



Discovery Education Assessment RESEARCH

What is Predictive Assessment?

ALABAMA EXECUTIVE SUMMARY

1. Are Discovery Education Predictive Assessments reliable?

These benchmark assessments are highly reliable. For Grades 2 to 8 and 11 Reading tests over three time periods (Fall, Winter, Spring), the median reliability was .87 with a median sample size of 13,407. The median Mathematics reliability was .87 with a sample size of 13,322. The median Science reliability was .71 with a median sample size of 6,088.

2. Do Discovery Education Predictive Assessments have content validity?

These benchmark assessments model the objectives and skills of the *Alabama Mathematics and Reading Test (ARMT)* and *Stanford Achievement Test, Tenth Edition (Stanford 10)* for Reading, Mathematics, and Science.

3. Do Discovery Education Predictive Assessments match state standardized tests?

Individual student level data necessary to conduct criterion-related validity studies in Alabama have not been made available to Discovery Education Assessment as of 2007-2008. Until such data are released, Discovery Education Assessment refers current and potential Alabama customers to inspect the evidence gathered in previous updates or other states such as Florida and Tennessee.

4. Can Discovery Education Predictive Assessments predict proficiency levels?

Yes, there is a greater than 90% accuracy rate for predicting state proficiency levels. Due to our representation throughout the state of Alabama, direct comparisons between Discovery Test B (Spring 2007) and actual 2007 ARMT proficiency percentages were made for Grades 3 to 8 in Reading and Mathematics. The median Proficiency Prediction Score for Reading was 95%, and the median Proficiency Prediction Score for Mathematics was 96%.

5. Can the use of Discovery Education Predictive Assessments improve student learning?

Many factors contribute to the improvement of student learning. A comparison of improvement in student proficiency from the 2006 ARMT to the 2007 ARMT was conducted for Birmingham City Schools and the state of Alabama. Birmingham City had significant improvement (exceeding the state's improvement during the same time period) in Grades 4, 6, and 8 Reading and in Grades 6, 7, and 8 Mathematics.

6. Can Discovery Education Predictive Assessments be used to measure growth over time?

Yes. These benchmark assessments are scored on a vertical scale using state-of-the-art Rasch psychometric modeling. Thus, reliable estimates of student growth can be made over time.



Discovery Education Assessment RESEARCH

7. Are Discovery Education Predictive Assessments based on scientifically-based research advocated by the U. S. Department of Education?

Two matched control group studies—one in Birmingham, Alabama, and the other in Nashville, Tennessee—support the claim that Discovery Education Predictive Assessments help schools demonstrate significant improvement in student proficiency.



Discovery Education Assessment RESEARCH

What is Predictive Assessment?

ALABAMA

An Overview of Standards and Scientifically-Based Evidence Supporting the Discovery Education Assessment Test Series

Since its inception in 2000 by Vanderbilt University, ThinkLink Learning, now a part of Discovery Education, has focused on the use of formative assessments to improve K-12 student learning and performance. Bridging the gap between university research and classroom practice, Discovery Education Assessment offers effective and user-friendly assessment products that provide classroom teachers and students with the feedback needed to strategically adapt their teaching and learning activities throughout the school year.

Discovery Education Assessment through ThinkLink Learning has pioneered a unique approach to formative assessments using a scientifically research-based continuous improvement model that maps diagnostic assessments to each state's high stakes test. Discovery Education Assessment's Predictive State-Specific Benchmark tests are aligned to the content assessed by each state test allowing teachers to track student progress toward the standards and objectives used for accountability purposes.

Furthermore, Discovery Education Assessment subscribes to the *Standards for Educational and Psychological Testing* articulated by the consortium of the American Educational Research Association, the American Psychological Association, and the National Council on Measurement in Education. This document, "What is Predictive Assessment?", outlines how Discovery Education Assessment addresses the following quality testing standards:

1. Are Discovery Education Predictive Assessments reliable?

Test reliability provides evidence that test questions are consistently measuring a given construct, such as mathematics ability or reading comprehension. Furthermore, high test reliability indicates that the measurement error for a test is low.

2. Do Discovery Education Predictive Assessments have content validity?

Content validity evidence shows that test content is appropriate for the particular constructs that are being measured. Content validity is measured by agreement among subject matter experts about test material and alignment to state standards, by highly reliable training procedures for item writers, by thorough reviews of test material for accuracy and lack of bias, and by examination of depth of knowledge of test questions.

3. Do Discovery Education Predictive Assessments match state standardized tests?

Criterion validity evidence demonstrates that test scores predict scores on an important criterion variable, such as a state's standardized test.



Discovery Education Assessment RESEARCH

4. Can Discovery Education Predictive Assessments predict proficiency levels?

Proficiency predictive validity evidence supports the claim that a test can predict a state's proficiency levels. High accuracy levels show that a high degree of confidence can be placed in the vendor's prediction of student proficiency.

5. Can the use of Discovery Education Predictive Assessments improve student learning?

Consequential validity outlines how the use of these predictive assessments facilitates important consequences, such as the improvement of student learning and student performance on state standardized tests.

6. Can Discovery Education Predictive Assessments be used to measure growth over time?

Growth models depend on a highly rigorous and valid vertical scale to measure student performance over time. A vendor's vertical scales should be constructed using advanced statistical methodologies such as Rasch measurement models and other state-of-the-art psychometric techniques.

7. Are Discovery Education Predictive Assessments based on scientifically-based research advocated by the U. S. Department of Education?

In the *No Child Left Behind Act of 2001*, the U.S. Department of Education outlined six major criteria for "scientifically-based research" to be used by consumers of educational measurements and interventions. Accordingly, a vendor's test

- (i) employs systematic, empirical methods that draw on observation and experiment;
- (ii) involves rigorous data analyses that are adequate to test the stated hypotheses and justify the general conclusions drawn;
- (iii) relies on measurements or observational methods that provide reliable and valid data across evaluators and observers, across multiple measurements and observations, and across studies by the same or different investigators;
- (iv) is evaluated using experimental or quasi-experimental designs in which individuals, entities, programs or activities are assigned to different conditions and with appropriate controls to evaluate the effects of the condition of interest, with a preference for random-assignment experiments, or other designs to the extent that those designs contain within-condition or across-condition control.
- (v) ensures experimental studies are presented in sufficient detail and clarity to allow for replication or, at a minimum, offer the opportunity to build systematically on their finding;
- (vi) has been accepted by a peer-reviewed journal or approved by a panel of independent experts through a comparably rigorous, objective and scientific review;



Discovery Education Assessment RESEARCH

TEST RELIABILITY

1. Are Discovery Education Predictive Assessments reliable?

Test reliability provides evidence that test questions are consistently measuring a given construct, such as mathematics ability or reading comprehension. Furthermore, high test reliability indicates that the measurement error for a test is low. Reliabilities are calculated using Cronbach’s alpha.

Table 1, Table 2, and Table 3 present test reliabilities and sample sizes for Discovery Education Predictive Assessments for three time periods—Fall, Winter, and Spring—in the subject areas of Reading, Mathematics, and Science.

The overall median Reading reliability was .87 with a median sample size of 13,407. The overall median Mathematics reliability was .87 with a sample size of 13,322. The overall median Science reliability was .71 with a median sample size of 6088.

Table 1: Test Reliabilities for Reading, Mathematics, and Science Fall 0708.

	Alabama – Test P					
	Reading	N	Mathematics	N	Science	N
Grade 2	.92	10,688	.86	10,527		
Grade 3	.88	16,797	.87	16,959	.82	7,014
Grade 4	.86	17,214	.89	17,031	.65	6,828
Grade 5	.88	17,202	.87	17,128	.72	10,291
Grade 6	.82	12,600	.86	12,562	.77	6,489
Grade 7	.85	11,615	.84	11,512	.73	7,496
Grade 8	.85	11,817	.87	11,319	.69	6,544
Median	.86	12,600	.87	12,562	.73	6,921

Table 2: Test Reliabilities for Reading, Mathematics, and Science Winter 0708.

	Alabama – Test A					
	Reading	N	Mathematics	N	Science	N
Grade 2	.87	9,818	.83	9,763		
Grade 3	.85	16,379	.86	16,779	.69	6,088
Grade 4	.88	16,499	.87	16,640	.71	6,102
Grade 5	.88	16,161	.84	16,311	.72	9,705
Grade 6	.88	13,718	.87	13,638	.67	6,083
Grade 7	.86	13,005	.86	13,034	.69	7,602
Grade 8	.84	13,323	.84	12,762	.66	5,971
Grade 11	.82	4,844	.73	4,043	.64	463
Median	.87	13,521	.85	13,336	.69	6,088



Discovery Education Assessment RESEARCH

Table 3: Test Reliabilities for Reading, Mathematics, and Science Spring 0708.

Alabama – Test B						
	Reading	N	Mathematics	N	Science	N
Grade 2	.87	9,138	.86	8,814		
Grade 3	.89	16,496	.89	16,815	.69	5,871
Grade 4	.88	17,431	.88	17,472	.72	5,726
Grade 5	.86	17,207	.87	17,233	.74	10,291
Grade 6	.87	13,926	.88	13,536	.73	5,309
Grade 7	.89	12,888	.88	13,107	.69	7,313
Grade 8	.85	12,750	.88	12,103	.68	5,042
Grade 11	.82	2,561	.82	2,286	.71	447
Median	.87	13,407	.88	13,322	.71	5,726



Discovery Education Assessment RESEARCH

CONTENT VALIDITY

2. Do Discovery Education Predictive Assessments have content validity?

Content validity evidence shows that test content is appropriate for the particular constructs that are being measured. Content validity is measured by agreement among subject matter experts about test material and alignment to state standards, by highly reliable training procedures for item writers, by thorough reviews of test material for accuracy and lack of bias, and by examination of depth of knowledge of test questions.

To ensure **content validity** of all tests, Discovery Education Assessment carefully aligns the content of its assessments to a given state's content standards and the content sampled by the respective high stakes test. Discovery Education Assessment hereby employs one of the leading alignment research methodologies, the **Webb Alignment Tool (WAT)**, which has continually supported the alignment of our tests to state specific content standards both in breadth (i.e., amount of standards and objectives sampled) and depth (i.e., cognitive complexity of standards and objectives). All Discovery Education Assessment tests are thus **state specific** and feature **matching reporting categories** of a given state's large-scale assessment used for accountability purposes.

The following reporting categories are used Grades 2 to 8 on Discovery Education Predictive Assessments for Alabama in Reading, Mathematics, and Science. These objectives and reporting categories are based on the *Alabama Reading and Mathematics Test (ARMT)* and the *Stanford Achievement Test, Tenth Edition (Stanford 10)* for Grades 2 to 8 and 11. We continually update our assessments to reflect the most current version of a state's standards.

Alabama Reading Reporting Categories (ARMT)

Decoding/Vocabulary/Word Knowledge	Principles of Grammar
Strategies in Literary and Recreational Reading	Writing Process
Strategies in Textual and Functional Reading	Literary Elements and Devices & Details in Informational Text
Literary Elements and Devices	Elements in Textual/Informational/Functional
Evaluate Impact on Theme	Distinguish Subcategories of Poetry

Alabama Reading Reporting Categories (Stanford 10)

Structural Analysis	Language Expression
Phonetic Analysis: Consonant Sounds	Initial Understanding
Phonetic Analysis: Vowel Sounds	Interpretation
Language Mechanics	Critical Analysis
Expression: Sentence Structure	Expression: Content and Organization
Strategies	



Discovery Education Assessment RESEARCH

Alabama Mathematics Reporting Categories (ARMT)

Number and Operations

Geometry

Measurement

Data Analysis and Probability

Algebra

Alabama Mathematics Reporting Categories (Stanford 10)

Number Sense and Operations

Communications and Representation

Geometry and Measurement

Estimation

Statistics, Probability, and Data Analysis

Math Connections

Algebra, Patterns, and Relationships

Reasoning and Problem Solving

Alabama Science Reporting Categories

Physical Science

Earth and Space Science

Life Science



Discovery Education Assessment RESEARCH

CRITERION VALIDITY

3. Do Discovery Education Predictive Assessments *match* state standardized tests?

Criterion validity evidence demonstrates that test scores predict scores on an important criterion variable, such as a state's standardized test. Scientifically-based research presents evidence that there is a significant correlation between Discovery Education Predictive Assessments and a state test, at the overall test score level and also at a specific skill level. Significant correlations show that high scores on these predictive assessments predict high scores on a state's test.

Discovery Education Assessment has accumulated significant amounts of criterion-related validity evidence for multiple test cycles in several states. Individual student level data necessary to conduct comparable validity studies in Alabama have not been made available to Discovery Education Assessment as of 2007-2008. Until such data are released, Discovery Education Assessment refers current and potential Alabama customers to inspect the evidence gathered in previous updates or other states such as Florida and Tennessee. You may also refer to our Multi-state document featuring a comprehensive reliability and validity overview across several states.



Discovery Education Assessment RESEARCH

PROFICIENCY PREDICTIVE VALIDITY

4. Can Discovery Education Predictive Assessments *predict* state proficiency levels?

Proficiency predictive validity supports the claim that a test can predict a state's proficiency levels. High accuracy levels show that a high degree of confidence can be placed in our test predictions of student proficiency. Two measures of predictive validity are calculated. If only summary data for a school or district are available, the *Proficiency Prediction Score* is tabulated. When individual student level data are available, then an additional index, the *Proficiency Success Rate*, is also calculated. Both measures are explained in the following sections with examples drawn from actual data from Alabama schools.

Proficiency Prediction Score

The Proficiency Prediction Score is used to determine the accuracy of predicted proficiency status. Under the NCLB legislation, it is important that states and school districts help students progress from a "Not Proficient" status to one of "Proficient". The Proficiency Prediction Score is based on the percentage of correct proficiency classifications (Not Proficient/Proficient). If a state uses two or more classifications for "Proficient" (such as "Proficient" and "Advanced"), the percentage of students in these two or more categories would be added together. Also, if a state uses two or more categories for "Not Proficient" (such as "Below Basic" and "Basic"), the percentage of students in these two or more categories would be added together. To see how to use this score, let's assume a school district had the following data based on its annual state test and a Discovery Education Assessment Spring benchmark assessment. Let's use data from a Grade 4 Mathematics Test as an example:

Predicted Percent Proficient or higher = 70%

Actual Percent Proficient or higher on the State Test = 80%

The error rate for these predictions is as follows:

Error Rate = /Actual Percent Proficient - Predicted Percent Proficient/

Error Rate = 80% - 70% = 10%

In this example, Discovery Education Assessment under predicted the percent of students proficient by 10%. The absolute value (the symbols //) of the error rate is used to account for cases where Discovery Education Assessment over predicts the percent of students proficient and the calculation is negative (e.g., Actual - Predicted = 70% - 80% = -10%; absolute value is 10%).

The Proficiency Prediction Score is calculated as follows:

Proficiency Prediction Score = 100% - Error Rate

In this example, the score is as follows:

Proficiency Prediction Score = 100% - 10% = 90%.



Discovery Education Assessment RESEARCH

A higher Proficiency Prediction Score indicates a larger number or percentage of correct proficiency predictions. In this example, Discovery Education Assessment had a score of 90%, which indicates 9 correct classifications for every 1 misclassification. Discovery Education Assessment uses information from these scores to improve its benchmark assessments every year.

Discovery Proficiency Predictions vs. ARMT Proficiency Predictions

Due to our representation throughout the state of Alabama, direct comparisons between Discovery Test B (Spring 2007) and actual 2007 ARMT proficiency percentages were made for Grades 3 to 8 in Reading and Mathematics.

The Proficiency Prediction Scores were calculated via the aforementioned formulas using the combined percentages of “Level III” (Meets Academic Content Standards) and “Level IV” (Exceeds Academic Content Standards). The results for Grade 3 through 8 in Reading and Mathematics are presented in Table 4. The median Proficiency Prediction Score for Reading was 95%, and the median Proficiency Prediction Score for Mathematics was 96%.

Table 4: Proficiency Prediction Scores for Reading and Mathematics.

	Reading Level III & Level IV		Mathematics Level III & Level IV	
	N	Proficiency Prediction Score	N	Proficiency Prediction Score
Grade 3	13,561	95%	12,413	96%
Grade 4	13,033	94%	11,402	91%
Grade 5	11,827	89%	10,758	90%
Grade 6	10,563	95%	9,370	98%
Grade 7	10,235	97%	7,600	96%
Grade 8	9,287	97%	8,542	97%
Median		95%		96%

To illustrate these Proficiency Prediction Scores, Figure 1 - 12 show the percentage of students at each proficiency level for Discovery versus ARMT for Grades 3 through 8 in Reading and Mathematics.



Discovery Education Assessment RESEARCH

Figure 1: Results of Proficiency Prediction Study by Proficiency Level for Grade 3 Reading.

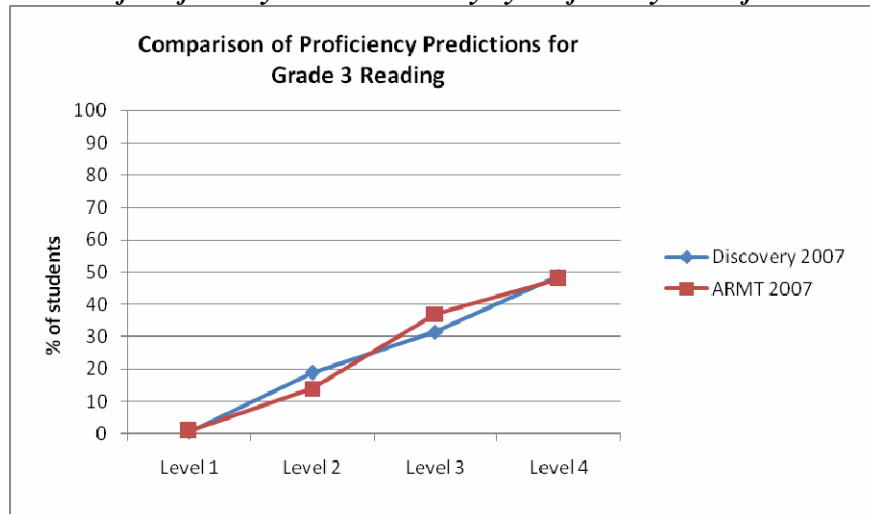
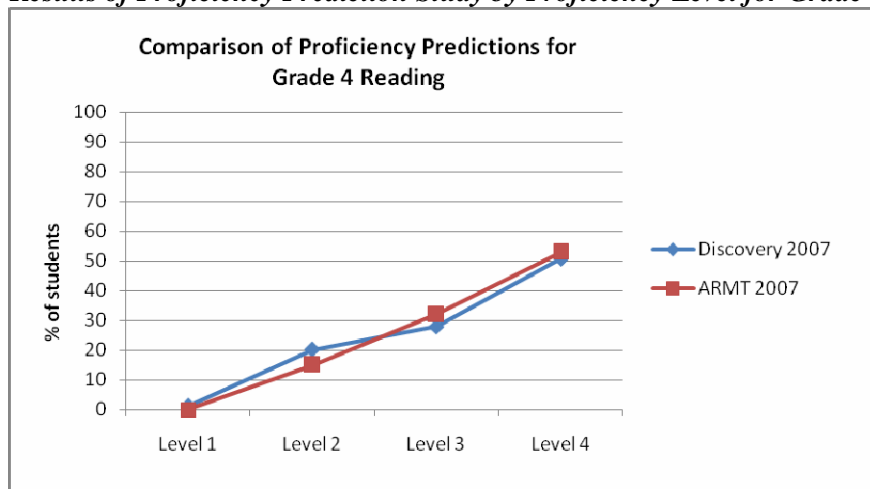


Figure 2: Results of Proficiency Prediction Study by Proficiency Level for Grade 4 Reading.





Discovery Education Assessment RESEARCH

Figure 3: Results of Proficiency Prediction Study by Proficiency Level for Grade 5 Reading.

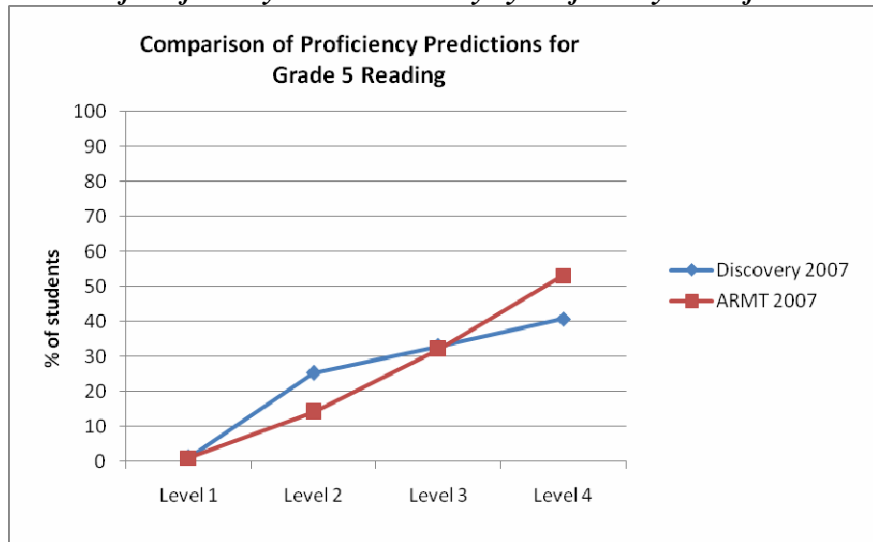
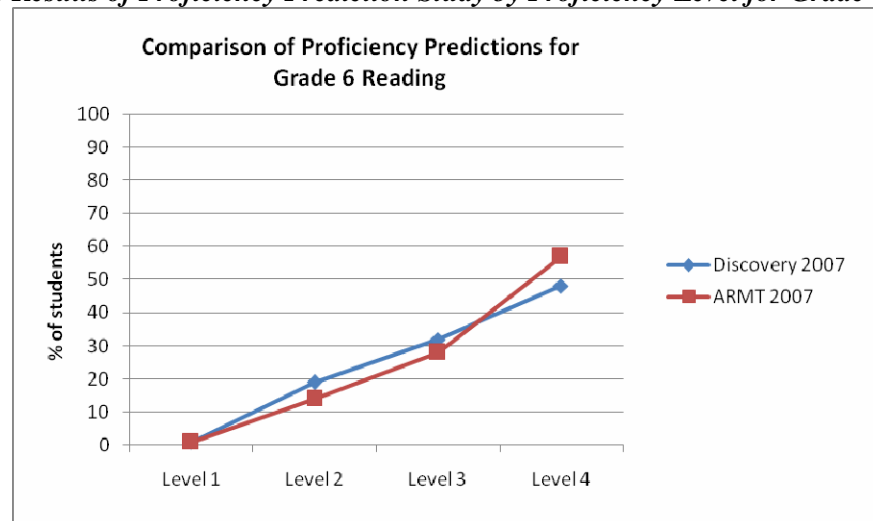


Figure 4: Results of Proficiency Prediction Study by Proficiency Level for Grade 6 Reading.





Discovery Education Assessment RESEARCH

Figure 5: Results of Proficiency Prediction Study by Proficiency Level for Grade 7 Reading.

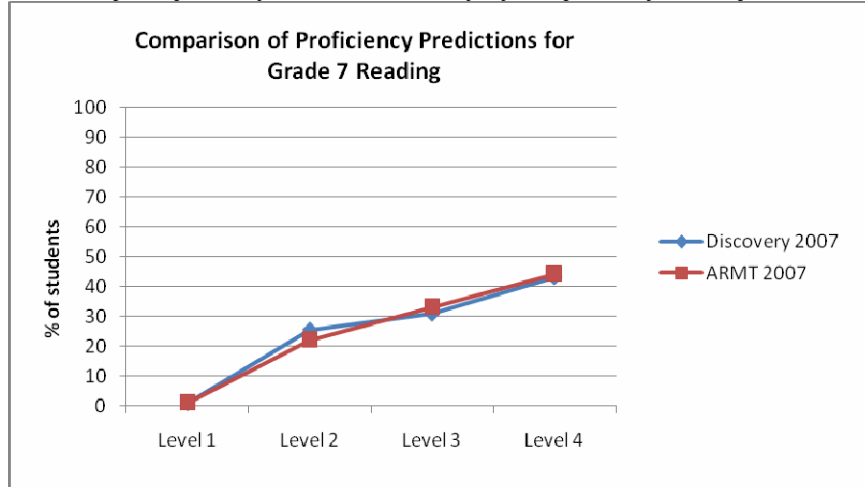
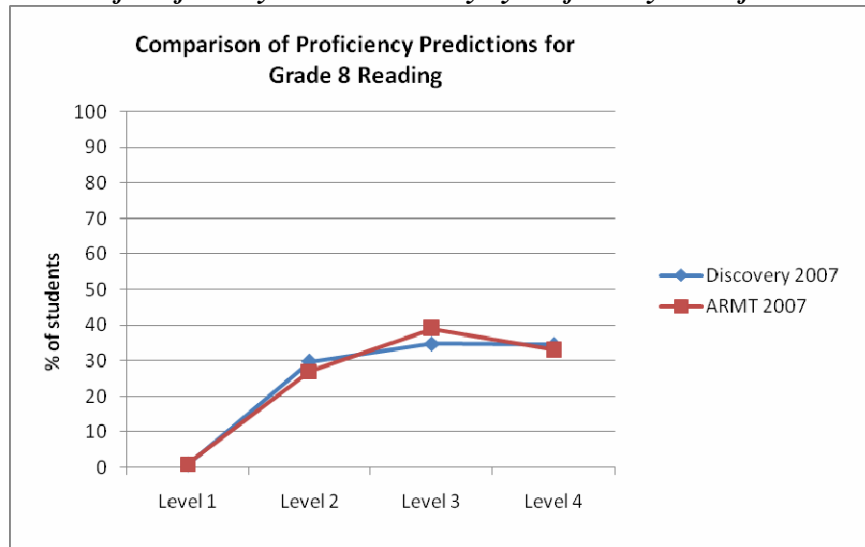


Figure 6: Results of Proficiency Prediction Study by Proficiency Level for Grade 8 Reading.





Discovery Education Assessment RESEARCH

Figure 7: Results of Proficiency Prediction Study by Proficiency Level for Grade 3 Mathematics.

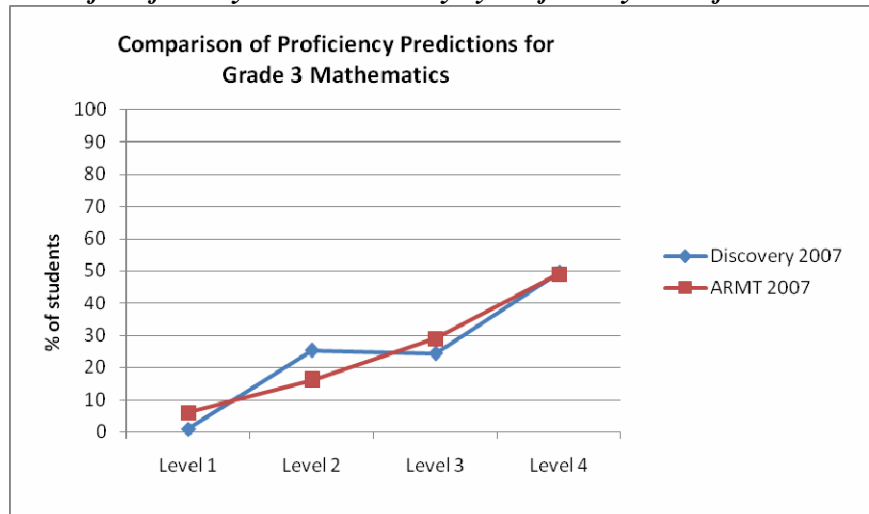
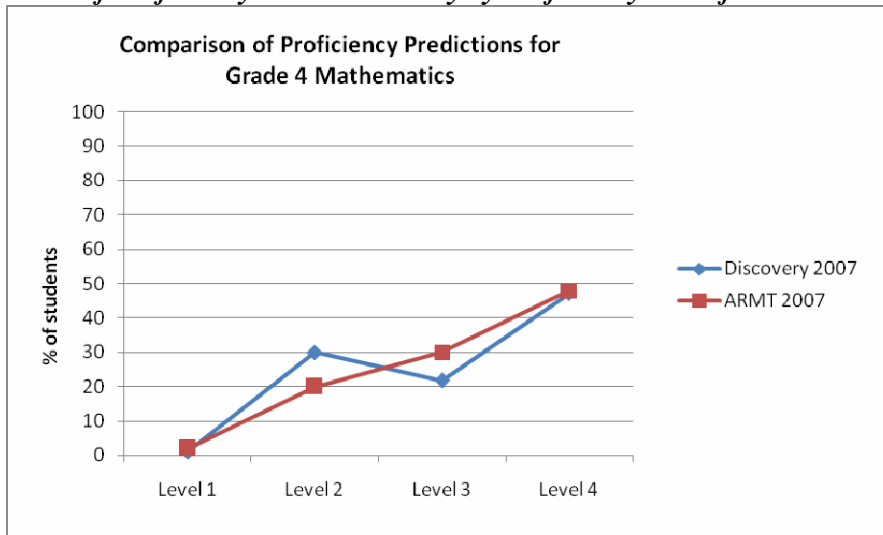


Figure 8: Results of Proficiency Prediction Study by Proficiency Level for Grade 4 Mathematics.





Discovery Education Assessment RESEARCH

Figure 9: Results of Proficiency Prediction Study by Proficiency Level for Grade 5 Mathematics.

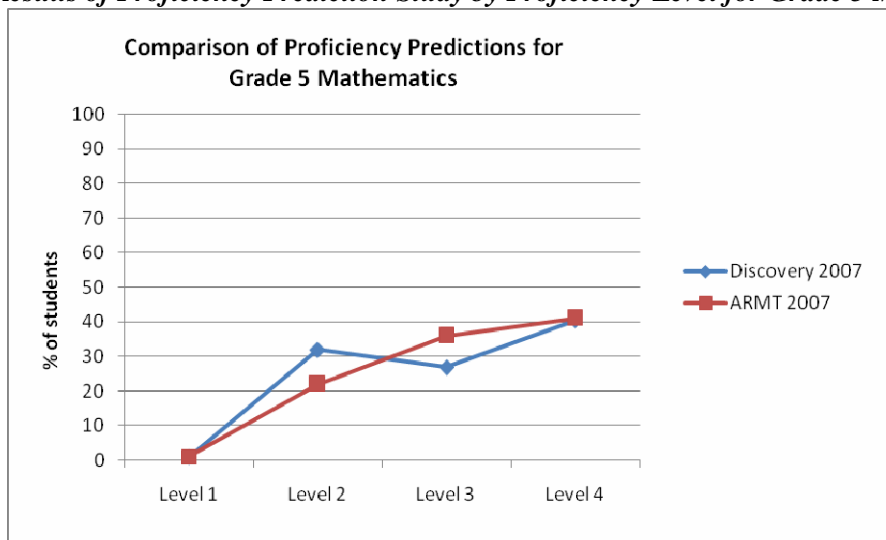
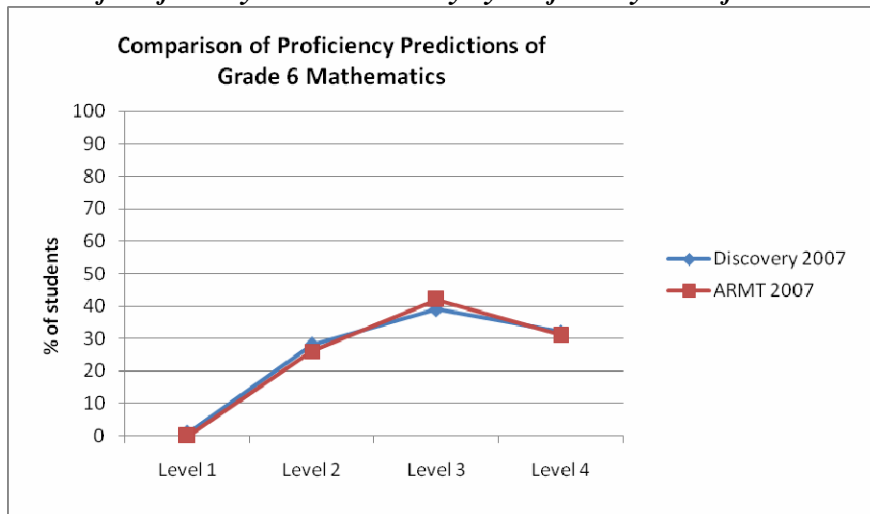


Figure 10: Results of Proficiency Prediction Study by Proficiency Level for Grade 6 Mathematics.





Discovery Education Assessment RESEARCH

Figure 11: Results of Proficiency Prediction Study by Proficiency Level for Grade 7 Mathematics.

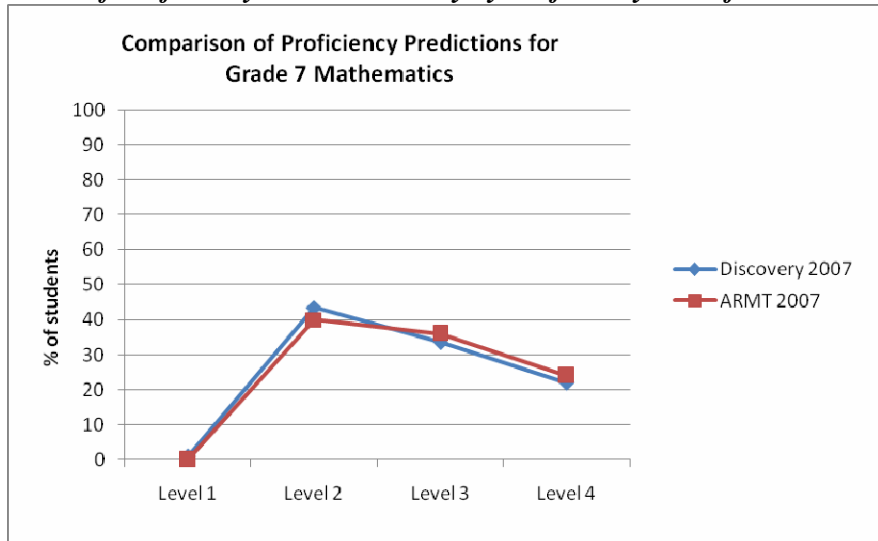
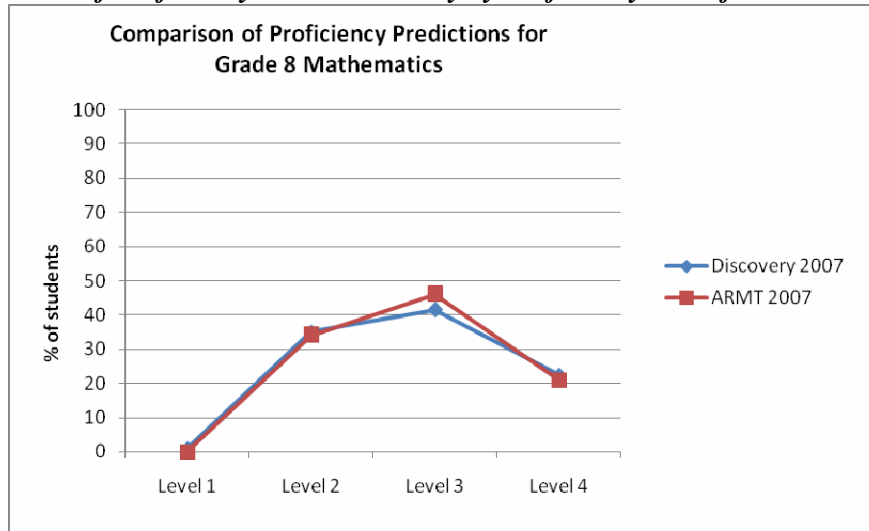


Figure 12: Results of Proficiency Prediction Study by Proficiency Level for Grade 8 Mathematics.





Discovery Education Assessment RESEARCH

CONSEQUENTIAL VALIDITY

5. Can the use of Discovery Education Predictive Assessments improve student learning?

Consequential validity outlines how the use of benchmark assessments facilitates important consequences, such as the improvement of student learning and student performance on state standardized tests.

The Birmingham City Schools participated in a consequential validity study. This system used Discovery Education Predictive Assessments during the 2006-2007 school year. The percent of students that were classified as “Level III” and “Level IV” on the 2006 ARMT was tabulated and compared with the percent of students that were classified as “Level III” and “Level IV” on the 2007 ARMT. The results for Grades 3 to 8, Reading and Mathematics, for the two years—2006 and 2007—are presented in Table 5 and 6. The “Difference” between 2006 and 2007 was also tabulated; a positive score indicates an increase in the percent of students proficient from 2006 to 2007. As a reference point, the improvement (or decline) in the percent of students classified as “Level III” and “Level IV” in the state of Alabama was compared to this Difference score.

The percentages are to be understood as follows. Take a look below at Grade 6 Reading. The percent of students proficient in 2006 was 68, and the percent proficient in 2007 was 73, a difference or improvement of 5%. However, Grade 6 Reading in the state of Alabama improved by only 2% during the same time. Therefore, the “Birmingham ⇔ AL State” calculation is actually 3%. That is, the Birmingham City Grade 3 Reading classes improved 3% in the percent of students proficient compared to the state of Alabama.

Table 5: Results of Consequential Validity Study for Birmingham City Schools in Reading.

Birmingham City, AL Reading				
Grade	2006	2007	Difference*	Birmingham ⇔ AL State*
3	74%	74%	0%	-1%
4	72%	73%	1%	1%
5	70%	74%	4%	0%
6	68%	73%	5%	3%
7	64%	66%	3%	0%
8	60%	62%	2%	1%

**Calculated based on exact not rounded percentages listed under 2006 and 2007.*



Discovery Education Assessment RESEARCH

Table 6: Results of Consequential Validity Study for Birmingham City Schools in Mathematics.

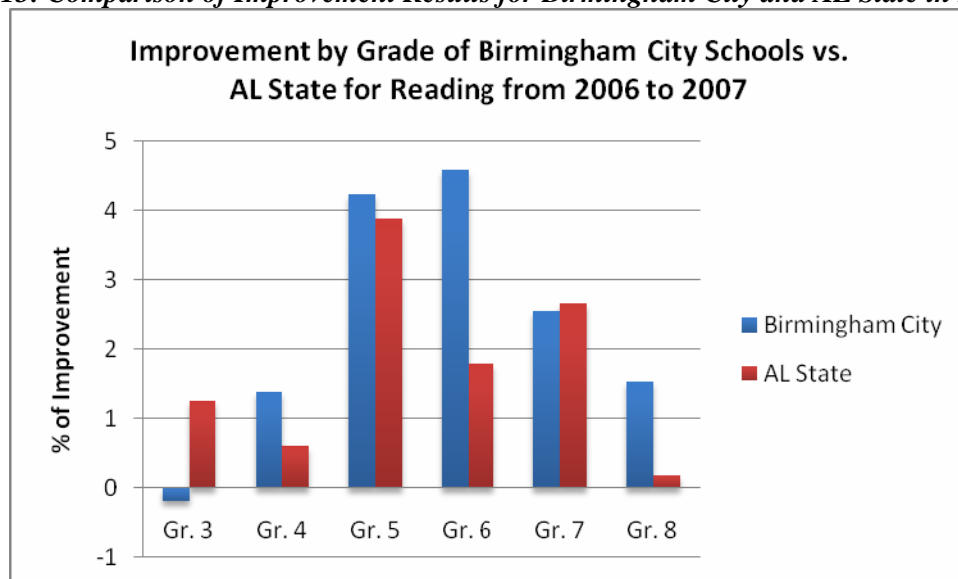
Birmingham City, AL Mathematics				
Grade	2006	2007	Difference*	Birmingham ↔ AL State*
3	68%	68%	0%	0%
4	69%	67%	-2%	-2%
5	69%	69%	0%	-1%
6	58%	60%	2%	4%
7	42%	48%	6%	5%
8	57%	58%	2%	3%

*Calculated based on exact not rounded percentages listed under 2006 and 2007.

Many factors contribute to the improvement of the percent of students proficient from year to year. Discovery Education Predictive Assessments are usually just one factor in school and district-wide improvement plans. Thus, these results should be considered in the light of these many factors.

The following figures graphically represent the data of Table 5 and 6. Birmingham City Schools had significant improvement (exceeding the state's improvement during the same time period) in Grades 4, 6, and 8 Reading and in Grades 6, 7, and 8 Mathematics (see Figure 13 and 14).

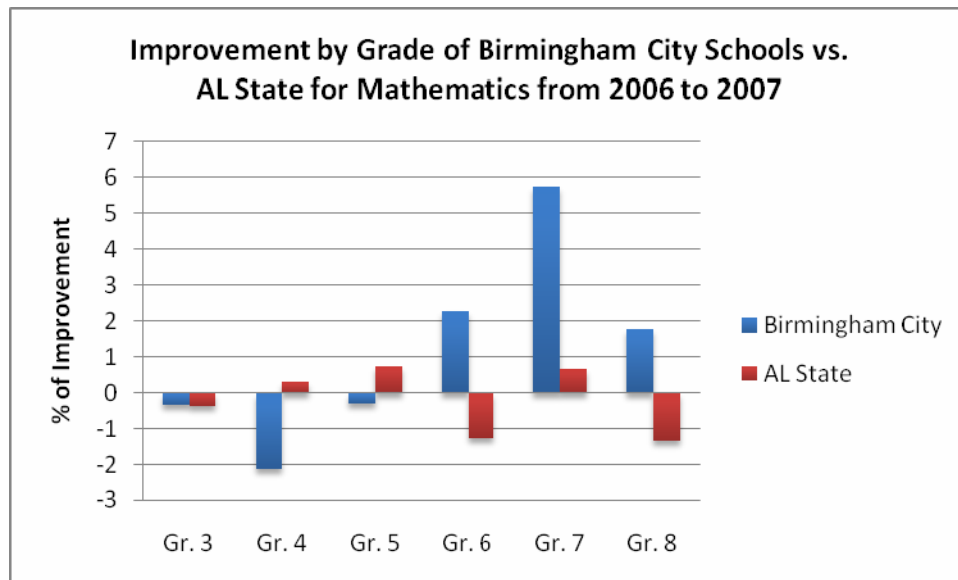
Figure 13: Comparison of Improvement Results for Birmingham City and AL State in Reading.





Discovery Education Assessment RESEARCH

Figure 14: Comparison of Improvement Results for Birmingham City and AL State in Mathematics.





Discovery Education Assessment RESEARCH

GROWTH MODELS

6. Can Discovery Education Predictive Assessments be used to measure growth over time?

Growth models depend on a highly rigorous and valid vertical scale to measure student performance over time. Discovery Education Assessment vertical scales are constructed using Rasch measurement models with state-of-the-art psychometric techniques.

The accurate measurement of student achievement over time is becoming increasingly important to parents, teachers, and school administrators. **Student “growth” within a grade and across grades** has also been sanctioned by the U. S. Department of Education as a reliable way to measure student proficiency in Reading and Mathematics and to **satisfy the requirements of Adequate Yearly Progress (AYP)** under the No Child Left Behind Act. Accurate measurement and recording of individual student achievement can also help with **issues of student mobility**: as students move within a district or state, records of individual student achievement can help new schools administer to the needs of this mobile population.

The assessment of student achievement over time is even more important with the use of benchmark tests. Discovery Education Assessment Benchmark tests provide a snapshot of student progress toward state standards at up to four points during the school year. These benchmark tests are scientifically linked, so that the reporting of student proficiency levels is both reliable and valid.

How is the growth score created?

Discovery Education Assessment has added a scientifically based vertical scaled growth score to its family of benchmark tests in 2007-08. These growth scores are based on the Rasch measurement model, a state-of-the-art psychometric technique for scaling ability (e.g., Wright & Stone, 1979; Wright & Masters, 1982; Linacre 1999; Smith & Smith, 2004; Wilson, 2005). To accomplish vertical scaling, common items are embedded across assessments to enable the psychometric linking of tests at different points in time. For example, a Grade 3 mathematics benchmark test administered mid-year might contain below grade level and above grade level items. Performance on these off grade level items provides an accurate measurement of how much growth occurs across grades. Furthermore, benchmark tests within a grade are also linked with common items, once again to assess change at different points in time within a grade. Discovery Education Assessment is using established psychometric procedures to build calibrated item banks and linked tests (i.e., Ingebo, 1997; Kolen & Brennan, 2004).

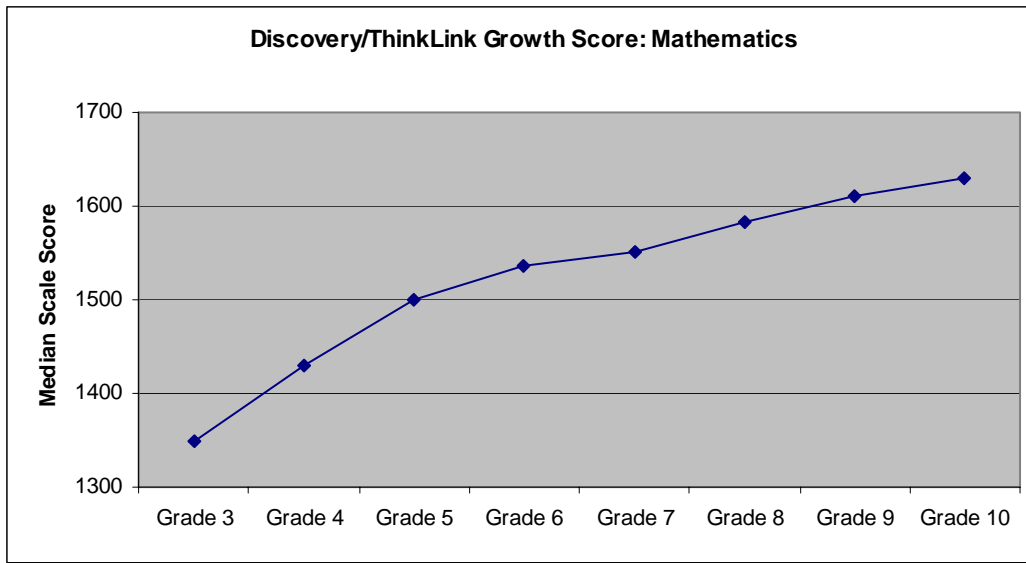
Why use such a rigorous vertical scale?

Isn't student growth similar across grades? Don't students change as much from Grade 3 to Grade 4 as they do from Grade 7 to Grade 8? Previous research on the use of vertical scales has demonstrated that **student growth is not linear**; that is, growth in student achievement is different from grade to grade (see Young 2006). For instance, Figure 15 on the next page shows preliminary Discovery Education Assessment vertically scaled growth results. This graph shows growth from Grades 3 to 10 in Mathematics as measured by Discovery's Spring benchmark tests. Typically, students have larger gains in mathematics achievement in elementary grades with growth somewhat slowing in middle and high school, as published by other major testing companies.



Discovery Education Assessment RESEARCH

Figure 15: Vertically Scaled Growth Results for Discovery Education Assessment Mathematics Tests.



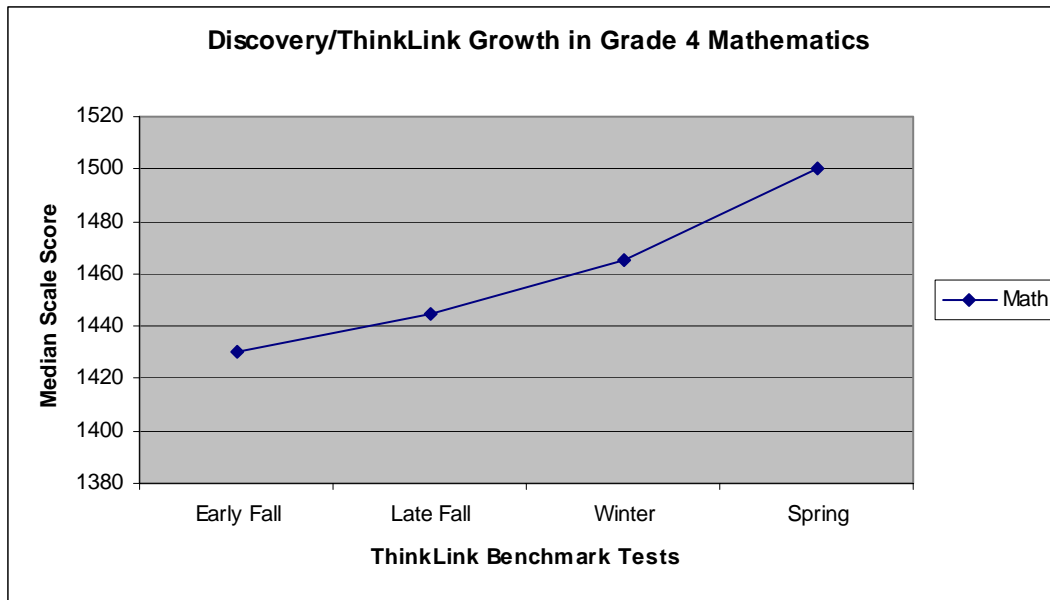
What is unique about the Discovery Education Assessment vertical growth scores?

Student growth can now be accurately measured at four points in time in each grade level. Discovery Education Assessment benchmark tests are administered up to four times yearly: Early Fall, Late Fall, Winter, and Spring. For each time period, we report scale scores and accompanying statistics. Most testing companies only allow the measurement of student growth at two points in time: Fall and Spring. Discovery Education Assessment benchmark tests provide normative information to assess student growth multiple times each year. Figure 16 illustrates this growth for Grade 4 Mathematics using our benchmark assessments.



Discovery Education Assessment RESEARCH

Figure 16: Within-Year Growth Results for Discovery Education Assessment Mathematics Tests.



Alabama Growth Scale

The following tables and figures illustrate the test difficulty on the Discovery Education Assessment vertical growth scale for Reading and Mathematics tests between three time periods, Fall 0708 (Test P), Winter 0708 (Test A), and Spring 0708 (Test B).

Table 7: Vertical Growth Score Comparisons for Fall, Winter, and Spring of 0708 in Reading.

Alabama 0708 Test Difficulty Comparisons							
Reading							
	Gr. 2	Gr. 3	Gr. 4	Gr. 5	Gr. 6	Gr. 7	Gr. 8
Fall	1317	1409	1405	1480	1508	1543	1565
Winter	1393	1413	1463	1503	1516	1527	1571
Spring	1397	1396	1483	1512	1539	1536	1610

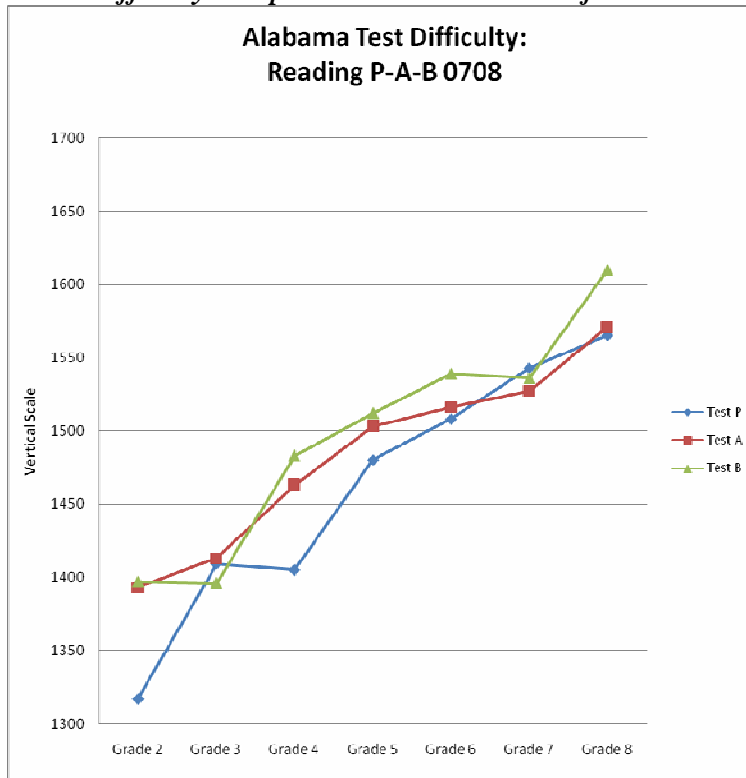


Discovery Education Assessment RESEARCH

Table 8: Vertical Growth Score Comparisons for Fall, Winter, and Spring of 0708 in Mathematics.

Alabama 0708 Test Difficulty Comparisons Mathematics							
	Gr. 2	Gr. 3	Gr. 4	Gr. 5	Gr. 6	Gr. 7	Gr. 8
Fall	1285	1326	1411	1440	1507	1544	1572
Winter	1289	1388	1434	1507	1518	1587	1588
Spring	1315	1380	1458	1491	1538	1562	1588

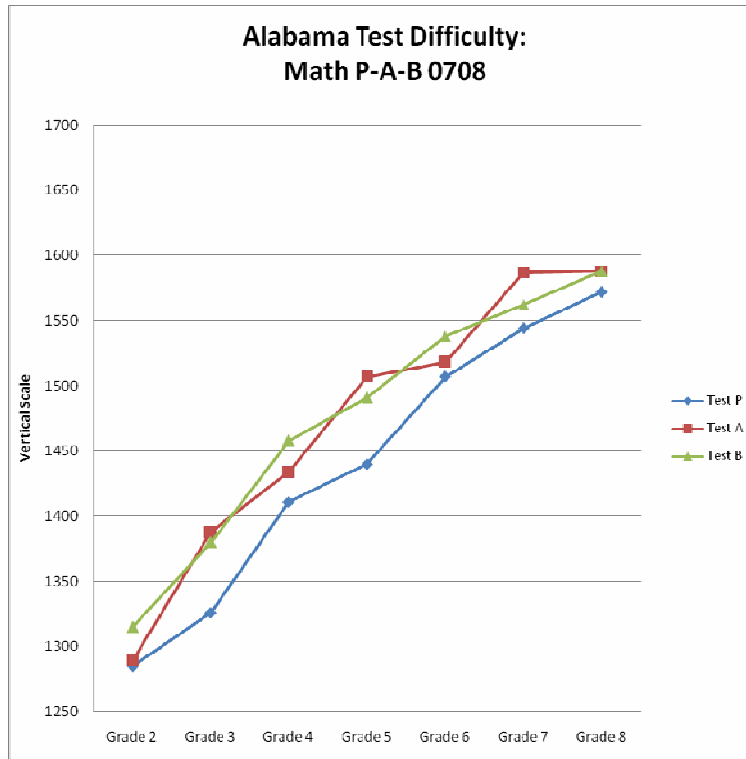
Figure 17: Test Difficulty Comparison on Vertical Scale for 0708 Reading Tests.





Discovery Education Assessment RESEARCH

Figure 18: Test Difficulty Comparison on Vertical Scale for 0708 Mathematics Tests.





Discovery Education Assessment RESEARCH

NCLB SCIENTIFICALLY-BASED RESEARCH

7. Are Discovery Education Predictive Assessments based on scientifically-based research advocated by the U. S. Department of Education?

Discovery Education Assessment has also adhered to the criteria for “scientifically-based research” put forth in the *No Child Left Behind Act of 2001*. “What is Predictive Assessment?” has outlined how Discovery Education Predictive Assessments test reliability and validity meets the following criteria for scientifically-based research set forth by NCLB:

- (i) *employs systematic, empirical methods that draw on observation and experiment;*
- (ii) *involves rigorous data analyses that are adequate to test the stated hypotheses and justify the general conclusions drawn;*
- (iii) *relies on measurements or observational methods that provide reliable and valid data across evaluators and observers, across multiple measurements and observations, and across studies by the same or different investigators;*

Discovery Education Assessment also provides evidence of meeting the following scientifically-based research criterion:

- (iv) *is evaluated using experimental or quasi-experimental designs in which individuals, entities, programs or activities are assigned to different conditions and with appropriate controls to evaluate the effects of the condition of interest, with a preference for random-assignment experiments, or other designs to the extent that those designs contain within-condition or across-condition control.*

Case Study One: Birmingham City Schools, Alabama

Larger schools and school districts typically do not participate in experimental or quasi-experimental studies due to logistical and ethical concerns. However, a unique situation in Birmingham, Alabama afforded Discovery Education Assessment with the opportunity to investigate the efficacy of its benchmark assessments in respect to a quasi-control group. In 2003/2004, approximately one-half of the schools in Birmingham City used Discovery Education Predictive Assessments whereas the other half did not. At the end of the school year, achievement results for both groups were compared revealing a significant improvement on the SAT10 for those schools that used the Discovery Education Predictive Assessments as opposed to those that did not. Discovery Education Assessment subsequently compiled a brief report titled the “Birmingham Case Study”. Excerpts from the case study are included below:

This study is based on data from elementary and middle schools in the City of Birmingham, Alabama. In 2002-03, no Birmingham Schools used Discovery Education’s Predictive Assessment Series. Starting in 2003-04, 20 elementary and 9 middle schools used the Discovery Education Assessment program. All Birmingham schools took the Stanford Achievement Test Tenth Edition (SAT10) at the end of both school years. The SAT10 is administered yearly as part of the State of Alabama’s School Accountability Program. The State of Alabama uses



Discovery Education Assessment RESEARCH

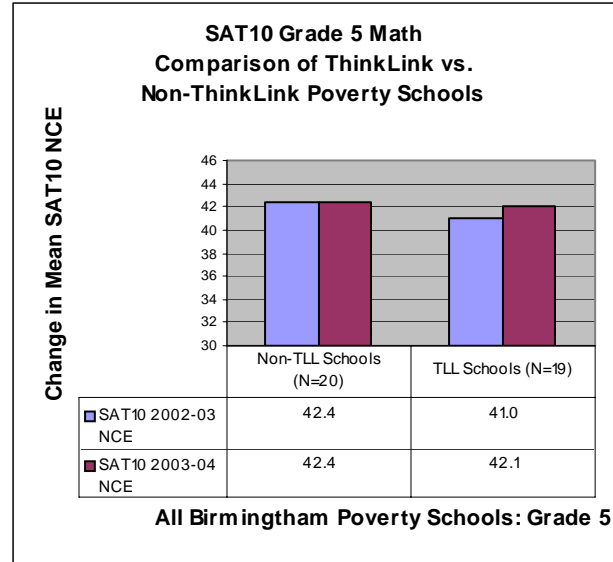
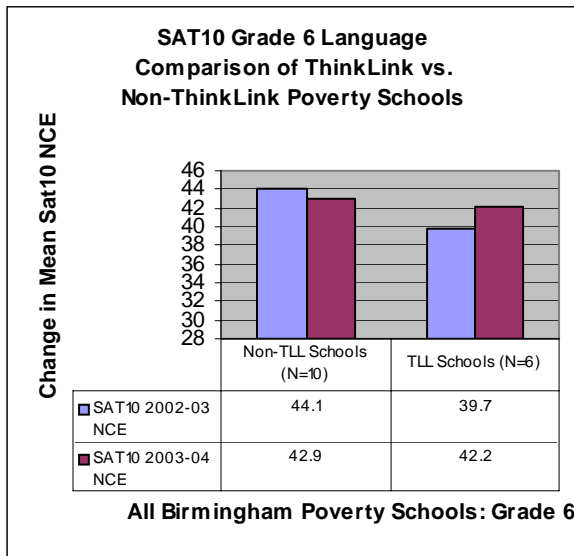
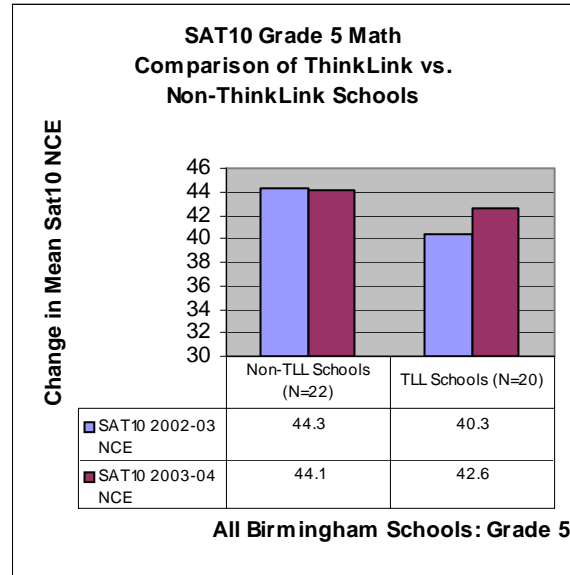
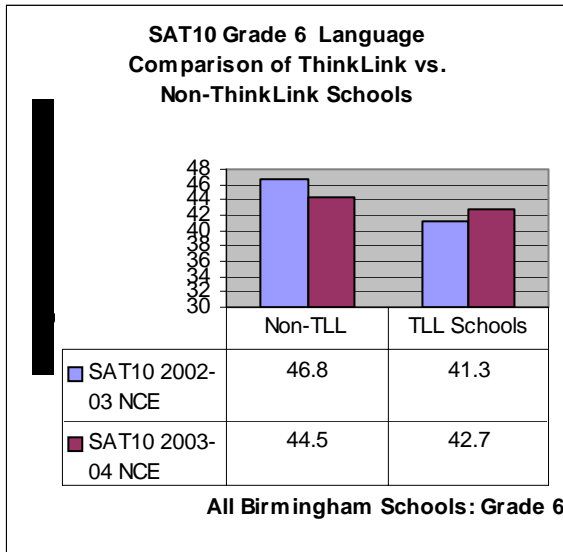
improvement in SAT10 percentiles to gauge school progress and as part of its NCLB reporting. National percentiles on the SAT10 are reported by subject and grade level. A single national percentile is reported for all students within a subject and grade level (this analysis is subsequently referred as ALL STUDENTS). Furthermore, national percentiles are disaggregated by various subgroups within a school. For the comparisons that follow, the national percentiles for students classified as utilizing free and reduced lunch (referred to below as POVERTY) were used. All percentiles have been converted to Normal Curve Equivalents (NCE) to allow for averaging of results.

The Discovery schools comprise the experimental group in this study. The Birmingham schools that did not use Discovery comprise the matched comparison group. The following charts show SAT10 National Percentile changes for Discovery Schools vs. Non-Discovery Schools in two grades levels (Grades 5 and 6) for three subjects (Language, Mathematics, and Reading) for two groups of students (ALL STUDENTS and POVERTY students). In general, there was a significant decline or no improvement in SAT10 scores from 2002-03 to 2003-04 for most non-Discovery schools. This trend however did not happen in the schools using Discovery: instead, there was a marked improvement with most grades scoring increases in language, math and reading. In grade levels where there was a decline in Discovery schools, it was a much lower decline in scores when compared to those schools that did not use Discovery.

As a result of the improvement that many of these schools made in school year 2003-04, the Birmingham City Schools selected Discovery to be used with *all* of the schools in school year 2004-05. The Birmingham City Schools also chose to provide professional development in each school to help all teachers become more familiar with the concepts of standardized assessment and better utilize data to focus instruction.

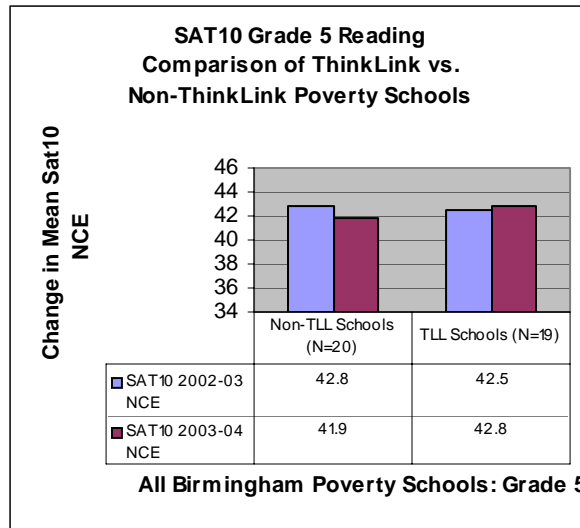
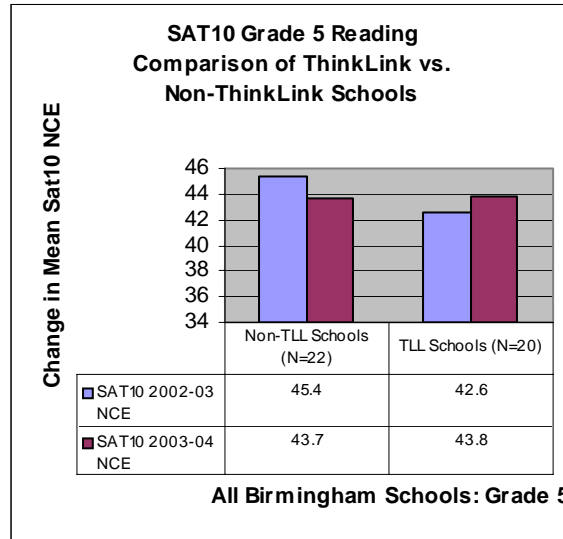


Discovery Education Assessment RESEARCH





Discovery Education Assessment RESEARCH

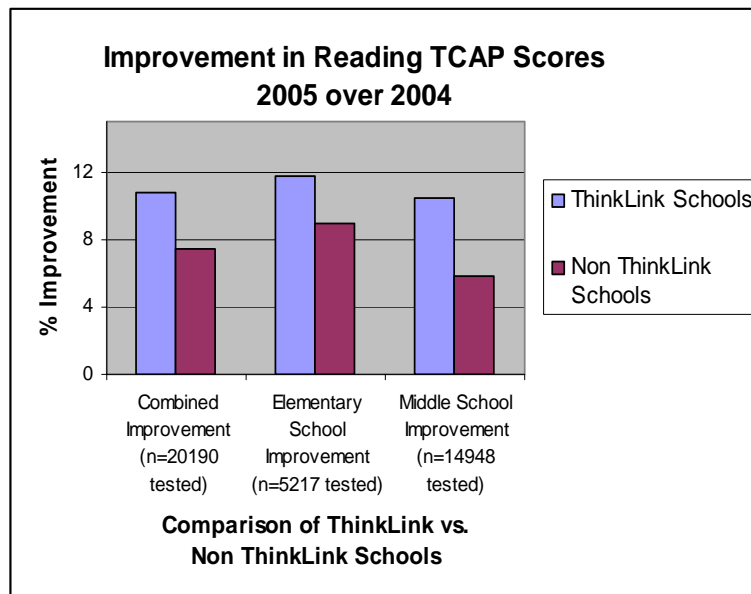




Discovery Education Assessment RESEARCH

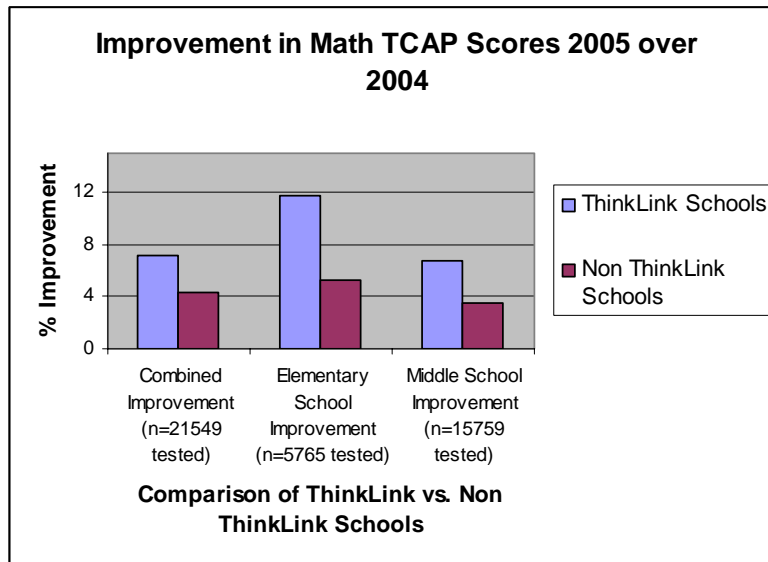
Case Study Two: Metro Nashville Public Schools, Tennessee

Metro Nashville schools that used Discovery Education Assessment made greater improvements in AYP than Metro Nashville schools that didn't use Discovery. During the 2004-2005 school year, sixty-five elementary and middle schools in Metro Nashville, representing over 20,000 students, used Discovery assessments. Fifty-two elementary and middle schools, representing over 10,000 students, did not use Discovery assessments. The improvement in the percent of students at the Proficient/Advanced level from 2004 to 2005 is presented in the graph below. The results compare Discovery schools versus Non-Discovery schools in Metro Nashville. Discovery schools showed more improvement in AYP status from 2004 to 2005 when schools are combined and analyzed separately at the elementary and middle school level.





Discovery Education Assessment RESEARCH



- (v) ensures experimental studies are presented in sufficient detail and clarity to allow for replication or, at a minimum, offer the opportunity to build systematically on their finding;

Consumers are encouraged to request additional data or further details for the examples listed in this overview. Discovery Education Assessment also compiles *Technical Manuals* specific to each school district and/or state. Accumulated data are of sufficient detail to permit adequate psychometric analyses, and their results have been consistently replicated across school districts and states. Past documents of interest include among others: “A Multi-State Comparison of Proficiency Predictions for Fall 2006” and “A Multi-State Look at ‘What is Predictive Assessment?’.” Furthermore, the “What is Predictive Assessment?” series of documents is available for multiple states. Please check the Discovery website www.thinklinklearning.com for document updates.

- (vi) has been accepted by a peer-reviewed journal or approved by a panel of independent experts through a comparably rigorous, objective and scientific review;

Discovery Education Assessment tests and results have been incorporated and analyzed in the following publications, conference proceedings, dissertations, research documents, and tests:

1. Publications

Shrago, J. B., & Smith, M.K. (2006). Online assessment in the K-12 classroom: formative assessment model for improving student performance on standardized tests. In S. Howell & M. Hricko (Eds.), *Online assessment and measurement: case studies from higher education, K-12 and corporate* (pp. 181-194). Hershey, PA: Information Science Publishing.



Discovery Education Assessment RESEARCH

2. Conference Proceedings

Shrago, J. B. chair. (2006, June). *Perspectives on large-scale formative assessment*. Presented at 36th annual nation conference on large-scale assessment hosted by the Council of Chief State School Officers. San Francisco, CA.

Hass, J. (2006, June). *Algebra I pilot project: West Virginia department of education*. Presented at 36th annual nation conference on large-scale assessment hosted by the Council of Chief State School Officers. San Francisco, CA.

Smith, M. K. (2006, June). *How can large scale formative assessment be research-based and valid?* Presented at 36th annual nation conference on large-scale assessment hosted by the Council of Chief State School Officers. San Francisco, CA.

Thompson, E. (2006, June). *Selecting a formative reading assessment: guiding classroom reading instruction and intervention strategies*. Presented at 36th annual nation conference on large-scale assessment hosted by the Council of Chief State School Officers. San Francisco, CA.

Vaughn-Neely, E., & Reed, M. (2005). *Reading findings*. Presented at Society for Research & Child Development. Atlanta, GA.

Vaughn-Neely, E., & Reed, M. (2006). *Reading findings*. Presented at Society on Scientific Study of Reading. Toronto, CA.

3. Dissertations

Johnson, J. (2005). *A multivariate analysis of the effects of the transition from elementary to middle school on the mathematics academic performance, personal achievement goal orientations, and achievement-related beliefs, perceptions and strategies of fifth grade students*. Unpublished doctoral dissertation, Union University, Jackson, TN.

4. Research Documents

Shrago, J. B., & Smith, M. K. (2006). *The uses of benchmark tests to improve student learning*. Nashville, TN: Discovery Education.

Smith, M. K. (2006). *Case study: Birmingham city school district, Ala.* Nashville, TN: Discovery Education.

Smith, M. K., & Kurz, A. (2008). *What is predictive assessment: Alabama?* Nashville, TN: Discovery Education.

Smith, M. K., & Kurz, A. (2008). *What is predictive assessment: Florida?* Nashville, TN: Discovery Education.



Discovery Education Assessment RESEARCH

Smith, M. K., & Kurz, A. (2008). *What is predictive assessment: Illinois?*
Nashville, TN: Discovery Education.

Smith, M. K., & Kurz, A. (2008). *What is predictive assessment: Kentucky?*
Nashville, TN: Discovery Education.

Smith, M. K., & Kurz, A. (2008). *What is predictive assessment: New York?*
Nashville, TN: Discovery Education.

Smith, M. K., & Kurz, A. (2008). *What is predictive assessment: Tennessee?*
Nashville, TN: Discovery Education.

5. Tests

Discovery Education Benchmark Assessments. *Kentucky test b: reading language grade eight.*
(2007). Nashville, TN: Discovery Education.

Please contact us for other specific information requests. We welcome your interest in the evidence supporting the efficacy of our Discovery Education Assessment tests.