The Efficacy of the Acuity Predictive Assessment Research Design: A Review of Empirical Data Studies

Prepared By CTB/McGraw-Hill

ACUITY

Introduction

Acuity™ Predictive Assessments comprise three state-aligned predictive forms—A, B, and C—for reading and math for Grades 3–8, as illustrated in Figure 1. Acuity is a state-aligned product; for a given state, each Acuity form is developed to reflect the content structure of the associated state NCLB test. For instance, the Acuity Predictive Assessment forms—administered approximately six weeks apart and prior to the state NCLB assessment—reflect the content structure and item formats of the subsequent state summative NCLB assessment.

Although each Predictive Assessment form includes content measuring the state standards in similar proportions to that of the state summative assessment, the predictive item selection strategy further supports the coverage of content that is appropriate for the specific time of testing during the school year. For example, the first Predictive Assessment includes content explicitly measuring standards from the prior year state assessment as well as content from early in the current year while the third Predictive Assessment fully measures on-grade content and standards.

Measurement error is decreased when students and forms are appropriately matched by ability and difficulty. This structure was intended to provide Acuity Predictive Assessments of appropriate difficulty. As students increase in ability, the forms also increase in difficulty, so that students and forms are appropriately matched.

The three Predictive Assessment forms per grade are designed to provide a measure of growth and progress toward end-of-year goals. This is accomplished by (a) providing a common scale for the Predictive Assessments within and across the grades of each content area and (b) providing predictive information toward the end-of-year goals measured by the state NCLB assessments.

The common scale is developed using Item Response Theory calibration, scaling, and equating during the first year of the testing program—the same rigorous psychometric methods that are applied to most state NCLB assessments. The common scale will provide students, teachers, parents, and administrators with a powerful tool to monitor and foster student achievement. The relationship established in the first year of administration between the three Acuity Predictive Assessment forms and the state NCLB assessment will be used to develop concordance and prediction tables that will provide teachers with empirical predictions from current performance to the end-of-year criteria.

Each form is on the common scale within and across grades of a content area. The common scale is indicated by the dashed lines connecting the forms in Figure 1. 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 Assessment forms in a given grade will be linked to the total score from that grade's state summative assessment results, indicated by the large arrows in Figure 1. The scaling and prediction studies will be based on the data collected in the baseline year.

Method, results, and discussion

CTB/McGraw-Hill developed and administered state-aligned Acuity forms for a number of Colorado and Missouri districts for Grades 3–8 during the 2007–2008 academic year. Data from the baseline administration contributed to an analysis of the quality of the assessments and selection of final forms. Empirical data from these administrations were analyzed and the results are presented here to support the efficacy of the Acuity forms. There are three forms per grade (A, B, and C) for six grades (3–8) in each content area (language arts and math) in each state (Colorado and Missouri). Each form includes approximately 35 selected-response and two constructed-response items (one 2-point and one 3-point item).

Form reliability and difficulty

Table 1 provides mean proportion correct and reliability information for the administered forms. The case counts range from 2500 to 4500 on each form. All reliabilities are in the mid to high .80s, which are good reliabilities for forms of this length. The average proportion of correct answers is similar for the three forms within a content area in each state. This data indicates that the forms are appropriately targeted, with average proportion correct ranging from .53 to .61, and that the forms did increase in difficulty, as intended, over the course of the year, and that the forms are appropriately matched to student ability.

Table 1. Reliability and Average Proportion Correct

		Average Reliability (Grades 3–8)		Average Proportion Correct (Grades 3–8)	
	Form	**Colorado**	**Missouri**	**Colorado**	**Missouri**
Language Arts	A	.88	.87	.60	.54
	B	.87	.87	.61	.57
	C	.88	.87	.61	.58
Math	A	.84	.85	.55	.53
	B	.85	.85	.58	.56
	C	.86	.85	.60	.56

Common scales within and across grades

In order to track student growth within and across school years, Item Response Theory calibrations, scaling, and equating were conducted to establish a common scale within and across grades. Figures 2–5 (appended) provides tabular and graphical information regarding the mean scale score for each form for Colorado and Missouri in language arts and math. These figures show that growth tends to occur regularly within and across grades, indicating that the Acuity forms measure content that is instructionally sensitive. We note that student growth frequently decreases from Form C of a grade to Form A of the next grade. This reflects the growth trend often referred to as "summer slump," an occasional decline in student growth during the summer vacation.

Table 2 indicates the mean growth within each grade in standard deviation units, also commonly referred to as effect size. Cohen (1992) described .2 as a small effect size, .5 as a medium effect size, and .8 as a large effect size. These results indicate medium to large effect sizes at the lower grades in all content areas and moderate to small effect sizes at the upper grades. This is similar to the growth trends commonly observed in norm referenced

standardized achievement tests. We also note that it is similar to the year-to-year growth measured by the Colorado Student Assessment Program (CSAP) Reading and Math assessments, which are also on a vertical scale, as indicated in Table 3 (Colorado Department of Education, 2006). Generally speaking, Acuity is observed to track student growth within and across grade.

Table 2. Acuity Within Grade Growth

Within Grade Growth Form A to C (standard deviation units/effect size)				
	Colorado		Missouri	
Grade	**ELA**	**Math**	**ELA**	**Math**
3	0.46	0.77	0.52	0.61
4	0.44	0.53	0.10	0.44
5	0.40	0.42	0.32	0.33
6	0.17	0.35	0.18	0.35
7	0.12	0.26	0.17	0.21
8	0.07	0.12	0.19	0.14

Table 3. CSAP Across Grade Growth

Across Grade Growth (standard deviation units/effect size)		
Grade	Reading	Math
3 to 4	0.35	0.40
4 to 5	0.38	0.36
5 to 6	0.19	0.23
6 to 7	0.17	0.26
7 to 8	0.16	0.13

Predictions

Matched data from each Acuity Predictive Assessment and the state NCLB assessment were gathered from a sample of districts in each state. The results, provided in Table 4, indicate average correlations of the Acuity form and NCLB scores between .75 and .80, which are strong associations for tests of this length. This indicates that the predictive relationship between Acuity and the state test is good and that Acuity Predictive Assessments can provide an early indicator of subsequent performance on the state test. Note that the standard error of prediction is identified on Acuity Predictive Reports, and thus, the degree of accuracy is indicated to support appropriate use of the data.

Table 4. Average Correlations Between Acuity Forms and State Summative NCLB Test

Average Correlations Grades 3-8				
	Colorado		Missouri	
Form	**ELA**	**Math**	**ELA**	**Math**
A	0.77	0.78	0.75	0.77
B	0.77	0.8	*	*

* *Study not yet conducted*

Summary

The data support the efficacy of the Acuity research design and implementation. The data indicate that the Acuity forms are reliable, of appropriate difficulty for students, track growth appropriately within and across grades, and support the prediction of subsequent student performance on the state summative NCLB assessment. Additional data will be made available when further analyses are complete.

References

Colorado Department of Education (2006). 2006–07 CSAP Technical Report. Downloaded from http://www.cde.state.co.us/cdeassess/publications.html , October, 2008.

Cohen, J. (1992). A power primer. Psychological Bulletin, 112, 155–159

Figure 1. Acuity Predictive Test Design

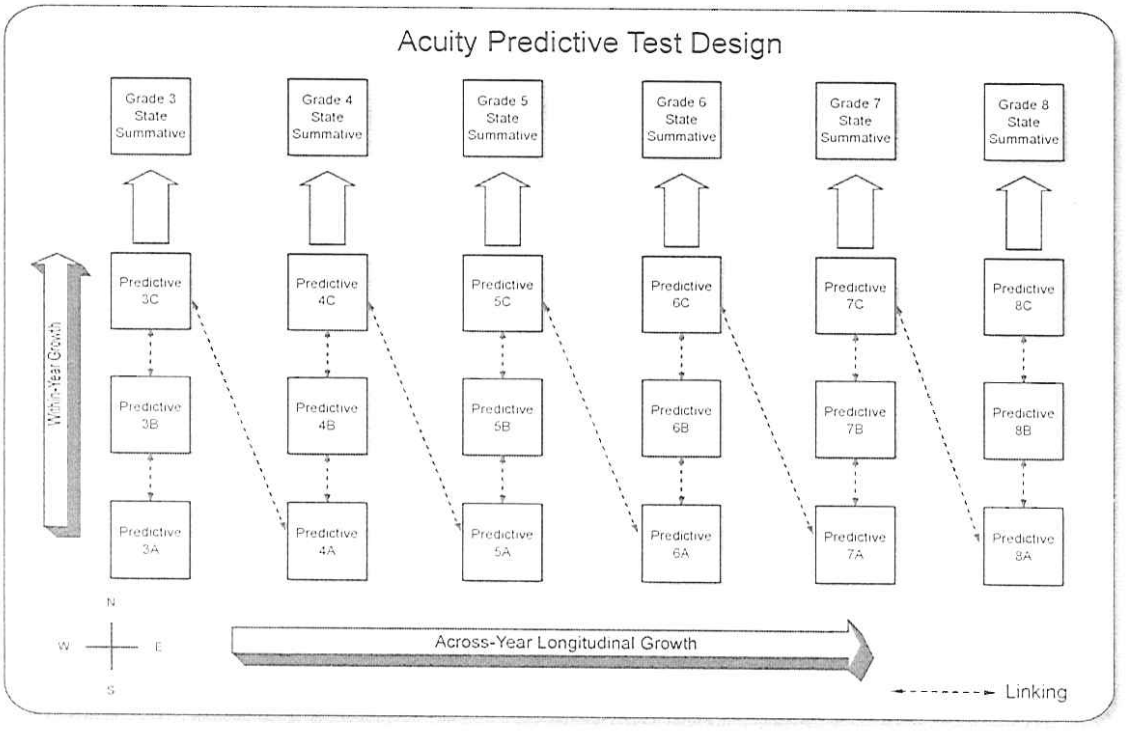

Figure 2. Colorado English Language Arts Growth by Form Across Grades 3–8

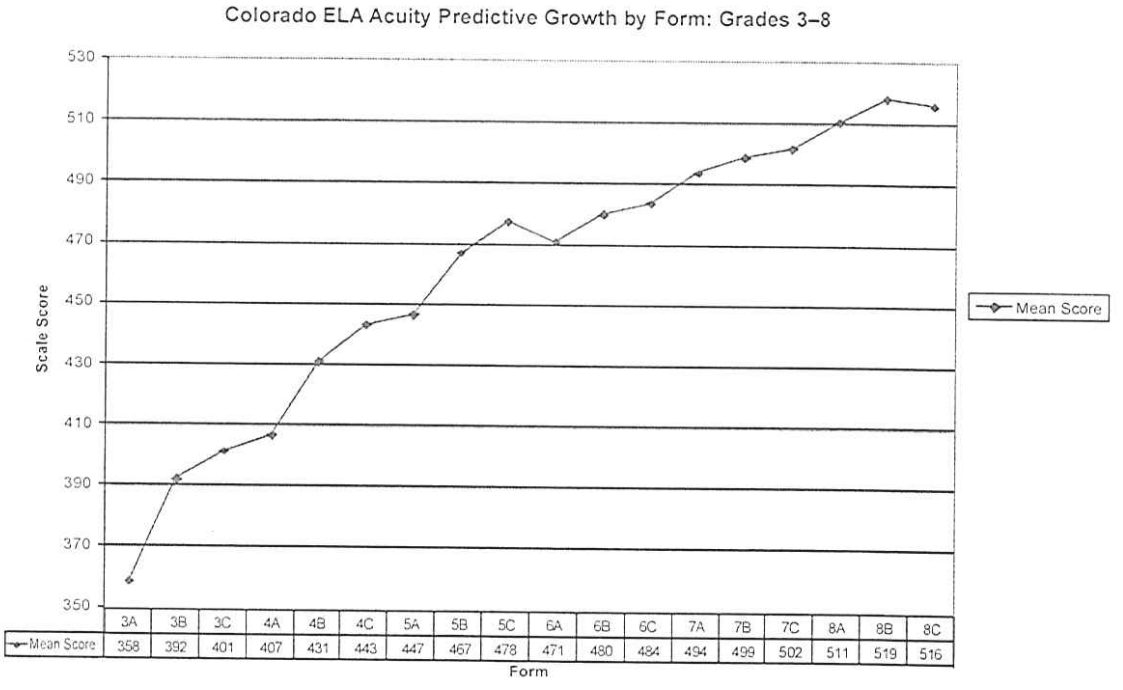

Figure 3. Colorado Math Growth by Form Across Grades 3–8

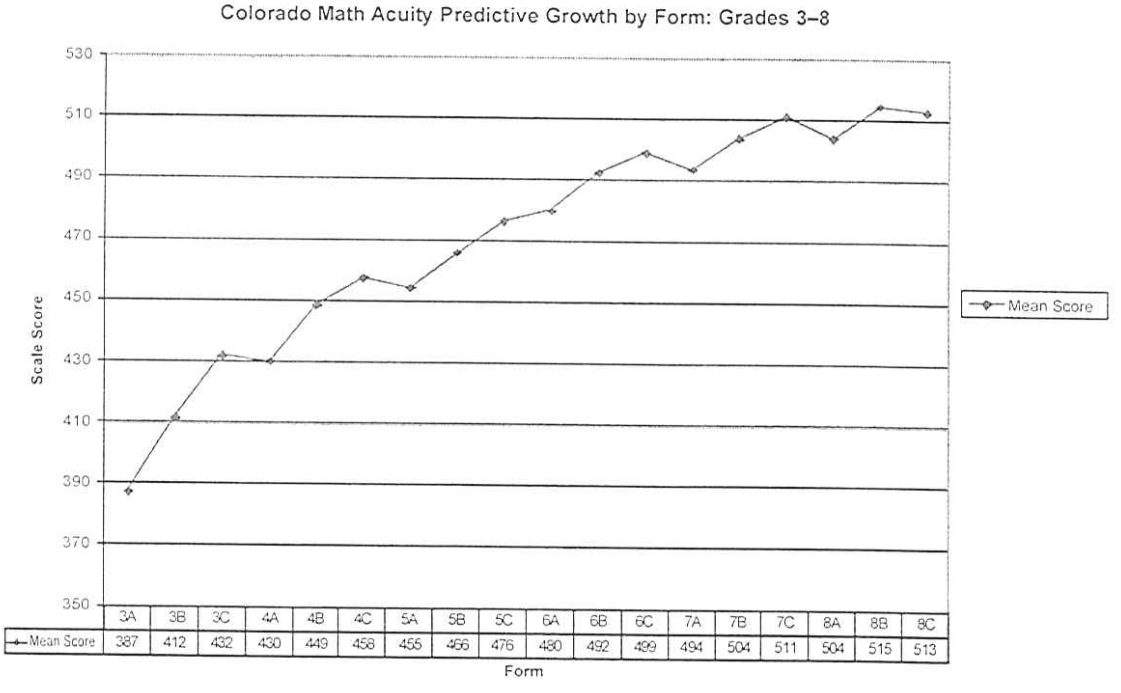

Figure 4. Missouri English Language Arts Growth by Form Across Grades 3–8

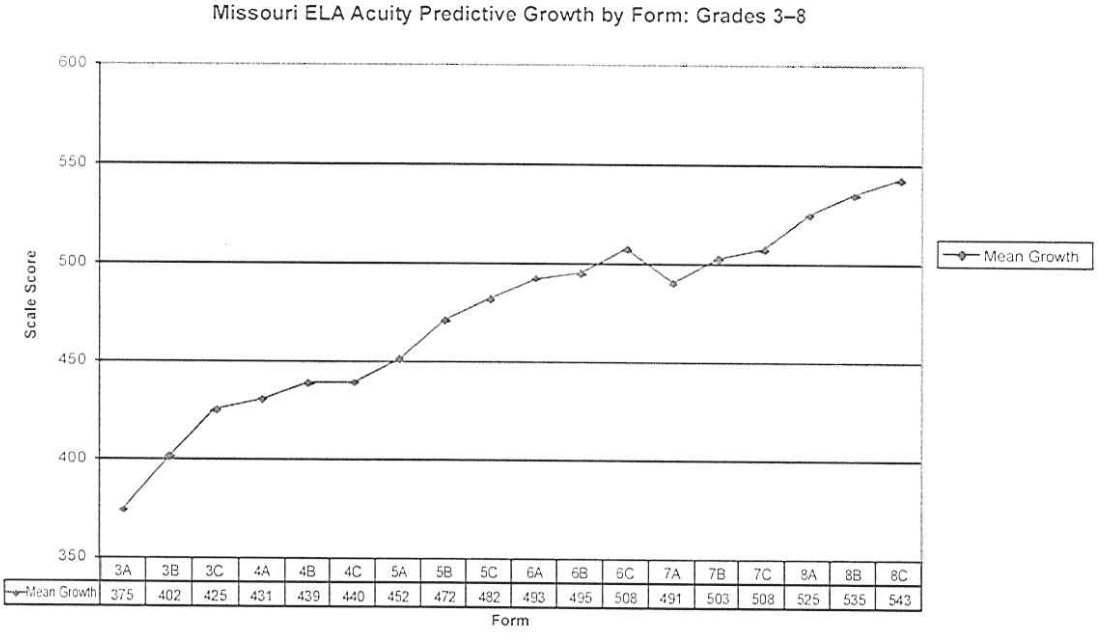

Figure 5. Missouri Math Growth by Form Across Grades 3–8