STUDENT ASSESSMENTS AND ASSOCIATED GROWTH MODELS FOR TEACHER AND PRINCIPAL EVALUATION

FORM *C*

PUBLICLY AVAILABLE SERVICES SUMMARY

This form will be posted on the New York State Education Department's Web site and distributed through other means for all applications that are approved in conjunction with this RFQ to allow districts and BOCES to understand proposed offerings in advance of directly contacting Assessment Providers regarding potential further procurements.

Assessment Provider Information	
Name of Assessment Provider:	Houghton Mifflin Harcourt Publishing Company
Assessment Provider Contact Information:	Tim Cooper
Name of Assessment:	The *Iowa Assessments*
Nature of Assessment:	**X** ASSESSMENT FOR USE WITH STUDENT LEARNING OBJECTIVES WITH A TARGET SETTING MODEL; OR **X** SUPPLEMENTAL ASSESSMENT WITH AN ASSOCIATED GROWTH MODEL: ☐ GAIN SCORE MODEL ☐ GROWTH-TO-PROFICIENCY MODEL ☐ STUDENT GROWTH PERCENTILES ☐ PROJECTION MODELS ☐ VALUE-ADDED MODELS X OTHER: Iowa Growth Model
What are the grade(s) for which the assessment can be used to generate a 0-20 APPR score?	Grades 3–12
What are the subject area(s) for which the assessment can be used to generate a 0-20 APPR score?	English Language Arts (Reading, Language) and Mathematics
What are the technology requirements associated with the assessment?	If the tests are administered on paper, no technology is needed. The minimal technology requirements to access score reports in *DataManager*, our web-based reporting system, or to administer the tests online are provided in Appendix D of this submission.
Is the assessment available, either for free or through purchase, to other districts or BOCES in New York State?	**X** YES ☐ NO

Please provide an overview of the assessment for districts and BOCES. Please include: • **A description of the assessment;** • **A description of how the assessment is administered;** • **A description of how scores are reported (include links to sample reports as appropriate);** • **A description of how the Assessment Provider supports implementation of the assessment, including any technical assistance. (3 pages max)**
NOTE TO REVIEWERS: Since this cell would not expand beyond one page, we have supplied the three-page description in Appendix E.

Please provide an overview of the student-level growth model or target setting model for SLOs for districts and BOCES, along with how student-level growth scores are aggregated to the create teacher-level scores, and how those teacher-level scores are converted to New York State's 0-20 metric.

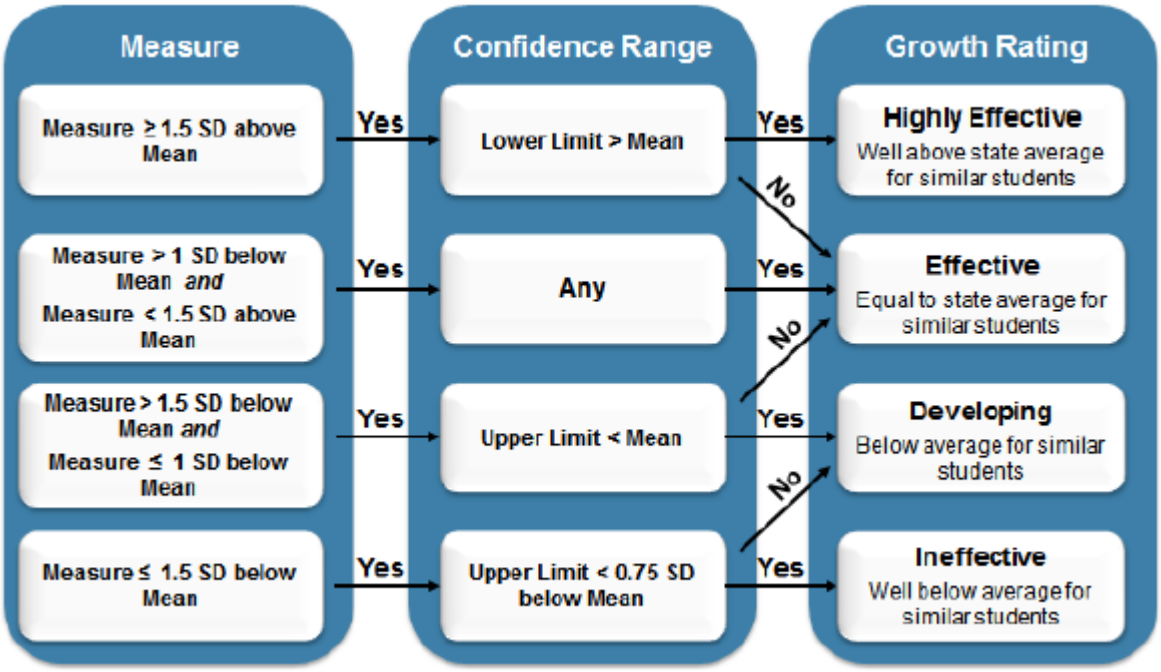
The *Iowa Assessments* provide a straightforward approach for tracking and summarizing growth through the use of a developmental score scale that spans all grade levels. This scale allows the calculation of a single teacher-level score by averaging the growth score for all students associated with a particular teacher. The growth score for an individual student is the difference between the observed score and the expected score (given a student's starting point). The expected score is a growth target for one year.

The essential pieces of this work are already completed and validated on a national scale, including a technically sound and aligned test battery with a vertical scale. The process described below will provide a crosswalk to the HEDI scale for use by New York teachers and principals as one part of their evaluation.

In the state of New York, teacher and principal growth scores represent the average student level score in each classroom and school, respectively. These scores are then converted to New York State's 0-20 Annual Professional Performance Review (APPR) Scale and are divided into four different reporting categories of educator effectiveness: Highly Effective, Effective, Developing, and Ineffective (HEDI). To calculate teacher growth scores, student level growth results are aggregated at the classroom level to determine the effect that the teacher had on students after controlling for prior achievement, English language learner status, students with disabilities status, and poverty status. This kind of model can easily be applied using the *Iowa Assessment* results.

The process to determine an educator's HEDI category using the *Iowa Assessments* is straightforward. First, student level growth scores from the *Iowa Assessments* are aggregated at the educator's level. These scores correspond to the value-added growth scores provided by the New York State Department of Education on the State-provided measures of student growth. In the same manner as the state-provided growth measures, these scores are then converted to New York State's 0-20 APPR scale and the same HEDI rating rules (as in Figure 1 below) are then applied. By converting growth scores to the New York MGP scale and using the same rules, similar proportion of educators should be identified in each of the four effectiveness (HEDI) categories. Accordingly, results from the Iowa Growth Model can be used to differentiate among New York educators and provide meaningful feedback for teachers and principals.

Figure 1. HEDI Rating Rules



New York State Next Generation Assessment Priorities Please provide detail on how the proposed supplemental assessment l or assessment to be used with SLOs addresses each of the Next Generation Assessment Priorities below.	
Characteristics of Good ELA and Math Assessments (only applicable to ELA and math assessments):	The *Iowa Assessments* are large-scale achievement tests that assess students' skills in Reading, Language, and Mathematics. Characteristics of these assessments include content alignment and appropriateness, solid technical characteristics such as validity and reliability, and valuable information being reported to students and educators. The *Iowa Assessments* provide information that can improve instruction and influence student learning. Teachers can use test results to inform parents of an individual student's progress and to evaluate the progress of an entire class. Educators can monitor student growth by comparing results from multiple administrations. The *Iowa Assessments* report student achievement and student growth data, and support inferences about college and career readiness. The *Iowa Assessments* have been empirically validated for each of these purposes. The appropriate supporting documentation for each of these purposes can be found in the *Iowa Assessments Guide to Research and Development*. The *Iowa Assessments* are a research-based and empirically validated and provide information in a fair, reliable and accurate manner. An integral component of the *Iowa Assessments* is the Iowa Growth Model. The Iowa Growth Model provides answers to important questions about student growth and changes to groups over time with a descriptive framework based on many years of research and development associated with the *Iowa Assessments*. Student growth information based on the Iowa Growth Model can be readily used for a variety of purposes in which the primary interpretation involves gain and improvement over time. Growth data based on the Iowa model are also amenable to various approaches for secondary analyses and scores that feed into proprietary methods. The Iowa Growth Model uses an underlying vertical score scale, the National Standard Score (NSS), that permits several approaches to describing growth. It is a metric that ranges numerically from 80 to 400 and spans a developmental continuum from Kindergarten to Grade 12 in major content domains such as reading, mathematics, and written expression. National research studies in the 2010–11 school year were conducted to validate the reference points on the NSS scale representing the medians for each grade level and the model-based inferences about the amount of

	growth typical of students at different achievement levels. The primary interpretations supported by the NSS scale have to do with (1) how much a student is growing from one assessment occasion to the next compared to his or her assessment peers (a relative growth interpretation), and (2) how much growth would be expected for this student's assessment peers (a normative growth interpretation). This basic information about growth can be used for a variety of purposes in student and program evaluation such as individual and group decisions about instructional interventions, and responses to interventions that can be gauged by the amount of growth achieved. Another key feature of the Iowa Growth Model and its backbone, the NSS scale, is the ability to track student growth over time to determine levels of proficiency or to research-based performance benchmarks that indicate college and career readiness. The model defines a longitudinal trajectory that at any given point in a student's educational development can be used to determine whether a student in on track or not to such benchmarks. The performance benchmark for the college and career interpretation of growth is the probability of student success in credit-bearing coursework in postsecondary education.
Assessments Woven Tightly Into the Curriculum:	Educators can use the results of the *Iowa Assessments* to improve instruction in a variety of ways. Teachers can use classroom level results to gauge how well students comprehended particular content areas and adjust instruction accordingly. Growth results can be used for instructional planning and curriculum changes (e.g., in a Response-to-Intervention or RTI framework). This information can be used to set goals for the upcoming school year that will lead to student growth that exceeds expectations. Growth results can also help in determining professional development opportunities.
Performance Assessment:	Not applicable.
Efficient Time-Saving Assessments:	The *Iowa Assessments* provide an effective and efficient assessment of students across a variety of subject areas. Our tests in the *Iowa Assessments* Complete balance efficient test administrations with rich reporting, including at the domain level. Customers who test online receive their results within 24 hours. Digital reporting through the *DataManager* platform is also available for paper-and-pencil testers, providing flexibility in reporting options as well as faster turnarounds.
Technology:	The *Iowa Assessments* offer flexible options for administration through both paper-based and online

	offerings. Customers who test with paper and pencil may elect to receive digital reporting through our *DataManager* platform.
Degree to which the growth model must differentiate across New York State's four levels of teacher effectiveness (only applicable to supplemental assessments):	The essential pieces of this work are already completed and validated on a national scale, including a technically sound and aligned test battery with a vertical scale. What would be required is a process to provide a crosswalk to the HEDI scale for use by New York teachers and principals as one part of their evaluation. In New York, teacher and principal growth scores represent the average student level score in each classroom and school, respectively. These scores are then converted to New York State's 0-20 Annual Professional Performance Review (APPR) Scale and are divided into four different reporting categories of educator effectiveness: Highly Effective, Effective, Developing, and Ineffective (HEDI). To calculate teacher growth scores, student level growth results are aggregated at the classroom level to determine the effect that the teacher had on students after controlling for prior achievement, English language learner status, students with disabilities status, and poverty status. This kind of model can easily be applied using the Iowa Assessment results. The process to determine an educator's HEDI category using the *Iowa Assessments* is straightforward. First, student level growth scores from the *Iowa Assessments* are aggregated at the educator's level. These scores correspond to the value-added growth scores provided by the New York State Department of Education on the state-provided measures of student growth. In the same manner as the state-provided growth measures, these scores are then converted to New York State's 0-20 APPR scale and the same HEDI rating rules are then applied. By converting growth scores to the New York MGP scale and using the same rules, similar proportion of educators should be identified in each of the four effectiveness (HEDI) categories. Accordingly, results from the Iowa Growth Model can be used to differentiate among New York educators and provide meaningful feedback for teachers and principals.

STUDENT ASSESSMENTS FOR TEACHER AND PRINCIPAL EVALUATION

FORM *G*

ATTESTATION OF TECHNICAL CRITERIA – SUPPLEMENTAL ASSESSMENTS WITH CORRESPONDING GROWTH MODELS

Please read each of the items below and check the corresponding box to ensure the fulfillment of the technical criteria outlined in the Technical Application on "FORM B-2".

PLEASE SUBMIT ONE "FORM G" FOR EACH APPLICANT. CO-APPLICANTS SHOULD SUBMIT SEPARATE FORMS.

COMPLETE THIS SECTION:

2.2(A) Narrative Overview of Proposed Supplemental Assessment and Associated Growth Model	
This application contains a short overview of the assessment being proposed, including the intended purpose of the assessment, and how the assessment is administered.	**X**
For supplemental assessments, this application contains a description of the growth model and how it is used in conjunction with the assessment.	**X**
For K-2 assessments, this application contains evidence that the proposed assessment is consistent with this RFQ's requirement that the assessment not be a "Traditional Standardized Assessment" as defined above in the section "Definitions of Key Terms Used in this RFQ."	**N/A**
2.2(B) Evidence of Capability	
This application provides an overview of services provided by the Assessment Provider, including a description of the range of support / technical assistance that the Assessment Provider would provide to an LEA if selected by an LEA for this service.	**X**
This application contains information as to whether the Applicant or Assessment Provider has been denied approval as a provider of assessment services in another state(s) and the reason(s) for such denial. If denied within New York State, the location and reason are indicated.	**X**
2.2(C): Evidence of Copyright Owner/Assessment Representative History of Assessment Development	
This application contains evidence that the Copyright Owner/Assessment Representative has a history of developing assessments of student learning (achievement or growth) for the purpose of making defensible judgments about educator effectiveness.	**X**

2.2(D)-i: Technical Documentation Related to Assessment and Student Growth Score Properties: RELIABILITY ***Both "minimum" and "desired" qualifications are listed. For the purposes of this RFQ, applications will only be rated against the "minimum" qualifications; however, NYSED's aspirational "desired" qualifications are also listed to identify possible future requirements for assessments and associated growth models.***	
For supplemental assessments used in conjunction with growth models: This application contains evidence of the *minimum* criteria for reliability: • Student test scores have adequate levels of reliability (e.g., coefficient alpha > 0.75). This application contains evidence of the *desired* criteria for reliability: • Standard errors provided for students growth scores. • Student growth classifications have adequate decision consistency. • Teacher effectiveness classifications demonstrate adequate consistency. *Examples include agreement statistics (e.g., kappa coefficients) based on simulation studies.*	**Check all that apply:** **X** **X** **X** **X**
2.2(D)-ii: Technical Documentation Related to Assessment and Student Growth Score Properties: VALIDITY – ALIGNMENT ***Both "minimum" and "desired" qualifications are listed. For the purposes of this RFQ, applications will only be rated against the "minimum" qualifications; however, NYSED's aspirational "desired" qualifications are also listed to identify possible future requirements for assessments and associated growth models.***	
For supplemental assessments used in conjunction with growth models: This application contains evidence of the *minimum* criteria for alignment validity: • Evidence that test content is sufficiently aligned with New York State Learning Standards and covers a range of measurable standards. Documentation that demonstrates that: (a) at least 80% of the test measures content aligned with NYS learning standards, (b) no more than 20% of test content is aligned with other learning standards or objectives, and (c) a range of content from the NYS learning standards is measured *Note: Other relevant standards can be proposed if NYS Learning Standards do not apply to subject area.* This application contains evidence of the *desired* criteria for alignment validity: • 100% alignment between NYS Learning Standards and assessment.	**Check all that apply:** **X** ☐
2.2(D)-iii: Technical Documentation Related to Assessment and Student Growth Score Properties: VALIDITY – RELATIONS TO OTHER VARIABLES ***Both "minimum" and "desired" qualifications are listed. For the purposes of this RFQ, applications will only be rated against the "minimum" qualifications; however, NYSED's aspirational "desired" qualifications are also listed to identify possible future requirements for assessments and associated growth models.***	
For supplemental assessments used in conjunction with growth models: This application contains evidence of the *minimum* criteria for validity in relation to other variables: • Evidence students' growth scores are correlated with other measures of student progress (e.g., r > .5 with measures such as the number of objectives mastered by a student over the course of the year, teachers' ratings of	**Check all that apply:**

students' progress, or scores from other assessments). This application contains evidence of the *desired* criteria for validity in relation to other variables: • Evidence teacher effectiveness ratings are positively correlated (e.g., r > .5) with other measures of teaching effectiveness.	**X** **X**
2.2(D)-iv: Technical Documentation Related to Assessment and Student Growth Score Properties: VALIDITY – INTERNAL STRUCTURE ***Both "minimum" and "desired" qualifications are listed. For the purposes of this RFQ, applications will only be rated against the "minimum" qualifications; however, NYSED's aspirational "desired" qualifications are also listed to identify possible future requirements for assessments and associated growth models.***	
For supplemental assessments used in conjunction with growth models: This application contains evidence of the *minimum* criteria for validity of internal structure: • Scale properties appropriate for growth model used (*see notes*). Total scores and subscores on student assessments should be supported by dimensionality analyses (e.g., IRT residual analyses, factor analyses). This application contains evidence of the *desired* criteria for validity of internal structure: • Evidence students' scores are on an interval scale. **Notes: If **gain score model** is used, evidence is needed that students' pretest and posttest scores are on the same scale. If **student growth percentile** model used, justification for the number of years included in the model should be provided. If **growth-to-proficiency**, projection, or **value-added models** are used, evidence is needed that the model explains a significant amount of variability in student achievement. Also, models should demonstrate robustness to missing data.*	**Check all that apply:** **X** **X**
2.2(D)-v: Technical Documentation Related to Assessment and Student Growth Score Properties: UTILITY AND COMPREHENSIBILITY ***Both "minimum" and "desired" qualifications are listed. For the purposes of this RFQ, applications will only be rated against the "minimum" qualifications; however, NYSED's aspirational "desired" qualifications are also listed to identify possible future requirements for assessments and associated growth models.***	
For supplemental assessments used in conjunction with growth models: This application contains evidence of the *minimum* criteria for utility and comprehensibility: • Technical documentation that describes how student growth and educator effectiveness are calculated. This application contains evidence of the *desired* criteria for utility and comprehensibility: • Student growth reports support instructional improvement. Resources and supporting materials available to the field.	**Check all that apply:** **X** **X**
2.2(E)-i: Technical Documentation Related to Aggregating Student-Level Growth Scores to Teacher-Level Scores: CREATION OF TEACHER LEVEL SCORES	

For supplemental assessments used in conjunction with growth models: This application includes a narrative description of how student-level scores are aggregated to create a single teacher-level score for each teacher.	**X**
2.2(E)-ii: Technical Documentation Related to Aggregating Student-Level Growth Scores to Teacher-Level Scores: EXCLUSION RULES	
This application includes a description of any exclusion rules that remove students associated with a given teacher from the teacher's teacher-level score (either through a growth model or in conjunction with an SLO).	**X**
2.2(F): Technical Documentation Related to Converting Teacher-Level Growth Score to New York State's 0-20 APPR Scale	
This application includes a crosswalk that maps scores on the assessment's aggregated teacher-level growth score to the required New York State teacher and principal evaluation metric, which ranges from 0-20.	**X**
This application includes procedures for converting teacher-level growth scores to the 0-20 APPR scale comply with the New York Standards for each evaluation rating category, which are based on the following definitions.	**X**
For supplemental assessments used in conjunction with growth models: This application includes an explanation of the assignment of HEDI rating categories based on the following ranges: • Highly Effective: results are well-above State average* for similar students • Effective: results meet State average* for similar students • Developing: results are below State average* for similar students • Ineffective: Results are well-below State average* for similar students	**X**
2.2(G)-i: Technical Documentation Related to Fairness: TEST TAKERS Consistent with the new Testing Standards (2014), there is an increased focus in the industry on fairness of assessments and their uses. Please provide evidence of fairness for both the proposed assessment and, if applicable, the proposed growth model.	
This application includes evidence that the proposed assessments are fair to all test takers (e.g., Differential Item Functioning [DIF] / bias information, fairness evaluation / sensitivity review plan.)	**X**
2.2(G)-ii: Technical Documentation Related to Fairness: TEACHER GROWTH SCORES	
This application includes evidence of fairness of the proposed aggregated teacher growth scores (e.g., lack of correlation between aggregated teacher growth scores and student demographics).	**X**
The evidence of fairness of the proposed aggregated teacher growth scores includes an explanation of how the growth model incorporates (a) prior academic history, (b) poverty, (c) students with disabilities, and (d) English language learners.	**X**

To be completed by the Copyright Owner/Assessment Representative of the assessment being proposed and, where necessary, the co-applicant LEA:

Houghton Mifflin Harcourt 1. Name of Organization (PLEASE PRINT/TYPE)	 4. Signature of Authorized Representative (PLEASE USE **BLUE** INK)
Shawn Weirather 2. Name of Authorized Representative (PLEASE PRINT/TYPE)	February 18, 2016 5. Date Signed
Senior Director, Business Desk 3. Title of Authorized Representative (PLEASE PRINT/TYPE)	

N/A 1. Name of LEA (PLEASE PRINT/TYPE)	4. Signature of School Representative (PLEASE USE **BLUE** INK)
2. School Representative's Name (PLEASE PRINT/TYPE)	5. Date Signed
3. Title of School Representative (PLEASE PRINT/TYPE)	

FORM C: OVERVIEW OF THE *IOWA ASSESSMENTS*

Featuring a four-color design, flexibility in administration, and a host of other enhancements, the *Iowa Assessments* Forms E and F include content that reflects today's challenging educational landscape, such as the Common Core State Standards (CCSS). The *Iowa Assessments* are designed to provide a thorough assessment of each student's progress in skills and standards that are essential to successful learning and college and career readiness.

The exceptional quality of the tests comes in part from its unique, collaborative development process. The tests were written by researchers from The University of Iowa, who contribute extensive curriculum and measurement expertise, as well as experience gained from managing an ongoing state-level testing program. HMH's staff adds editorial expertise, as well as extensive design, art, and pre-press experience to the development effort. These combined talents have created tests that are widely known, trusted, and used throughout the world. Over the course of a typical school year, the *Iowa Assessments* are administered to nearly 10 million students who are in every state in the nation.

Development of the Tests

As the tests were developed, the *Iowa Assessments*' authors and HMH's Product Development team worked to ensure that test items, which are all multiple-choice, were clearly written and that all distractors challenge students without being "tricky" so that students' academic achievement will be assessed as accurately as possible. Above all, the fundamental guiding principle for the development of the *Iowa Assessments* was to make time spent on testing instructionally useful for all students. For example, passages selected for inclusion in the Reading tests yield good comprehension questions and cover a range of reading interests.

All test items are extensively reviewed and edited by the authors and HMH's staff with the above goal in mind, and the quality and functionality of the items was verified by national item tryouts. Review criteria include age and grade appropriateness; ethnic, racial, and gender balance; good style, grammar, and syntax; and good item characteristics. Furthermore, all items were examined and controlled for potential bias by qualified reviewers to make sure clarity and functionality occurred across subpopulations. The feedback from bias reviewers was thoroughly considered in conjunction with the known performance of the items across the subpopulations. Problem items identified during reviews or tryouts were either rewritten or eliminated.

During the development of test specifications, the test authors worked closely with national experts and educators. Review of local, state, and national guidelines for curriculum in each subject is an ongoing activity of the author team. The *Iowa Assessments* were developed to be consistent with recent shifts in curriculum and instructional practice and to be attuned to current curricular objectives at specified grades. The authors and HMH's team incorporated feedback into their test design from participants in both the preliminary and final item tryouts, from fairness and content review panels, and from exchanges of ideas and information with other curriculum and testing professionals. These processes ensure that tests are developmentally appropriate and aligned with a consensus of up-to-date content standards, including the CCSS.

During the research studies, items were tried out at a span of levels. The research then provided p-values that the authors and HMH's team used to determine the best placement of

items at each grade and test level. Appropriate placement not only ensures that there is an adequate floor and ceiling for every test level, so the majority of students are able to finish the test, but also that the test is challenging enough for high-achieving students to demonstrate what they know and can do. In short, with this measurement approach, accurate data is obtained for the entire achievement continuum, ranging from the least able students to the gifted and talented.

The formats of the *Iowa Assessments* items and directions are based upon applied research that has demonstrated their efficacy in measuring the skills selected for the tests. HMH pays close attention to the format and design characteristics of the test booklets and all related materials. The four-color test booklets incorporate design and layout features that are engaging, developmentally appropriate, and both student- and teacher-friendly at all grade levels. The visual design—including artwork, photographs, font size and style, and page layouts—adheres to the Principles of Universal Design and makes the tests look more like the materials students customarily see in their classrooms every day. Reviewers should note that care was taken, however, to ensure that design elements are not distracting and do not otherwise adversely affect student scores for the sake of appearance.

The *Iowa Assessments*' directions for students and test administrators are also extensively reviewed internally by the authors and HMH's team. The clarity and functionality of the directions were then verified by the administration of the tests during the initial pilot studies, followed by national item tryouts and the national standardization research study.

How the *Iowa Assessments* Are Used

The *Iowa Assessments* answer the needs of schools to provide better ways to measure achievement and impact instruction to help students prepare for college and career. Specifically, Form E can be used to:

- Screen for Response to Intervention (RtI) grouping
- Measure growth from year to year
- Evaluate instructional programs
- Plan for instruction
- Predict success on the state test
- Predict college readiness indicators
- Help districts monitor student learning in terms of the CCSS

How Scores Are Reported

A full range of derived scores, including Grade Equivalents, Percentile Ranks, and Standard Scores that can provide important information about student performance in each content area assessed, is available.

The *Iowa Assessments* offer many specific innovations to support educators' needs. Reporting features include:

- National comparison data for individuals and groups
- Item level data
- Skill level data
- Cognitive level data
- Test total data
- 16 paper reports
- Reporting aligned to CCSS
- Web-based interactive reporting
    - 42 web-based reports
    - Improved speed and performance
    - Greater utility and more flexibility
    - Ability to switch alignments on the fly
    - Select Common Core Skill Domain Alignments
    - Select *Iowa Assessments* Skill Domain Alignments
    - Cognitive difficulty levels reported with other test data to show how students perform within each level, by skill or category
    - Ability to export data to clients' student information system

Implementation and Technical Assistance

A key component of HMH's approach to the management of an assessment program is a comprehensive system of customer support, which includes providing training in the administration of the test and in the use of the results. HMH's regional sales and support staff has extensive experience presenting and supporting both pre- and post-test training workshops for a wide variety of constituencies and groups. They will work with New York LEAs that use the *Iowa Assessments* to determine the nature, purpose, and depth desired for pre- and/or post-test workshops.