigor of the CCSS by employing a robust suite of traditional and nontraditional item types, including:

- **Constructed- and extended-response items:** These items challenge students to draw upon higher-order thinking and cognitive processes. For example, they may require the student to solve multi-step mathematics problems or craft an extended response to a writing prompt. A mix of technology-based scoring that uses automated scoring engines and human scoring in which rubrics guide the assignment of scores will be used.
- **Technology-enhanced items:** Typically administered on a computer, these items require students to make use of complex thought processes and responses. By taking advantage of the many features in today’s computer-based technologies, these items can be interactive. They may include unique response interfaces such as hot spots, drag-and-drop, point-and-click, fill-in-the-blank, and graphing; or require students to provide or select multiple responses to a single question. These items are machine-scored.
- **Multiple-choice:** These items are efficient to administer and offer strong technical properties. As evidenced by the items in our current *Iowa Assessments*, multiple-choice items can be written to address varying levels of cognitive complexity. The multiple-choice items in the *HMH-Riverside Next Generation Assessment Suite* measure students’ skills and knowledge at three cognitive levels. This item type will serve as part of the summative assessment that is specifically designed to efficiently measure student learning in terms of the CCSS.

Summative Assessment Content

The ELA assessments will address reading, language, and writing. The following tables present a blueprint for the reading, language, and writing tests within the ELA assessment. Note that the number of items listed below is approximate; the final number of each type of item will be determined as the tests are finalized.

In the reading and language tests, text complexity will be addressed for each passage through quantitative and qualitative measures. Quantitative measures are aspects of text complexity that are unlikely to be evaluated by a subject matter expert reliably, and

therefore computer software is used. The quantitative measure relevant to CCSS passage development is the *Lexile*[®] score.

Riverside and the authors at The University of Iowa are following the guidelines for *Lexile* ranges found on the Common Core State Standards Initiative’s website.

Qualitative measures are best determined by a subject matter expert who can evaluate the use, organization, language appropriateness, and the likely understanding of the target reader. The qualitative measures relevant to CCSS passage development are documented in a passage review checklist and evaluated by a minimum of two independent subject matter experts during the development process.

Table 3. ELA Blueprint—Section 1, Reading

Grade	Multiple-Choice Items	Technology-Enhanced Items	Constructed-Response Items
3	29–31	2–4	3–5
4	30–32	2–4	3–5
5	31–33	2–4	3–5
6	32–34	2–4	3–5
7	33–35	2–4	3–5
8	34–36	2–4	3–5
9	29–31	2–4	3–5
10	29–31	2–4	3–5
11	29–31	2–4	3–5

Table 4. ELA Blueprint—Section 2, Language

Grade	Multiple-Choice Items	Technology-Enhanced Items
3	28–30	5–7
4	31–33	5–7
5	33–35	5–7
6	35–37	6–8
7	36–38	7–9
8	38–40	8–10
9	44–46	8–10
10	44–46	8–10
11	44–46	8–10

The third section of the ELA assessment is the writing test, which will include a single writing prompt. The responses will be scored by two independent scorers and the scores assigned will be summed, so the writing test can yield a maximum score of 12 points.

Table 5. ELA Blueprint—Section 3, Writing

Grade	Extended -Response Items	Maximum Points Possible
3	Persuasive, Narrative	12
4	Explanatory, Narrative	12
5	Persuasive, Narrative	12
6	Explanatory, Persuasive	12
7	Explanatory, Narrative	12
8	Explanatory, Persuasive	12
9	Explanatory, Persuasive	12
10	Explanatory, Persuasive	12
11	Explanatory, Persuasive	12

The *HMH-Riverside Next Generation Assessment Suite* Mathematics tests will be rigorous, assessing what students can do with what they have learned. Items included in the assessments will be carefully selected from the full range of content of the CCSS and will require a range of cognitive skills.

Students will be required to demonstrate their understanding of concepts and procedures, to solve problems, analyze data, and communicate their results. At the high school level, the Mathematics assessments are being designed to assess the indicators as they are organized in Appendix A of the CCSS for Math, which is found at:

http://www.corestandards.org/assets/CCSSI_Mathematics_Appendix_A.pdf.

For high school grades, Riverside and the authors at The University of Iowa are developing and expanding end-of-course assessments in Algebra I, II, and Geometry, which will align with the CCSS.

The following tables present blueprints for the Mathematics assessments. Note that the number of items listed below is approximate; the final number of each type of item will be determined as the tests are finalized.

Table 6. Mathematics Blueprint: Grades 3–8

Grade	Multiple-Choice Items	Technology-Enhanced Items	Constructed-Response Items
3	37	4	3–5
4	42	4	3–5
5	47	4	3–5
6	51	5	3–5
7	56	5	3–5
8	61	5	3–5

Table 7. Mathematics Blueprint: High School End-of-Course Tests

Course	Multiple-Choice Items	Technology-Enhanced Items	Constructed-Response Items
Algebra I	20–25	5–8	6–8
Algebra II	20–25	5–8	6–8
Geometry	20–25	5–8	6–8

Benchmark Assessments

The *Suite’s* computer-adaptive benchmark assessments will span Grades K–11 in ELA and mathematics. The benchmarks can be administered up to three times per year, in a Fall/Winter/Spring battery and will link to the summative assessment. Each test will comprise two parts that will take 40 to 45 minutes. In ELA, there will be 40 to 45 items across all grade levels, of which two to three per level will be technology-enhanced items. In math, there will be 25–35 items per grade, with approximately two to four (about 10 percent) technology-enhanced items.

b. Describe the timeline for the development of the assessments to ensure full implementation by the 2014 – 2015 academic year.

The components of the *HMH-Riverside Next Generation Assessment Suite* are being released for use across 2014 and 2015. In September 2014, computer-adaptive benchmark assessments will be available; summative tests for Grades 3–11 will be available in in Spring 2015 in both paper and online administration modes. In addition, we will start adding new technology-enhanced items to our formative item bank in 2014.

The summative assessments are the result of an extended, iterative process during which draft test materials are developed and administered to national and state samples to evaluate their measurement quality and appropriateness. All materials have been part of an extensive field-testing program.

To obtain an operational item pool of adaptive benchmark items, we will conduct a pair of calibration events—one in Fall 2013 and one in Winter 2014. In addition, a validity study will take place parallel to the field-test event in Spring 2014.

- c. Describe how, and the degree to which, the assessments are specifically aligned to Arizona’s academic standards in mathematics and English language arts (reading and writing), include any alignment studies, if available.**

Alignment to the CCSS is a guiding principle for the development of the *HMH-Riverside Next Generation Assessment Suite*. Because the Arizona standards are based on the CCSS, these new assessments will also align well to Arizona’s standards.

To produce items that are aligned to the CCSS, Riverside and the authors at The University of Iowa follow a well-defined development process that helps to ensure the appropriate balance and representation of content. This process includes the following steps:

- Creation of test specifications that define the content areas and cognitive processes to be measured by the assessments
- Development of test materials (items and scoring rubrics) that measure critical aspects of the CCSS
- Alignment of individual items to the CCSS during the item development process
- Verification of these alignments by focus groups of educators who are actively teaching English language arts and mathematics at the appropriate grade levels

Traditionally, alignment methodologies have been based on content and cognitive definitions by examining an item-to-standard alignment. However, as with any assessment, this process cannot be completed until the item pool or forms have been

completed. We are confident that new and traditional methodologies that address the complex challenge of aligning assessments to the CCSS will appropriately validate the coverage and rigor of the CCSS and the claims and targets defined by the tests.

d. Describe how the assessments’ results can be compared to other states’ criterion-referenced assessments expected to be in use beginning in the 2014-2015 academic year.

Through an analysis conducted by Riverside and the assessments’ authors at The University of Iowa, we have determined that the test blueprints for the *HMH-Riverside Next Generation Assessment Suite* compare favorably in overall item count, test configuration (number and percentage of multiple-choice, technology-enhanced, and constructed- and extended-response items), and cognitive complexity to competing CCSS assessments. Our new suite will assess student achievement and report results in English Language Arts and Mathematics from the same set of standards as do competing assessments, including those from the consortia. In addition, the *HMH-Riverside Next Generation Assessment Suite* will have the advantage of reporting both subject-area scores (e.g., English Language Arts) and domain levels (e.g., Key Ideas and Details). Therefore, policymakers and educators will be able to see how Arizona’s students are performing on the same standards and domains as students taking other states’ assessments.

e. Describe how the assessments are aligned to college / career expectations. Describe the validation process, including the role of post-secondary education in establishing the readiness expectations. Include any alignment studies, if available.

College readiness information can help educators and families determine whether students are on track to successfully complete first-year post-secondary coursework upon graduation from high school, or whether additional remedial coursework and preparation may be necessary. It allows families and educators to monitor student progress from middle school through high school and allows flexibility to determine the appropriate improvement and support strategies for students as they plan for postsecondary education opportunities. This information should be used to inform instruction and enhance learning for students as they prepare for postsecondary education

opportunities. In addition, it can complement existing admission criteria and processes already used by higher education institutions.

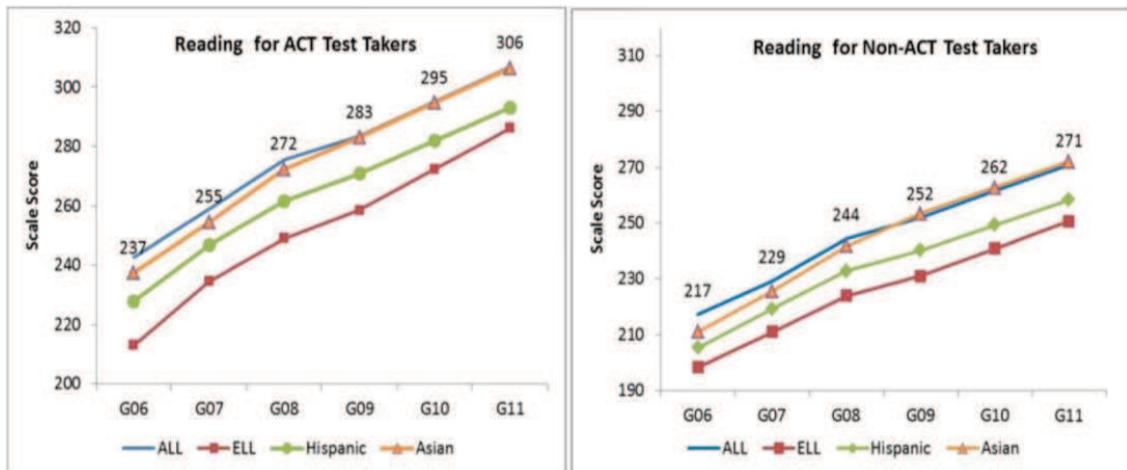
The relationship between performance on the pool of items being readied for the *HMH-Riverside Next Generation Assessment Suite* summative assessments and college admissions test scores was investigated by Furgol, Fina, and Welch (2011)¹ in a matched longitudinal cohort of more than 25,000 students in Grades 5 through 11 who tested annually over a five-year period. These items are on the *Iowa Assessments*, which serve as the basis for the *HMH-Riverside Next Generation Assessment Suite*. Evidence of a strong relationship between scores on the *Iowa Assessments* and the *ACT* composite score suggests that the *Iowa Assessments* and college readiness measures assess very similar—if not the *same*—achievement domains. This relationship sustains itself and strengthens from Grade 5 to Grade 11. Furgol, Fina, and Welch also reported the correlations between *ACT* and *Iowa* subject-area test scores for approximately 18,000 students in Grades 8–11. These correlations are generally high across all grade levels and content areas, providing supporting evidence for the use of the Grade 11 scores to predict whether students are likely to meet or exceed the *ACT* College-Readiness Benchmarks. Note that the correlations between the Grade 11 *Iowa* subject-area tests and the corresponding *ACT* tests are as high as or higher than those between corresponding subject-area tests on *EXPLORE*[®] and *ACT*.

This research will continue following introduction of the *HMH-Riverside Next Generation Assessment Suite*, and the college readiness standards will be statistically linked to the new summative assessments. In addition, as students who have taken the tests complete their high school education and enroll in public (two-year and four-year) and private institutions, researchers at The University of Iowa will validate the predictive validity of the *HMH-Riverside Next Generation Assessment Suite* by tracking student performance throughout their postsecondary experiences.

¹ Furgol, K., Fina, A., & Welch, C. (2011). *Establishing validity evidence to assess college readiness through a vertical scale*. Paper presented at the 2011 Annual Meeting of the National Council on Measurement in Education, New Orleans, LA.

A study by Wang, Chen, and Welch (2012)² examined group differences in the empirical trajectories of performance and established that growth trends for culturally (e.g., Asian and Hispanic) and linguistically diverse (e.g., English Language Learners) test-takers run parallel to the college readiness trajectories identified by Furgol, et al. (2011). All effect sizes for departure from parallel trajectories were extremely small, as suggested by the results shown in Figure 1 below. Such results provide evidence of the appropriateness of using the *NGIA* scale to track the college readiness of all students, including those belonging to subgroups.

Figure 1. College Readiness in Mathematics



a. Describe the available accessibility features, as well as assessment accommodations for individuals with disabilities and English Language Learners.

Students taking the *HMH-Riverside Next Generation Assessment Suite* tests will have access to a number of accommodations, which should be determined by the student’s Individualized Education Plan (IEP). The following table identifies some of the available accommodations by administration mode.

² Wang, M., Chen, K., & Welch, C. (2012). *Evaluating College Readiness for English Language Learners and Hispanic and Asian Students*. Paper presented at the 2012 Annual Meeting of the American Educational Research Association, Vancouver, Canada.

Table 8. Accommodation Supports for Students

Area of Support	Available in Online Administrations	Available in Paper/Pencil Administrations
Presentation of Testing Materials	<ul style="list-style-type: none"> • Audio read-alouds • Magnification • Line Readers 	<ul style="list-style-type: none"> • Read-alouds • Large Print • Braille
Student Interaction with Testing Materials	<ul style="list-style-type: none"> • Highlighting • Scratch Paper • Online Calculator or Personal Calculator 	<ul style="list-style-type: none"> • Highlighting • Scratch Paper • Personal Calculator • Student dictates responses to transcriber
Testing Conditions	<ul style="list-style-type: none"> • Extra time • Flexible configuration of test order, breaks, environment • Individual or group administration 	<ul style="list-style-type: none"> • Extra time • Flexible configuration of test order, breaks, environment • Individual or group administration

b. Describe any practice and/or sample assessments that are available.

Practice assessments will be available for the summative assessments when they are published in 2015. In addition, the benchmark assessments will include similar item types as, and a degree of prediction to, the summative assessments. Therefore, students who take the benchmark assessments will also receive practice in technology-enhanced and constructed-response items that are similar in rigor and content to the items on the summative assessments and they will also have practice using the same online interface that will deliver the online summative assessment.

3. Computer-based assessments

a. If applicable, describe the computer-based option(s) for the administration of the assessments.

The tests in the *HMH-Riverside Next Generation Assessment Suite* can be administered in either a paper-based or online mode, with the decision residing at the district level. This approach gives schools flexibility in their use of technology resources and scheduling so that they can administer tests in ways that meet their needs.

The online system is web-based and will provide all needed security features as well as full support for the test-taking experience. Each student will have a logon and password

to the tests within the online testing system. Teachers will facilitate the administration just as they do with a paper-based administration. Once the student completes the test, the responses are submitted via the online system for scoring.

b. If applicable, describe the technological specifications for the administration of the assessments. This should include specifications for computer hardware, input devices, security requirements, bandwidth, web browser requirements, and platform software.

One key advantage of a web-based system is that it does not require the installation of extensive software locally. It supports current releases of most major operation systems and browsers. The list of supported versions will continue to evolve between now and the introduction of these assessments in the 2014–15 school year.

All devices must have the administrative tools and capabilities to temporarily disable features, functionalities, and applications that could present a security risk and question the integrity of the results. These include, but are not limited to, such functions as unrestricted Internet access, cameras, e-mail and instant messaging, and screen captures.

The table below identifies the system requirements for laptops and PCs, as currently configured. Note that the benchmark assessments can also be administered on a tablet.

Table 9. Online Testing System Requirements

General System Requirements			
Operating System	Web Browser	Software	Screen Resolution
Windows 7	Microsoft Internet Explorer 8 and up* JavaScript must be enabled (Note: for Audio Test Administrations, Internet Explorer 8 also needs Flash 10 or higher installed)	Adobe Reader 9.0 or higher Adobe Flash player for training videos	Minimum: 1024x768 screen resolution
Windows Vista			
Windows XP			
Mac OS X 10.5 Leopard or higher	Safari 5* JavaScript must be enabled	Adobe Reader 9.0 or higher Adobe Flash player for training videos	
System Requirements for Student Testing PC: 1 GHz or faster processor PC and Mac: 1 GB RAM or greater			
System Requirements for Reporting PC: 2 GHz or faster processor PC and Mac: 2 GB RAM or more is recommended for optimal performance			
Internet/Network Requirements High-Speed Internet connection			

* Administrators/educators can also use Firefox 5 or higher on their workstations

c. Describe any available training and technical support that is available for the computer-based assessments.

Riverside will provide a number of training opportunities related to using the *HMH-Riverside Next Generation Assessment Suite*. The following training sessions will be available.

Implementation and Training Planning Meeting—On Site

Participants: Riverside sales representative, District’s main implementation contact, District Testing Coordinators, District IT Representative, District-identified representatives, and HMH Assessment Implementation Specialists

Purpose: The purpose of the Implementation and Training Planning Meeting is to create an implementation plan that identifies assessment dates, milestones, and required actions that need to occur from responsible parties.

Length: 4 hours

Outcome: Implementation and Training Plan

Roster Requirement Meeting—Webinar

Participants: Riverside Technical Assessment Specialist, District main implementation contact, District IT Representative, and District Student Information System Specialist

Purpose: In this session, the Riverside Technical Assessment Specialist will describe and review all system requirements and the layout and field requirements for staff and student files and answer any technical questions.

Length: 1.5 hours

Outcome: District Technical Representatives will understand and be ready to submit the needed roster files to begin use of the system.

Assessment Set Up and Product Training—On Site

Participants: District main implementation contact, District Testing Coordinators, and District-identified representatives

Purpose: The Assessment Set Up and Product Training sessions provide a system navigation overview and best practices for Test Event Set Up; Test Assignment; administration processes; and adding users, locations, and students during a testing window. The session will also include a simulation of the student testing experience. This is a hands-on session that is best suited for 20–25 participants per session.

Length: 3 hours

Outcome: District representatives will be able to access the testing system to create test events, assignments, and sessions. This is the initial implementation step for administration of the assessments via the system.

Reports Interpretation Training—On Site

Participants: District main implementation contact, District Testing Coordinators, Principals, Teachers, and District-identified representatives

Purpose: The Reports Interpretation Training sessions focus on report navigation and generation of assessment reports. Assessment data will be discussed with an emphasis on the CCSS and report interpretation and usage to drive instruction in the district, schools, and classrooms. This is a hands-on session best suited for 20 – 25 participants per session.

Length: 3 hours

Outcome: District representatives will be able to successfully generate and interpret assessment reports in the system. Participants will leave with strategies they can immediately employ in their classroom based on class and student data.

Utilizing Assessment Data—On Site

Participants: District main implementation contact, District Testing Coordinators, Principals, Teachers, and District-identified representatives

Purpose: This session focuses on the reasons for test administration, how the system reports support educators; it includes details on the different scores. Recommendations for using the score report data are shared by the different audience attendees: teachers, administrators, special needs programs staff, and parents. The session will also dive deeper into assessment results with participants leaving the training with strategies they can immediately employ in their classroom. This is a hands-on session best suited for 20–25 participants per session.

Length: 3 hours

Outcome: District representatives will gain a deeper understanding of the assessments, allowing them to better interpret the score reports. Participants will leave with a stronger

knowledge of assessment and strategies they can immediately employ in their classroom based on class and student data.

Technical Support

Riverside offers our customers ready access to technical support through a toll-free phone number as well as e-mail. Our reputation for customer support is outstanding and is one of our company's greatest strengths. All contact with Riverside can be made directly through the Customer Service team via e-mail or a toll-free number. Live technical assistance telephone hours are 5:00 AM to 7:00 PM Mountain Time, Monday through Friday, and messages can be left 24/7, via voice- or e-mail systems. Customer Service team members are the initial contacts for e-mail and telephone support. Generally, our Customer Service team answers all calls immediately. During peak times, incoming calls may occasionally go to voicemail. Voicemail is checked frequently and the customer promptly receives a call back. During off-hours and on weekends, incoming calls and e-mails are typically answered or returned no later than the next business day.

d. If the assessment is computer-based is there a paper / pencil option?

As noted above, the summative assessments will also be available in a paper-pencil format for those schools/classes that do not wish to use the web-based online administration mode.

4. Assessment administration

a. Describe the total anticipated testing time for each assessment (mathematics, reading, and writing) by grade level. If computer-based, include the calculated student to device ratio.

The summative tests require one hour for each test; students will also have approximately 40 minutes to write an essay for the writing test. The benchmark tests are administered in two 40–45 minutes sessions per subject area.

Each student taking either the benchmark assessments or the online version of the summative assessments must have access to a computer that is connected to the Internet. The benchmark assessments can also be administered on a tablet.

b. Identify the anticipated testing window for each assessment by grade level.

The summative assessments are intended for use at the end of the school year; the end-of-course tests are used at the end of the course, whether that is at the end of a semester (if a block schedule is used) or the school year.

The benchmark assessments are designed to be administered two or three times per school year in a Fall/Winter/Spring pattern. Because they link to the summative tests, the Spring test window may use the summative assessment instead of a benchmark test.

c. Describe the training needs and available training for teachers and administrators pertaining to the administration of the assessments.

Please see our response to question c in Section 3 above for descriptions of the training we will offer.

5. Assessment standards setting and scoring

a. Describe the standards setting process.

Riverside will work with each state to set appropriate performance standards for the *HMH-Riverside Next Generation Assessment Suite*, based on state policies and requirements. We can use any of a number of standard-setting methods, including the Bookmark method and the Body of Work method, which have been used with Arizona's assessments. Final plans and a design for the standard setting would be determined in collaboration with the Arizona Department of Education.

Note that the *HMH-Riverside Next Generation Assessment Suite* is being designed so that we can align reporting with the proficiency levels that will be determined for the PARCC and Smarter Balanced consortia's new assessments.

b. If already established, describe the performance levels and the performance level descriptors for each category.

Performance levels and performance-level descriptors have not been established or defined for our new assessment suite. We will work with the Department to write performance-level descriptors as part of the standard-setting planning process.

- c. Describe the score reports available to teachers, students and parents. The description should include:**
- i. How the reports illustrate a student’s progress on the continuum toward college and career readiness, grade by grade, and course by course; and**
 - ii. How the reports are instructionally valuable, easy to understand by all audiences, and are delivered in time to provide useful, actionable data to students, parents, and teachers.**

The summative tests in the *HMH-Riverside Next Generation Assessment Suite* are designed to support aggregate scores in ELA and mathematics for accountability based on proficiency levels. However, that is only the starting point for the information available. Diagnostic information will be provided through scores based on specific “claims” about what students know and are able to do, as well as domain scores that reflect student performance on the major strands of the CCSS. The domain scores will be useful for teachers and schools in evaluating instruction relative to the CCSS.

Other scores for growth and college or career readiness are based on continuing research at The University of Iowa in the measurement of growth and preparation for the content of credit-bearing college courses and for postsecondary training in career-oriented programs. Finally, metrics familiar to parents and policymakers that reference student performance relative to national benchmarks are available for comparative purposes.

Scoring and reporting are integral aspects of the validity of information deriving from an assessment, and the approach for the *HMH-Riverside Next Generation Assessment Suite* is to offer multiple levels of reporting to reflect the likely multiple uses of assessment information. For a summative, accountability assessment, scoring and reporting for proficiency against the ELA and mathematics domains of the CCSS is critical, as is comparability of proficiency-based inferences within-grade and between grades *for all students*. Our new summative assessments will report results in terms of claim scores, domain scores, proficiency levels, growth indicators, college/career readiness, and national comparison (national percentile ranks). *Lexile* and *Quantile*[®] scores will be reported for the ELA and Mathematics assessments, respectively.

The summative assessments in the *HMH-Riverside Next Generation Assessment Suite* will report a variety of metrics to assist various audiences in understanding and using the results. The reporting content categories in the table below will define the domain scores that will be reported.

Table 10. ELA Reporting Domain Categories

Test	Reporting Categories	Grade						
		3	4	5	6	7	8	HS
ELA Part 1, Reading	Total Reading Score	•	•	•	•	•	•	•
	Key Ideas	•	•	•	•	•	•	•
	Craft and Structure	•	•	•	•	•	•	•
	Integration of Knowledge and Language	•	•	•	•	•	•	•
ELA Part 2, Language	Total Language Score	•	•	•	•	•	•	•
	Conventions of Spoken English	•	•	•	•	•	•	•
	Knowledge of Language	•	•	•	•	•	•	•
	Vocabulary Acquisition and Use	•	•	•	•	•	•	•
ELA Part 3, Writing	Total Writing Score	•	•	•	•	•	•	•
Composite	Total Score	•	•	•	•	•	•	•

Table 11. Math Reporting Domain Categories, Grades 3– 8

Reporting Categories	Grade						
	3	4	5	6	7	8	
Operations and Algebraic Thinking	•	•	•				
Number and Operations in Base Ten	•	•	•				
Number and Operations – Fractions	•	•	•				
Measurement and Data	•	•	•				
Geometry	•	•	•	•	•	•	
Ratios and Proportional Relationships				•	•		
The Number System				•	•	•	
Expressions and Equations				•	•	•	
Statistics and Probability				•	•	•	
Functions						•	
Mathematics Problem Solving and Data Interpretation	•	•	•	•	•	•	
Mathematics Concepts	•	•	•	•	•	•	

The benchmark tests will report an overall test score, a percentile rank, and either a *Lexile* or *Quantile* score immediately after the student ends the test.

Score Reports

A number of reports will be provided to meet the needs of various audiences, including teachers, administrators, parents, and students. Reports will be available in our web-based reporting system; they can be printed and saved from that system. Reports will be phased in across 2014 and 2015 and will include:

- Student Roster Report
- Group Roster Report
- Subgroup Roster Report
- Student Proficiency Profile Report
- Group Proficiency Profile Report
- Student Progress Monitoring Report
- Group Progress Monitoring Report
- Multimeasure Student Roster
- Student Report to Families
- Item Response Record
- Group Item Analysis

d. Describe the process and timelines for scoring the assessments. Include computer-based and pencil / paper processes and timelines, as applicable.

Student responses to the summative assessments, whether taken online or on paper, will be scored centrally by Riverside. When the tests are taken online, responses are captured in the online system; responses to paper-based tests are scanned at Riverside's Scoring Center. With online administrations, responses to multiple-choice items and any items that are scored with artificial intelligence (AI) can be scored quickly. Responses to constructed-response items that are scored by human readers require more time. We anticipate that comprehensive results will be available approximately four weeks after receipt of students' responses, either by scanning paper documents or through our online

testing system. Final turnaround time for results will depend on the number of students testing in each administration mode and the number of constructed-response items included in each test at each grade level.

Riverside’s Scoring Process

Riverside’s commitment to producing high-quality products and services is reflected in our ongoing efforts to develop and refine methods that ensure not only the accuracy of our test items, but also the individual and summary data that make up our score reports. Individual student and summary data are examined against a variety of sources as well as cross-referenced to ensure completeness and accuracy. We have instituted a multi-step scoring system that stresses quality, efficiency, flexibility, and security. Documentation and logs detailing the series of checks conducted throughout the data operations process are compiled and can be provided to the State or districts. Handscoring of responses to constructed-response items follows established procedures, including careful training and qualifying of readers and ongoing monitoring of scores and readers to ensure adherence to the scoring rubric.

Before the release of assessment results, the accuracy and completeness of the data that make up the score reports are scrutinized. The quality checks conducted on the final data encompass not only the final output of score reports, but also the production scoring system:

- **Answer keys:** Frequency distributions are produced for actual student response data to verify the accuracy of the multiple-choice answer keys.
- **Scoring and reporting system:** An independent statistical program is used to verify the scoring process.
- **Data used to produce the final reports:** All data elements are checked to ensure that no data are missing, that processing and scoring rules were applied correctly, and that the file layout is correct.

Because the benchmark assessments are computer-adaptive, responses will be scored by the computer as they are entered. Subsequent items will be selected by the system based on each student’s response to the current item. A final score will be computed at the end of the test.

e. Describe how scores on the assessments will be comparable to other common college/career ready assessments.

As previously stated, the test blueprints for the *HMH-Riverside Next Generation Assessment Suite* compare favorably in overall item count, test configuration, and cognitive complexity to other college/career ready assessments. Our new assessment suite will assess student achievement and report results based on the same set of standards as competing assessments, including those from the consortia. We anticipate that the *HMH-Riverside Next Generation Assessment Suite* will provide more metrics/scores than will be available from other common college/career ready assessments. Not only will our tests report domain-level information, but they will also report scores that can be used to compare students’ performance to their peers locally and across the nation.

6. Assessment development

a. Describe how the development of the assessments will adhere to the principles of universal design, so that the testing interface, whether paper- or technology-based, does not impede student performance.

Riverside’s content specialists and the assessment developers at The University of Iowa are experienced in employing the Principles of Universal Design when developing items and tests so that all students have accessibility to the assessments in an equitable manner. This approach better ensures that all students can participate in the assessments. According to the NCEO Synthesis Report 44, there are seven elements of universally designed assessments. These elements are:

- Inclusive assessment population
- Precisely defined constructs
- Accessible, nonbiased items

- Amenable to accommodations
- Simple, clear, and intuitive instructions and procedures
- Maximum readability and comprehensibility
- Maximum legibility

All items for the *HMH-Riverside Next Generation Assessment Suite* are being or will be developed while taking into account these elements. We ensure development of items in accordance with these principles in the following manner:

- We develop items to include a wide array of contexts and cultures. By developing these types of items, we ensure the participation of students with diverse learning needs.
- The test and item specifications serve as a map for precisely defining the constructs that the tests will measure. These specifications indicate to the item writer, content reviewer, and test development specialist exactly what is to be measured. The item could be assessing a particular part of a standard or a combination of elements within a standard. The reviews serve as a method for eliminating items that include assessment of knowledge outside of the standard. For example, mathematics items will have nonmathematical vocabulary below grade level; otherwise the tests would also assess reading ability, decreasing the validity of the measure.
- The review of items will serve to ensure that all items are accessible. Teacher review committees will have an opportunity to review the instructions to ensure that they are appropriate for the grade levels and subject areas of the students.
- Finally, we have experience in developing text, art, tables, maps, and diagrams with maximum legibility. We ensure the recruitment and inclusion of expertise on how text, art, tables, maps, and diagrams may be more easily understood by students with special needs.

b. Describe any comparability studies between the paper/pencil and computer-based assessments.

In the Spring of 2012 and the Spring of 2013, comparability studies with the item pool that is being used to construct the *HMH-Riverside Next Generation Assessment Suite* comparing paper versus an online administration were conducted. The data were collected through a counterbalanced design which required that the same students take both formats of the items but in different sequences to balance any practice effect. Students were randomly assigned (1) paper-and-pencil version first and computerized version second or (2) the computerized version first, paper-and-pencil version second. Approximately 12,000 student responses were obtained for each item in the pool in each administration format.

The study sought to determine the equivalence in both construct and scores (average student performance) across formats. In addition, differences in visual presentation, item position, and complexity of graphics were compared. Analysis included confirmatory factor analysis, differential item functioning statistics, item-level p -value and biserial analyses, and overall differences in performance.

The difference in performance between the two modes was minimal. No evidence was found to suggest any systematic advantage of taking the items in one mode over the other.

c. Describe the processes for item development. The description should include:

- i. How the reading and writing items will require students to demonstrate a range of higher-order, analytical thinking and performance skills in reading, writing and research based on the depth and complexity of the standards, allowing robust information to be gathered for students with varied levels of achievement; and**
- ii. How the mathematics items will require students to demonstrate a range of performance based on the depth and complexity of the standards, allowing robust information to be gathered for students with varied levels of achievement.**

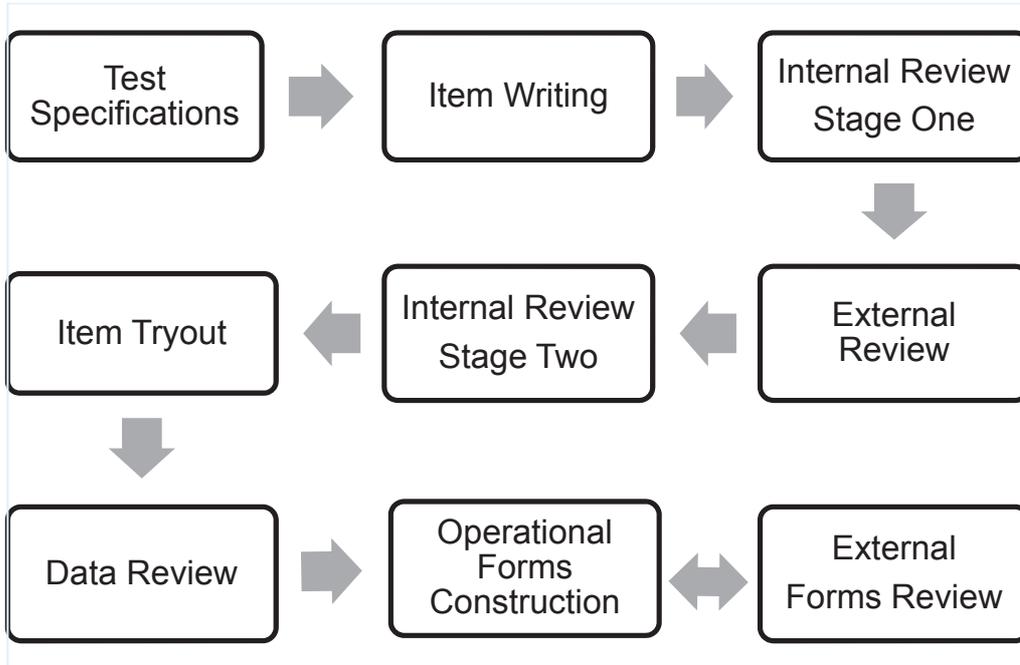
The *HMH-Riverside Next Generation Assessment Suite* has been designed to provide a comprehensive assessment system in mathematics, reading, and writing that is aligned to

the Common Core State Standards with the goal of providing useful, actionable information to students and educators. Key to the design and development processes has been the articulation of the test specifications for this new suite. The specifications of the *HMH-Riverside Next Generation Assessment Suite* have been developed through an extended, iterative process that has included the steps below. Every step is informed empirically through educator input (teacher item writing, review, and scoring) and student performance (evaluation of student responses). The resulting specifications provide a balance of content, a range of cognitive depth and complexity, and a variety of item types that most appropriately measure the content and depth/complexity. Our development steps are:

1. Identify prioritized content for each type of assessment (summative versus benchmark versus formative)
2. Within each assessment, identify the targets for measuring the CCSS
3. Identify the appropriate types of items to measure the range of higher-order, analytical thinking across content areas of the CCSS. These item types are required to elicit evidence of student performance on challenging tasks.
4. Identify the appropriate reporting categories that will inform instruction, monitor growth, and report defensible information on the CCSS. These reports are designed to provide feedback to students and educators that can be used to improve success of instruction and performance.
5. Draft specifications that reflect the appropriate balance of content, range of cognitive complexity, item type, and reporting information.
6. Validate the specifications, item types, and reporting categories through extensive research studies during which draft test materials are developed and administered to national and state samples to evaluate their measurement quality and appropriateness. All materials have been part of an extensive field-testing program.

Specifics concerning the item design and development process are illustrated in Figure 2 and articulated in the text that follows.

**Figure 2. Development Steps for the
HMH-Riverside Next Generation Assessment Suite**



Test Specifications

First, we create test specifications that outline (among other attributes) the statistical requirements for a test; the distribution of content, skills, item formats, and cognitive levels across the test form; the test’s organization; and any special accommodations and other conditions of test administration. Establishing these parameters beforehand helps to ensure the new forms are comparable to existing forms to the degree desired.

For this new assessment suite, comparability and continuity with proficiency expectations and college or career readiness benchmarks are part of the plan for both development and psychometric research. As test development proceeds, the test specifications are continually revisited and evaluated in an iterative process with educator involvement to ensure that the materials available for assembly of final forms reflect the evolving purposes of the assessments.

Item Writing

Items and stimuli/item sets (reading passages, graphs, maps, tables, and so on that support a group of items) are created according to the test specifications. Our content specialists and the authors at The University of Iowa convene item-writing workshops and train educators on sound item writing practices. Of late, such efforts have emphasized best practices in the authoring of unique item types—among them technology-enhanced and both constructed- and extended-response items—that reflect the rigor associated with the Common Core State Standards.

Whether in reading, writing, or mathematics, these items feature a range of cognitive complexity that extends from comprehension (Depth of Knowledge Level 1) through analysis (Depth of Knowledge Level 2) and on into higher-order skills such as application (Depth of Knowledge Level 3). Our test blueprints have been developed with these levels of complexity in mind. As a result, students must demonstrate both a breadth and depth of cognition to successfully answer items.

Further, the adaptive nature of the benchmark tests within our new assessment suite will provide students with tests in which items are selected for each student based on the student’s individual testing performance. For example, a student who answers an item correctly will next receive an item of greater difficulty; likewise, a student who answers an item incorrectly will next receive an item of lesser difficulty. Such adaptive adjustments provide administrators and educators with information about a student’s true level of achievement.

Educators are assigned to write items in the content areas and grade levels that best align with their experience in the classroom. Item production goals ensure a significant surplus of items across subject areas at each cognitive level so that the pool of available items in each subject and at each level is far greater than is needed to build each test. This surplus allows content experts to retain those items which exhibit the characteristics desired.

Internal Review—Stage One

After items are written, content specialists review them for content accuracy, fairness, and Universal Design considerations. The goal of these reviews is to ensure the items are accurate, fair, and accessible to all student subgroups in the diverse population of test takers. The items and associated materials are edited to ensure that they are clearly written and that reading loads are grade appropriate. The items are also copyedited for grammar and spelling at this stage in the process.

External Review

Once the items have been reviewed internally, we convene panels of educators to review the items and associated stimuli (reading passages, tables, graphs, maps, and so on). After a formal training session in the review process, educators review the items for grade-level appropriateness, content relevance, and accuracy. Since they have not been involved in the development process up to this point, external reviewers provide an objective “cold read” of potential test materials. A main goal of the educator review is to confirm that the items are appropriate for the intended grade level and content area. Since the adoption of the Common Core in 2010, approximately 400 teachers have reviewed and aligned tryout items.

Internal Review—Stage Two

Development staff members at Riverside and The University of Iowa review the items again after the educator panel review. This review focuses on edits made to the items during previous steps in the process and again checks for content accuracy, fairness, and Universal Design considerations.

Item Tryout

We collect data on the performance of the items that have successfully passed the review process by conducting a field test to determine how well the items are likely to perform operationally. Students complete the field tests in numbers sufficient to ensure the associated statistical results are sound. Tryout of test materials and alignment to the

CCSS provide data necessary to ensure optimal placement of items for the measurement of growth. Since the adoption of the Common Core in 2010, approximately 432,000 students have responded to tryout items.

Data Review

The data collected during the tryout are analyzed for technical qualities related to item difficulty and discrimination. This analysis determines whether the items are appropriate measures of students' knowledge and the extent to which they will contribute to the test's overall reliability. Other aspects of the data review include key checks and the analysis of distractor choices, subgroup differences, and correlations with operational test forms. Only items that display acceptable descriptive statistics are eligible to appear on operational forms.

Operational Forms Construction

Items that survive data review become part of the pool of items that are eligible for use on operational tests. Procedures for creating the test forms are designed to ensure adequate content coverage in each subject area while being meaningful to students of varying achievement levels, as well as diversity of the items with regards to skill alignment, cognitive level, and difficulty. Careful attention is paid to item selection so that the final tests follow the predetermined test specifications and meet psychometric targets for difficulty, discrimination, and reliability.

Forms Review

Once draft tests have been constructed, they undergo another round of external reviews. Educators are recruited to evaluate the materials from a variety of perspectives, including appropriateness for the intended audience. A second group of experts evaluates materials for perceived fairness and sensitivity concerns. Members of this second group are selected from various ethnic and racial groups, both genders, and from among those who have expertise in the needs of student subgroups such as English Language Learners, students with special needs, and students who are visually impaired.

d. Describe the procedures used to ensure all test items are properly aligned to applicable standards and avoid bias. Include the role of state representatives in these processes.

As described above, the test specifications are written to guide item and test development and to help ensure that the items align with the CCSS. Our test development procedures include steps and processes, including checklists and training, for review of all items and their content for any potential bias. Not only are item writers trained to be aware of and to avoid introducing potential bias, but we also conduct bias reviews, both with outside reviewers and as part of our internal reviews and edits. When the *HMH-Riverside Next Generation Assessment Suite* is finalized, a formal item-to-standard alignment study can be completed to validate the alignment. We would welcome the inclusion of Arizona representatives in this process.

e. Describe the procedures for ongoing assessment item development, including procedures to accommodate any future modifications to the relevant academic standards.

Ongoing research related to assessment and continual content development are key activities for the staff and assessment specialists at The University of Iowa. They monitor and contribute to developments related to standards and assessment, constantly seeking optimal ways to measure student learning and achievement. Those activities support the *HMH-Riverside Next Generation Assessment Suite*. In addition, Riverside will monitor content needs for the benchmark assessments and the formative item bank to ensure that sufficient items are available to support the assessments. New items can be developed and field-tested by embedding them into operational assessments, as needed, to ensure the item pool is robust and adequately covers the standards being assessed.

f. Describe field testing procedures and timelines for the assessments. Include information regarding field testing of Arizona students, if applicable.

As described above, we have included two field-test events in the development of the benchmark assessments: Fall 2013 and Winter/early Spring 2014. Castle Dome Middle School in Yuma Elementary District 1, Yuma, Arizona, participated in the Fall field test.

Additional items will be field tested in the future by embedding them in the operational benchmark assessments. The summative assessment development processes includes ongoing item tryouts.

7. Assessment Cost

a. Describe the pricing structure including the cost per student.

Riverside offers the summative tests from the *HMH-Riverside Next Generation Assessment Suite* in both paper and online formats. The per-student price that includes both the ELA and Math assessments is \$25.00 for paper-based testing and \$20.00 for online testing. That price includes test materials (or online test presentation), scoring, and reporting for a single test administration per student.

The price for the benchmark tests is \$10.00 per student. These assessments are priced on an annual subscription basis, so districts can elect when and how often (up to three times) they will administer the tests throughout the school year. The price for these computer-adaptive assessments includes the online presentation of the tests themselves as well as scoring and reporting.

Riverside is a full-service assessment company that possesses the in-house capabilities to provide a broad range of services and activities needed to support large-scale testing programs. Such services include program management, psychometric services, materials design and delivery, scoring and reporting, software development, and professional development workshops. Prices for any custom services needed to support the statewide use of the *HMH-Riverside Next Generation Assessment Suite* are dependent on the final specifications for the program.