

The Research Foundations of Acuity Predictive Assessments in Mathematics and Reading/Language Arts

Prepared By CTB/McGraw-Hill



This document provides an overview of the features and research foundations of the **Acuity**[™] predictive assessments in Mathematics and Reading/Language Arts. To frame our discussion, we first consider accepted definitions of the categories of assessments delivered by **Acuity**. After conducting a review of related literature, we focus our attention on two relevant sources—a 2007 report by the *National Center for the Improvement of Educational Assessment* (Perie, Marion, & Gong, 2007) and the work resulting from the *Council of Chief State School Officers' Formative Assessment for Students and Teachers State Collaborative on Assessment and Student Standards* (FAST SCASS; CCSSO (2007)).

Perie, Marion, & Gong (2007) define formative assessment as follows:

An assessment is formative to the extent that information from the assessment is used, during the instructional segment in which the assessment occurred, to adjust instruction with the intent of better meeting the needs of the students assessed (page 4).

The *Council of Chief State School Officers' Formative Assessment for Students and Teachers State Collaborative on Assessment and Student Standards* (FAST SCASS) group adopted the following definition of formative assessment (CCSSO, 2007):

Formative assessment is a process used by teachers and students during instruction that provides feedback to adjust ongoing teaching and learning to improve students' achievement of intended instructional outcomes.

On their Web site, the FAST SCASS group further defines interim assessment as follows:

Assessments administered during instruction to evaluate students' knowledge and skills relative to a specific set of academic goals in order to inform policymaker or educator decisions at the classroom, school, or district level. The specific interim assessment designs are driven by the purpose and intended uses, but the results of any interim assessment must be reported in a manner allowing aggregation across students, occasions, or concepts.

Based on these definitions, **Acuity** features characteristics of both Interim and Formative assessments that help educators improve student achievement relative to the mandates and objectives of the No Child Left Behind Act (NCLB). NCLB increases educator demand for research-based assessments that support teachers and students in preparing for their state NCLB assessments and guide instructional decision-making at the classroom, school, district, or even state level. **Acuity** supports educators by meeting these demands through interim and formative assessment in Math and Reading/Language Arts for grades 3–8, and Algebra for grades 6–12. Additional **Acuity** modules, including **Acuity** Science, will be available in 2008.

The **Acuity** Math and Reading/Language Arts predictive assessments measure content strictly aligned to related state content standards and deliver informative reports that teachers can use in the classroom to identify student strengths and weaknesses, assign personalized instructional resources, monitor growth, and track expected achievement on the subsequent state NCLB assessment. In addition, this information can be aggregated so it supports district-level instructional and curriculum decision making.



In this paper, we will explicate the research foundations of the **Acuity** predictive assessments in Math and Reading/Language Arts for grades 3–8.

Assessment Designed to Measure Proficiency, Track Progress, Target Instruction, Predict Performance

Acuity moves beyond traditional assessments to help educators make important decisions that can have a significant and positive effect on student achievement levels in schools and across districts nationwide. **Acuity** helps improve student learning with research-based assessments delivered in online and paper-and-pencil formats (many districts use a combination of both) and features predictive assessments, diagnostic assessments, standards-aligned item banks, and instructional resources.

The **Acuity** predictive assessments meet the demand for assessments that track progress relative to the state NCLB assessment. Accordingly, the **Acuity** predictive assessments mirror the look and feel of the state NCLB assessment, align to state standards, and deliver critically important predictions of expected student achievement on subsequent state NCLB assessments. This information can help teachers identify students' strengths and weaknesses, and allows educators to support student achievement through targeted instruction. Importantly, this information also helps shape district- or state-level initiatives and policy decisions.

The **Acuity** diagnostic assessments also track progress and, like the predictive assessments, are aligned to state standards. These assessments are also intended to provide timely assessment of recently taught curriculum and are developed according to state-level pacing guides, scope and sequence, and other available documentation related to common state curriculum.

Acuity supports the creation of custom, state-specific assessments with an item bank of more than 20,000 questions aligned to each given state's standards. Educators can create and share assessments with other educators in their school or district.

Acuity's state-correlated item bank includes assessment items across multiple grades, content areas and standards.

The **Acuity** item bank:

- Includes more than 20,000 state-specific items
- Allows teachers to create their own custom tests by selecting items and test length
- Provides intervention opportunities
- Allows for creation of assessments that are targeted to the particular needs of the students and classes

Instruction is accessible directly from assessment results and can be electronically assigned to a class as a whole, or to individual students, using the “Assign” feature on **Acuity** reports.

Instructional resources are designed to:

- Engage students with interactive activities
- Improve student achievement in key areas
- Intervene in time to make a difference

Each engaging instructional activity provides extra practice and instruction and includes three important components:

- Introduction and explanation related to a given skill
- Guided practice
- Opportunity to evaluate progress throughout instruction



Predicting to State NCLB Assessments

The **Acuity** predictive assessments are comprised of three state-specific forms per grade for Math and Reading/Language Arts. These are labeled A, B, and C, and are illustrated in figure 1.

Each predictive form includes items designed to measure the state content standards in a manner that is proportional to what is measured by the state summative assessment. The **Acuity** predictive assessment item selection strategy results in content that is appropriate for the specific time of assessment. For example, the first predictive assessment (form A) includes some content that explicitly measures standards from the prior year's state assessment, as well as content from the current year, while the third predictive assessment (form C) measures content and standards for the grade level.

A Research-Based System that Measures Growth and Progress

CTB/McGraw-Hill utilizes industry-standard methods to develop, analyze, and score the **Acuity** assessments, including Classical Test Theory (CTT) and Item Response Theory (IRT). These methods are the same scientific, research-based, and documented methods used for most state large-scale assessments.

The three predictive assessments are designed to measure growth and progress toward end-of-year goals. This is accomplished by providing a single, common scale for the predictive benchmark assessments within and across the grades of each content area, and providing predictive information regarding standards measured by the state NCLB assessments.

The single, common scale can be developed using IRT calibration, scaling, and equating subsequent to the first year of the assessment program. These are the same rigorous methods that are applied to most state NCLB assessments. Using this common scale, educators and parents have a powerful tool to monitor and encourage student achievement. A link between the three **Acuity** predictive assessments and the state NCLB assessment will be established subsequent to the first year of administration to develop concordance and prediction tables. These will provide teachers with empirical predictions based on current performance to the subsequent NCLB assessment.

Anchor items link each assessment (form C, for example) with the prior assessment (form B) to create a common scale within and across grades of a content area. In figure 1, the links that will result in a common scale within and across grades are indicated by the dashed lines connecting the forms. The three within-grade forms will be linked to the state summative NCLB test using matched student data. That is, the total scores from the three **Acuity** predictive benchmark assessments in a given grade will be matched to the total score from that grade's state summative assessment results, as indicated by the arrows in figure 1. The scaling and prediction studies will result from data collected in the baseline year.

A common item design, illustrated in figures 2A and 2B, is utilized to support the development of the common scale within and across grades. Common item designs such as this are typically used on state NCLB assessments to equate alternate forms of the assessment. A set that includes approximately 12–15 items will be carried forward from assessment to assessment (form A to form B, form B to form C, and form C to form A of the next grade).

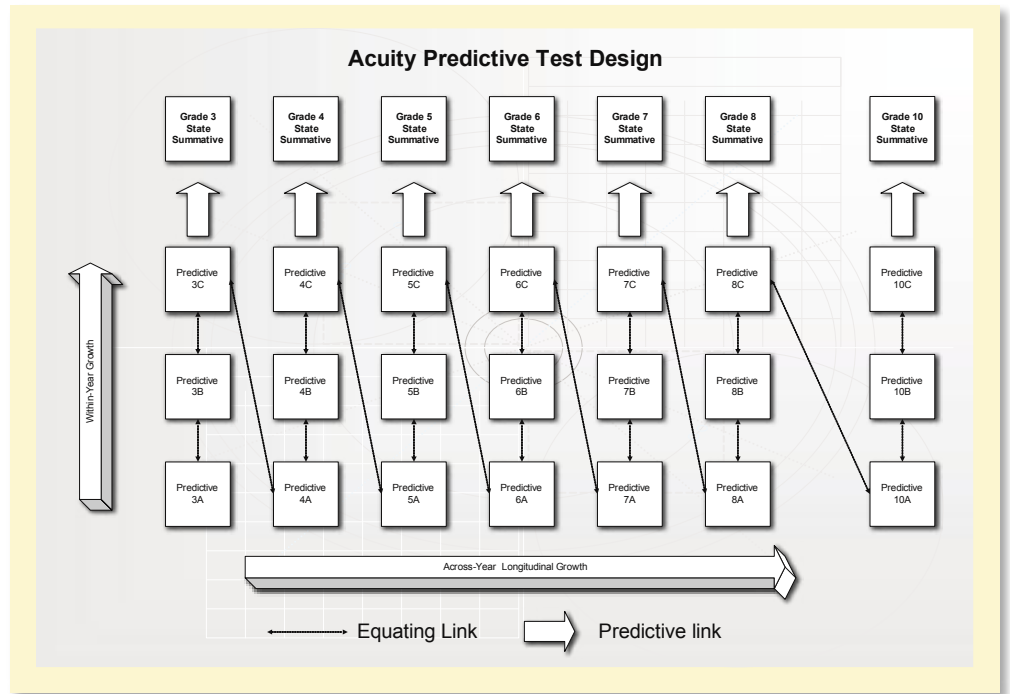


Figure 1 **Acuity** Predictive Test Design

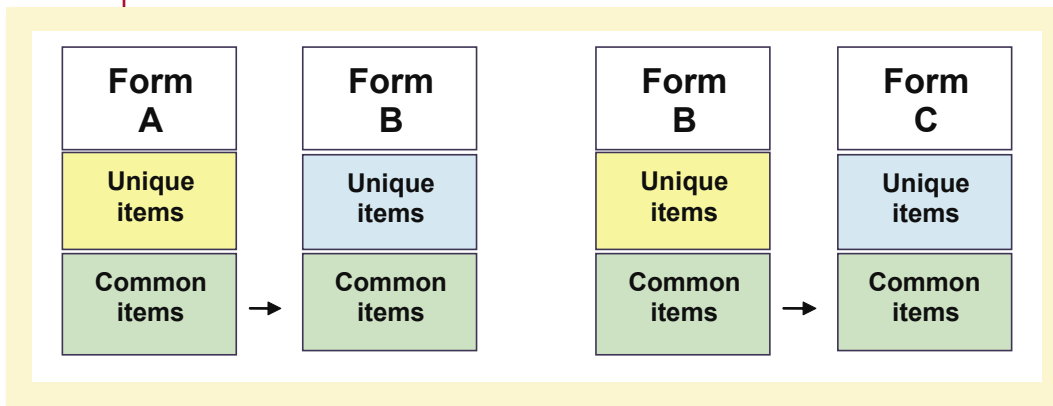


Figure 2A Within-grade Common Item Design

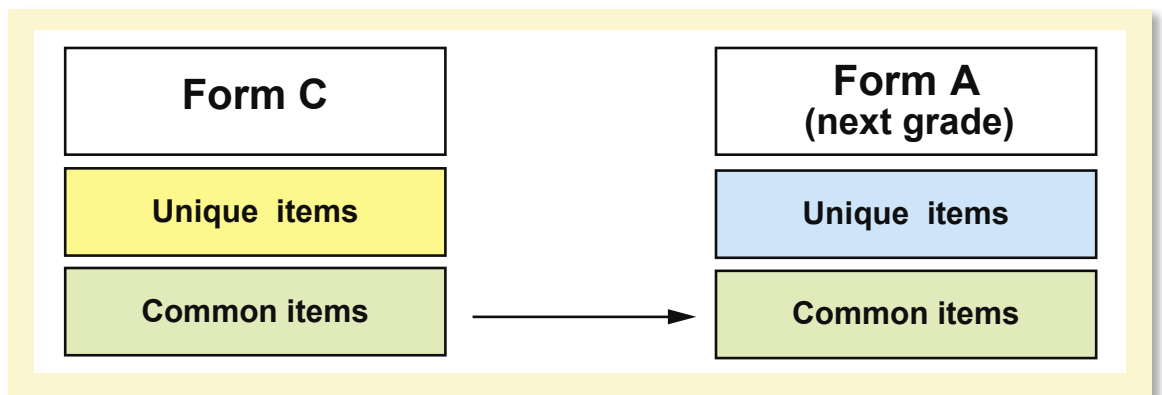


Figure 2B Across-grade Common Item Design



Item Analysis and Scale Development

Items in each predictive assessment will undergo rigorous CTT and IRT test and item analyses. Classical analyses will be used, such as p-values, distractor analyses, point biserial correlations, Mantel-Haenszel differential item functioning (bias indices as sample sizes permit), and test reliability coefficients. IRT analyses will include parameter estimates, model fit indices, checks of local item independence, and checks of the equating quality. Items will be analyzed after each predictive benchmark administration to assess their measurement properties and the final set of selected items will be calibrated using IRT to produce assessments on a common scale. Initial assessments will typically consist of 40 points including overage items used to support final form selections. Other state-specific forms with different numbers of items may be developed.

The common items form an anchor which is used to place the three assessments for each grade and content area on a common scale within and across grades. The IRT calibration, scaling, and equating methods used for **Acuity** are similar to the rigorous methods used to scale and equate the state NCLB assessments; these rigorous methods support the scientific validity of the assessments. CTB has a long history of excellence with respect to scale development and equating and brings that experience to bear on **Acuity** assessments.

Prediction Studies

A variety of analyses and methodologies will be examined to support the prediction studies. The models that best fit the data and provide the most accurate forecasts will be identified and implemented. Some of the models that will be considered are called univariate, multivariate, linear, and curvilinear. Assessment blueprints that reflect the content structure of the state NCLB assessments support the predictive validity of the three predictive benchmark forms. The adequacy of the links will be established and reported both in the technical documentation and in the form of error bands on student reports to indicate the strength of the relationship between **Acuity** and state NCLB assessments. Evidence for predictive validity will consist of an overall association between scores and will be supplemented by “information about the form of the association and the variability associated with that association in different ranges of test scores.” (*Standards for Educational and Psychological Testing*, 1999; Standard 1.15, p. 21).

Common Item Design

The common item design illustrated in figures 2A and 2B supports the development of the common scale, both within and across grades. Common item designs such as this are typically used on state NCLB assessments to equate alternate forms of the assessment.

Under this research design, students in the baseline and subsequent years interact with items in the same context. Specifically, students will see common content between adjacent forms, as this supports the development of scales for measuring growth. Memory effects for assessments such as these are not expected to substantially affect the equating or prediction results, since students in the baseline and subsequent years will interact with the items in similar contexts.

A comparison of students’ responses to the common items on adjacent assessments will provide a direct measure of growth. Some memory effects due to prior exposure may be observed. Simply seeing an item a second time will not provide students with the correct answer, however. Students who miss an item the first time, then respond successfully the



second, will have demonstrated growth; this will be appropriately reflected in the student's number-correct and scale scores. Under this design, the ability to accurately measure growth will be improved if teachers do not teach the specific items to students. Any district can choose to suppress the teaching of specific items by requesting that access to these items be restricted.

The predictive nature of the **Acuity** assessments is not expected to be adversely affected by the common item design because students in the baseline and subsequent years will interact with the items in similar contexts. Any effects to exposure will be accounted for in the prediction studies.

The **Acuity** predictive and diagnostic assessments and the in-depth reports provide educators with information that can support positive changes in instructional programs and impact classroom instruction.

The consistent classroom use of **Acuity** provides a research-based and technology-supported interim assessment program with formative benefits that can help teachers be more efficient and more effective. By engaging students in the assessment and learning processes and providing timely and meaningful information on related progress, new-generation assessment programs like **Acuity** are motivating students—and helping them achieve higher levels of proficiency.

References

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

CCSSO (2007). Formative Assessment for Students and Teachers. Retrieved from <http://www.ccsso.org>

Stocking, M. L., Lord, F. M., (1983). Developing a Common Metric in Item Response Theory. *Applied Psychological Measurement*, 7, 201–210.

Perie, M., Marion, S., & Gong, B. (2007). A Framework for Considering Interim Assessments. *National Center for the Improvement of Educational Assessment*.