What is Predictive Assessment?
NEW YORK EXECUTIVE SUMMARY

1. Are Discovery Education Predictive Assessments reliable?

These benchmark assessments are highly reliable. For the 0708 A and B Reading tests Grades 3 to 8, the median reliability was .84 with a median sample size of 322. The median 0708 Mathematics reliability for the A and B tests was .83 with a sample size of 395.

2. Do Discovery Education Predictive Assessments have content validity?

These benchmark assessments model the competencies and performance indicators of the New York State core curricula for English Language Arts and Mathematics.

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

Individual student level data necessary to conduct criterion-related validity studies in New York 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 New York customers to inspect the evidence gathered in other states such as Florida and Tennessee.

4. Can Discovery Education Predictive Assessments predict proficiency levels?

Yes, there is a greater than 80% accuracy rate for predicting state proficiency levels. While our current New York sample size is still limited, a proficiency prediction analysis of over 1200 students at Albion and William H. Golding middle school offered strong initial results: the median Proficiency Prediction Score for Test B English Language Arts (ELA) was 82%, and the median Proficiency Prediction Score for Mathematics was 91% for Test A and 88% for Test B.

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

In states with a large customer base, Discovery Education Assessment has conducted comparative analyses of the improvement of Discovery versus Non-Discovery schools. Evidence that improvement in Discovery schools is greater than in Non-Discovery schools has emerged in several large customer states such as Florida and Tennesse. Our assessment series in New York is built on the demonstrated success and efficacy of our Discovery Education Predictive Assessements in other states. Evidence to support consequential validity will be gathered and provided as soon as sufficient data will become available for New York.

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.

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.

What is Predictive Assessment?

NEW YORK

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 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.

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;

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 and 2 present current test reliabilities and sample sizes of the 0708 A and B test cycles—Fall and Spring— for our Discovery Education Predictive Assessments in the subject areas of English Language Arts (ELA) and Mathematics.

The median English Language Arts reliability was .84 with a median sample size of 322. The median Mathematics reliability was .83 with a sample size of 395.

Table 1: Test Reliabilities for English Language Arts and Mathematics Test A Fall 2007.

	New York – Test A 0708			
	ELA	N	Mathematics	N
Grade 3	.85	502	.83	346
Grade 4	.85	497	.83	433
Grade 5	.81	647	.78	499
Grade 6	.82	450	.83	561
Grade 7	.84	318	.82	356
Grade 8	.79	326	.75	352

Table 2: Test Reliabilities for English Language Arts and Mathematics Test B Spring 2008.

	New York – Test B 0708			
	ELA	N	Mathematics	N
Grade 3	.88	140	.85	513
Grade 4	.84	218	.84	510
Grade 5	.84	332	.81	633
Grade 6	.79	289	.85	339
Grade 7	.85	164	.87	157
Grade 8	.82	273	.84	176

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 objectives are used on Discovery Education Predictive Assessments for New York in English Language Arts and Mathematics. These objectives and reporting categories are based on New York State's core curricula. We continually update our assessments to reflect the most current version of a state's standards.

NY English Language Arts Reporting Categories	
Reading: Information and Understanding	Writing: Literary Response and Expression
Reading: Literary Response and Expression	Writing: Critical Analysis and Evaluation
Reading: Critical Analysis and Evaluation	Reading: Core Performance Indicators
Writing: Information and Understanding	Writing: Core Performance Indicators

NY Mathematics Reporting Categories	
Number Sense and Operations	Measurement
Algebra	Statistics and Probability
Geometry	

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 New York 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 New York customers to inspect the evidence gathered in 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.

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 is available, then an additional index, the *Proficiency Success Rate*, is also calculated. Both measures are explained in the following sections.

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 underpredicted 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 overpredicts 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%.

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.

Case Study: Albion and William H. Golding Middle School
Comparisons of Discovery Education Assessment proficiency predictions between the 0607 Test A and Test B results and actual 2007 NY State test results were made for our two largest middle school customers Grades 6-8 in English Language Arts and Mathematics.

The Proficiency Prediction Scores for Test A Grades 6-8 in Mathematics at Albion Middle are presented in Table 3. Figure 1 provides the respective proficiency averages by grade. The median Proficiency Prediction Score for Test A at Albion Middle was 91%. Table 4 and Figure 2 provide the Proficiency Prediction Scores and proficiency averages for Test B Grades 6-8 in Mathematics at Albion Middle. The median Proficiency Prediction Score for Test B at Albion Middle was 88%. Table 5 and Figure 3 provide the Proficiency Prediction Scores and proficiency averages for Test B Grades 6-8 in English Language Arts (ELA) at William H. Golding. The median Proficiency Prediction Score for Test B at William H. Golding was 82%.

Table 3: Albion Middle Test A Proficiency Prediction Scores for Mathematics.

	Mathematics Discovery Education Assessment Test A	
	N	Proficiency Prediction Score
Grade 6	170	96%
Grade 7	187	89%
Grade 8	207	91%
Median		**91%**

Figure 1: Results of Aggregated Proficiency Predictions for Mathematics Test A versus NY State.

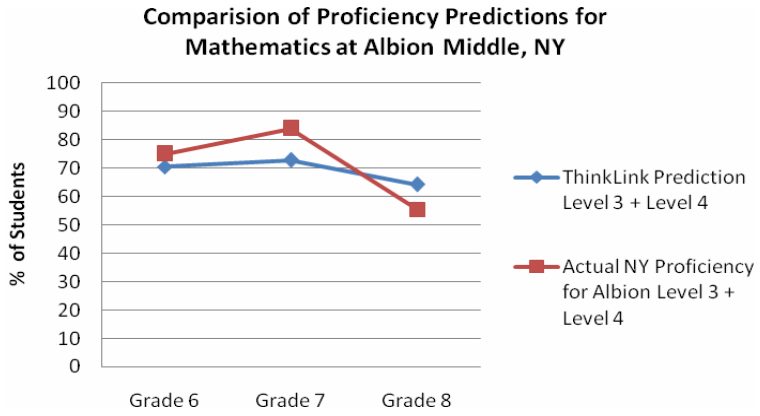


Table 4: Albion Middle Test B Proficiency Prediction Scores for Mathematics.

	Mathematics Discovery Education Assessment Test B	
	N	**Proficiency Prediction Score**
Grade 6	165	87%
Grade 7	185	96%
Grade 8	185	88%
Median		**88%**

Figure 2: Results of Aggregated Proficiency Predictions for Mathematics Test B versus NY State.

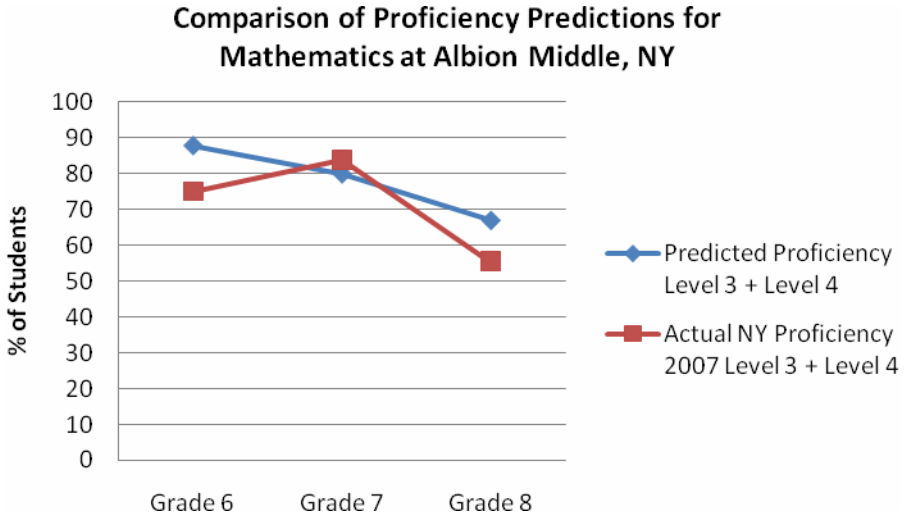
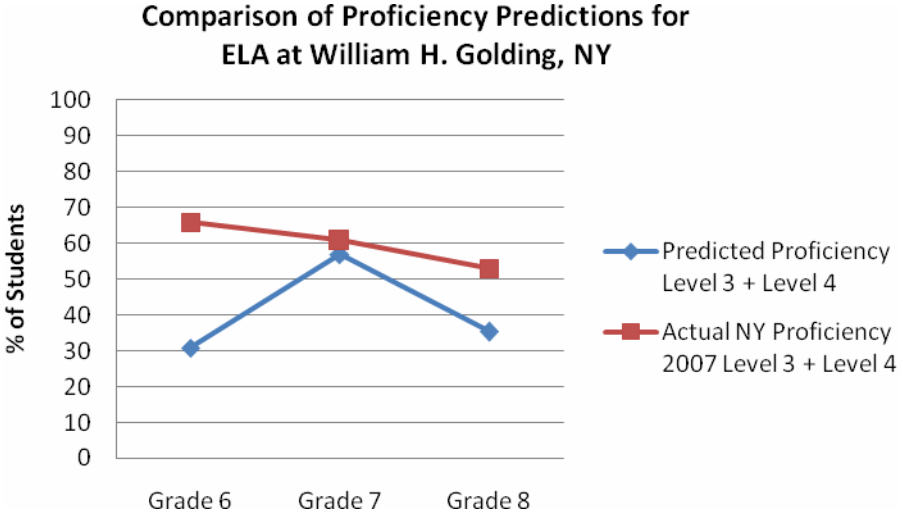


Table 5: William H. Golding Middle Test B Proficiency Prediction Scores for English Language Arts.

	English Language Arts Discovery Education Assessment Test B	
	N	**Proficiency Prediction Score**
Grade 6	52	65%
Grade 7	128	96%
Grade 8	161	82%
Median		**82%**

Figure 3: Results of Aggregated Proficiency Predictions for English Language Arts Test B versus NY State.



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.

Many factors contribute to the improvement of student learning. The use of Discovery Education Predictive Assessments is typically only one of several interventions used by a given state or school system. In states with a large customer base, Discovery Education Assessment has conducted comparative analyses of the improvement of Discovery versus Non-Discovery schools. Evidence that improvement is greater in Discovery schools compared to Non-Discovery schools has emerged in several large customer states such as Florida and Tennesse. Our assessment series in New York is built on the demonstrated success and efficacy of our Discovery Education Predictive Assessements in other states. Evidence to support consequential validity will be gathered and provided as soon as sufficient data will become available for New York.

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.

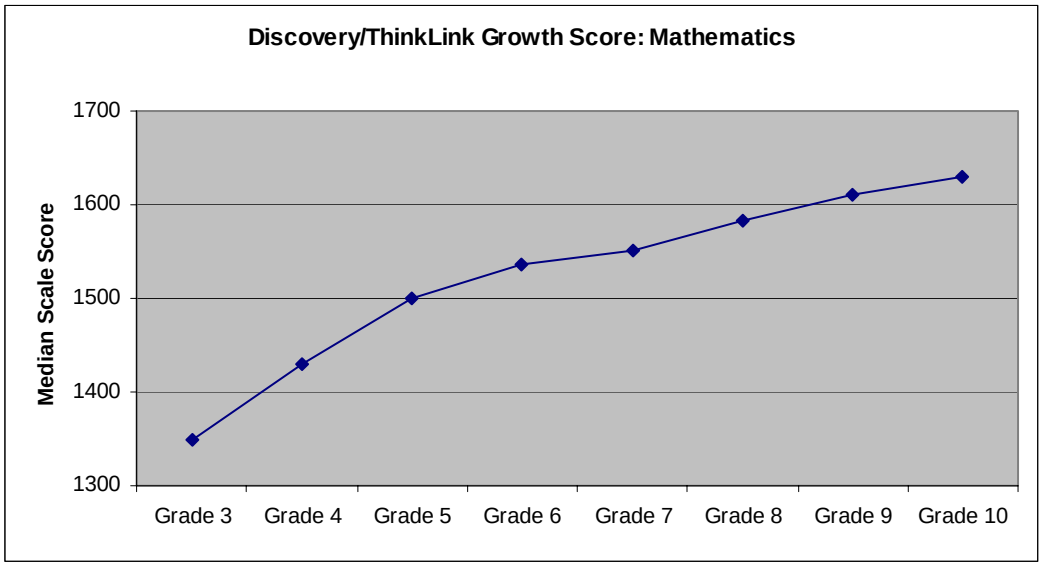
The accurate measurement of student achievement over time is becoming increasingly important to parents, teachers, and school administrators. **Student "growth" within a grade and across grades** has also been sanctioned by the U. S. Department of Education as a reliable way to measure student proficiency 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 benchmarks 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).

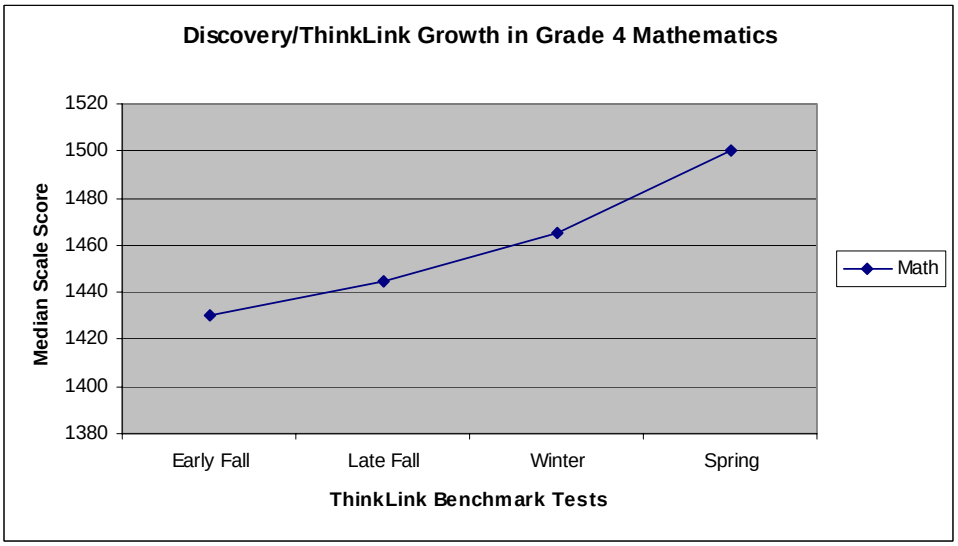
Why use such a rigorous vertical scale?
Isn't student growth similar across grades? Don't students change as much from Grade 3 to Grade 4 as they do from Grade 7 to Grade 8? Previous research on the use of vertical scales has demonstrated that **student growth is not linear**; that is, growth in student achievement is different from grade to grade (see Young 2006). For instance, Figure 4 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 Education Assessment'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.

Figure 4: 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 5 illustrates this growth for Grade 4 Mathematics using our benchmark assessments.

Figure 5: Within-Year Growth Results for Discovery Education Assessment Mathematics Tests.



New York Growth Scale

Tables 6 and 7 and figures 6 and 7 illustrate the Test Difficulty on the Discovery Education Assessment vertical growth scale for English Language Arts and Mathematics tests for two time periods, Test A 0708 and Test B 0708.

Table 6: Vertical Growth Score Comparisons for 0708 Test A and B in English Language Arts.

New York 0708 Test Difficulty Comparisons English Language Arts						
	Gr. 3	Gr. 4	Gr. 5	Gr. 6	Gr. 7	Gr. 8
Test A	1408	1442	1495	1511	1533	1596
Test B	1393	1457	1504	1532	1562	1589

Table 7: Vertical Growth Score Comparisons for 0708 Test and B in Mathematics.

New York 0708 Test Difficulty Comparisons Mathematics						
	Gr. 3	Gr. 4	Gr. 5	Gr. 6	Gr. 7	Gr. 8
Test A	1358	1443	1506	1552	1595	1636
Test B	1377	1456	1508	1566	1582	1621

Figure 6: Test Difficulty Comparisons on Vertical Scale for 0708 English Language Arts Tests.

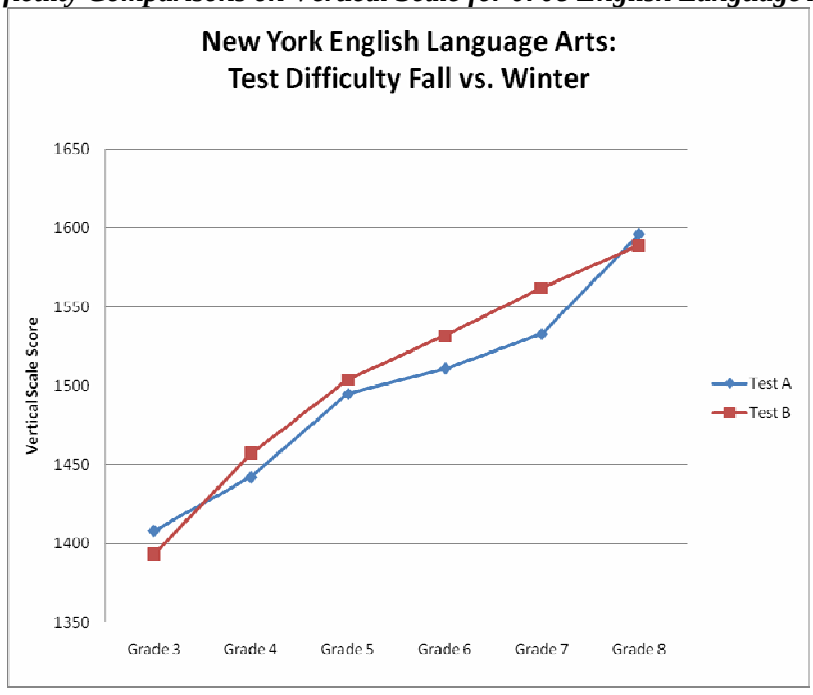
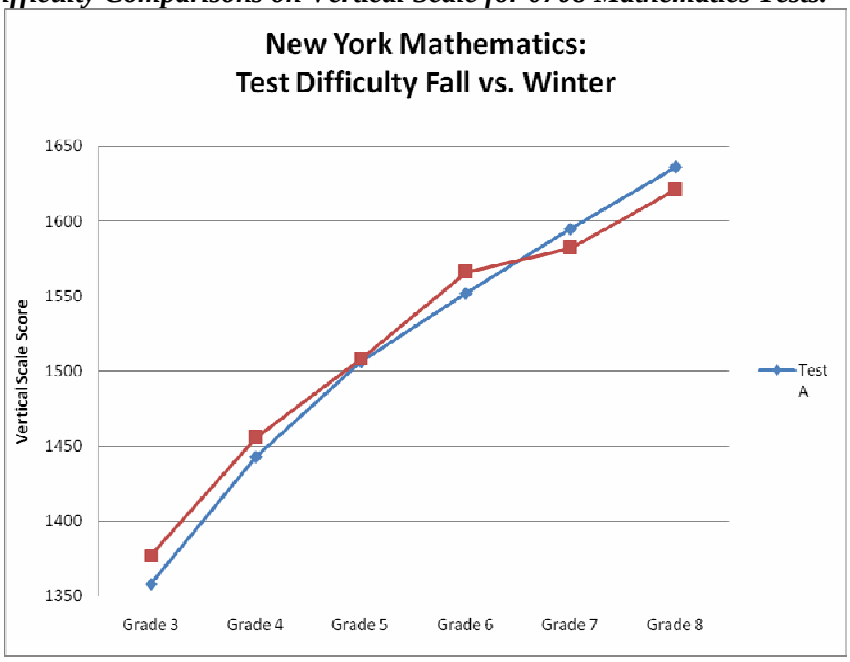


Figure 7: Test Difficulty Comparisons on Vertical Scale for 0708 Mathematics Tests.



Tables 8 and 9 and figures 8 and 9 illustrate the Student Test Averages on the Discovery Education Assessment vertical growth scale for English Language Arts and Mathematics tests for two time periods, Test A 0708 and Test B 0708.

Table 8: Vertical Growth Score Comparisons for 0708 Test A and B in English Language Arts.

New York 0708 Student Ability Comparisons English Language Arts						
	Gr. 3	Gr. 4	Gr. 5	Gr. 6	Gr. 7	Gr. 8
Test A	1436	1474	1525	1588	1566	1615
Test B	1452	1499	1534	1571	1609	1622

Table 9: Vertical Growth Score Comparisons for 0708 Test and B in Mathematics.

New York 0708 Student Ability Comparisons Mathematics						
	Gr. 3	Gr. 4	Gr. 5	Gr. 6	Gr. 7	Gr. 8
Test A	1366	1454	1519	1534	1594	1607
Test B	1428	1485	1529	1605	1630	1662

Figure 8: Student Ability Comparisons on Vertical Scale for 0708 English Language Arts Tests.

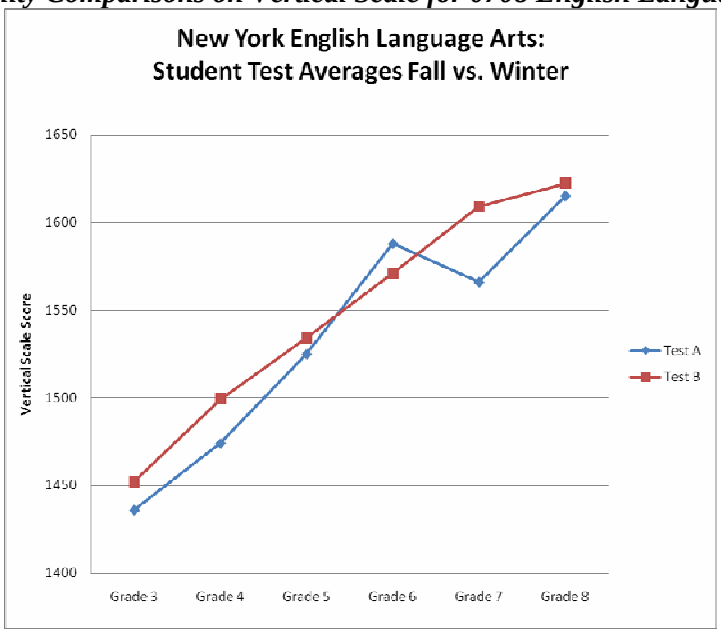
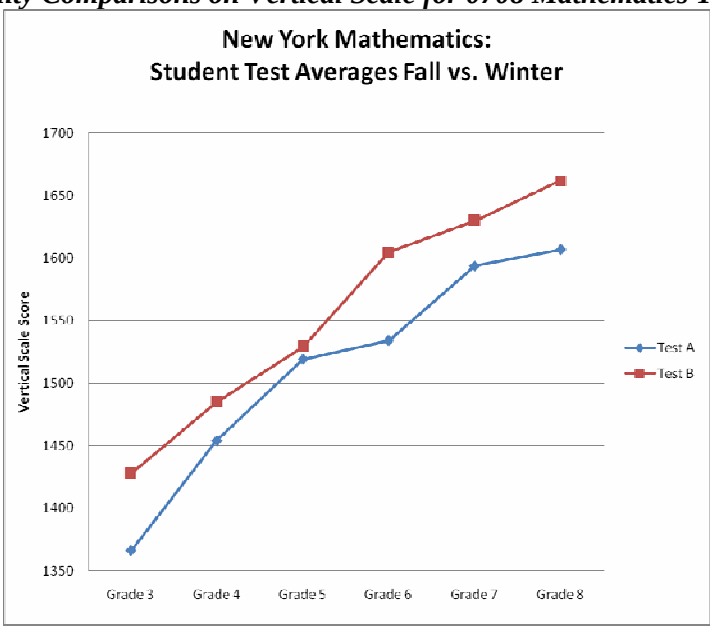


Figure 9: Student Ability Comparisons on Vertical Scale for 0708 Mathematics Tests.



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, Alabama City Schools

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 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

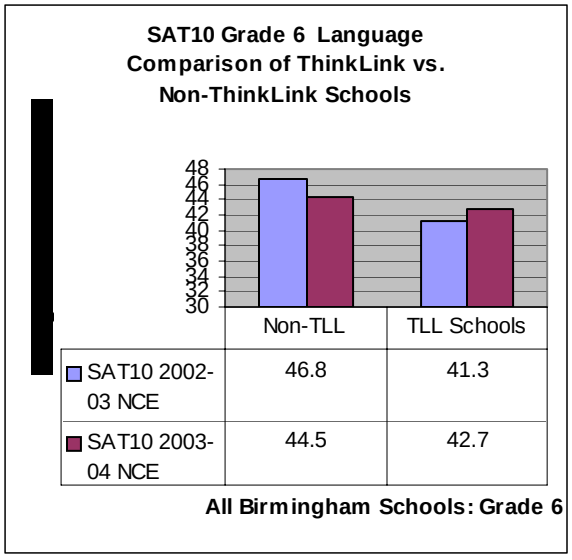
single national percentile is reported for all students within a subject and grade level (this analysis is subsequently referred as ALL STUDENTS). Furthermore, national percentiles are disaggregated by various subgroups within a school. For the comparisons that follow, the national percentiles for students

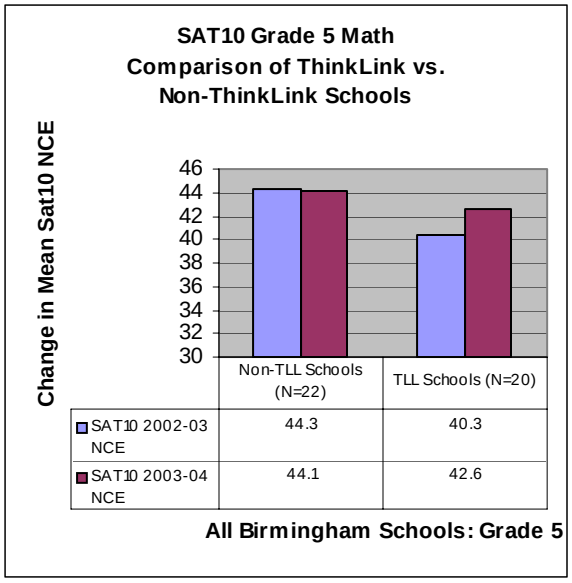
classified as utilizing free and reduced lunch (referred to below as POVERTY) were used. All percentiles have been converted to Normal Curve Equivalents (NCE) to allow for averaging of results.

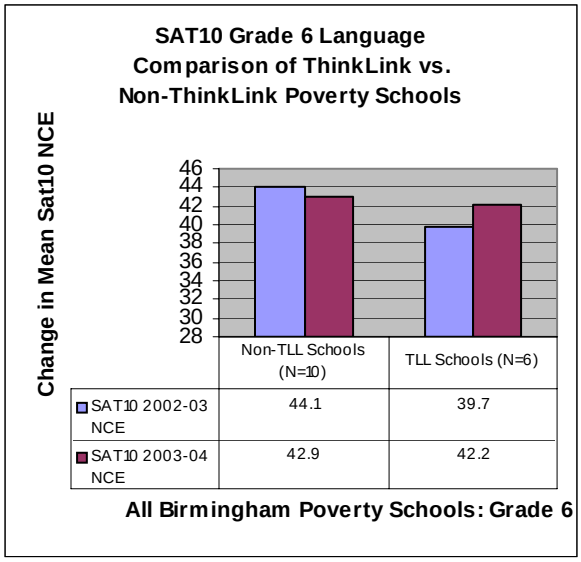
The Discovery Education Assessment schools comprise the experimental group in this study. The Birmingham schools that did not use Discovery Education Assessment comprise the matched comparison group. The following charts show SAT10 National Percentile changes for Discovery Education Assessment Schools vs. Non-Discovery Education Assessment 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 Education Assessment schools. This trend however did not happen in the schools using Discovery Education Assessment: 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 Education Assessment schools, it was a much lower decline in scores when compared to those schools that did not use Discovery Education Assessment.

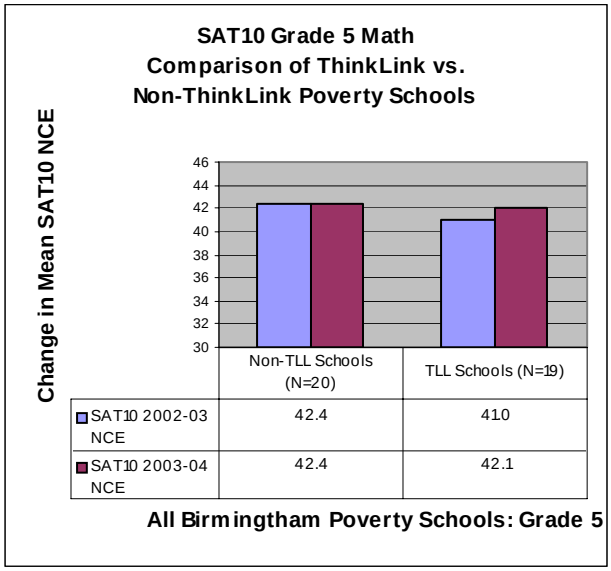
As a result of the improvement that many of these schools made in school year 2003-04, the Birmingham City Schools selected Discovery Education Assessment to be used with *all* of the schools in school year 2004-05. The Birmingham City Schools also chose to provide professional development in each school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

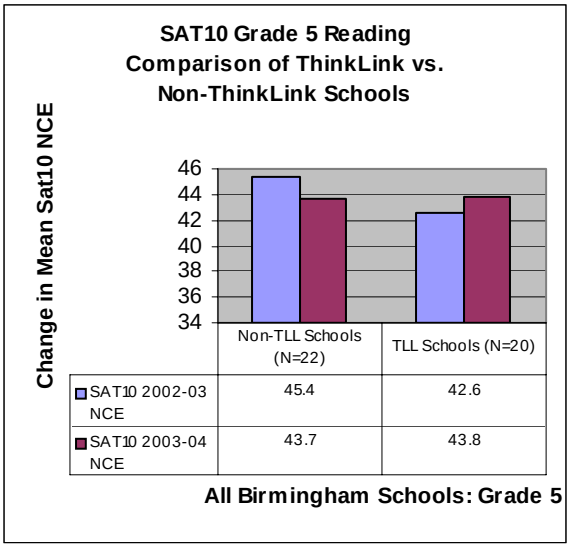
Discovery Education Assessment
RESEARCH

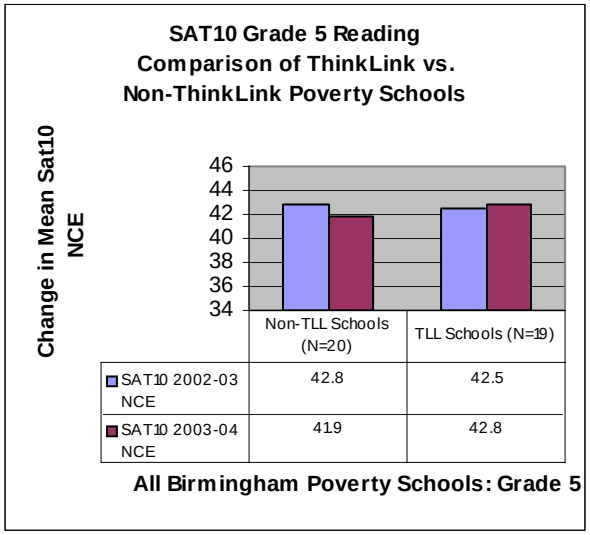






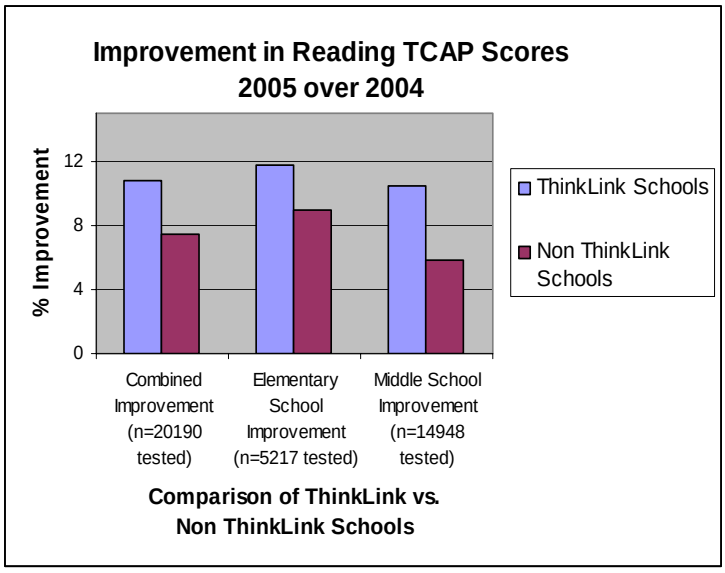


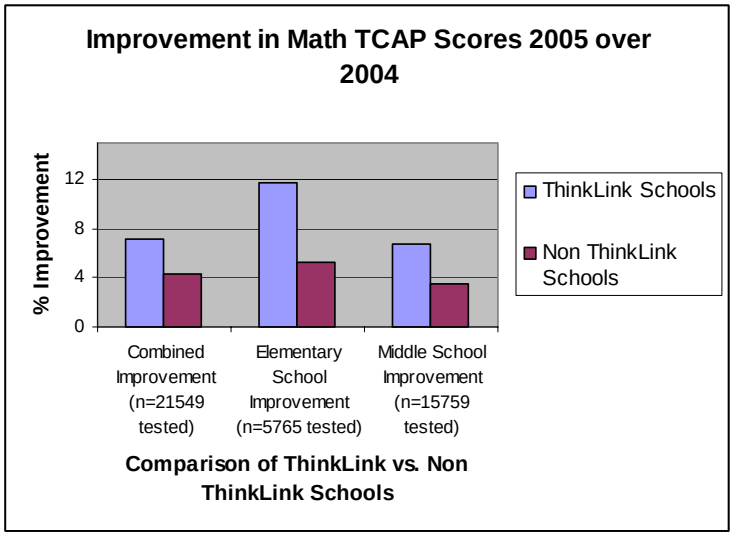

SAT10 Grade 5 Reading
Comparison of ThinkLink vs.
Non-ThinkLink Schools
Change in Mean Sat10 NCE
46
44
42
40
38
36
34
Non-TLL Schools (N=22)
TLL Schools (N=20)
SAT10 2002-03 NCE
45.4
42.6
SAT10 2003-04 NCE
43.7
43.8
All Birmingham Schools: Grade 5


SAT10 Grade 5 Reading
Comparison of ThinkLink vs.
Non-ThinkLink Poverty Schools
Change in Mean Sat10 NCE
46
44
42
40
38
36
34
Non-TLL Schools (N=20)
TLL Schools (N=19)
SAT10 2002-03 NCE
42.8
42.5
SAT10 2003-04 NCE
41.9
42.8
All Birmingham Poverty Schools: Grade 5

Case Study Two: Metro Nashville, Tennessee City Schools

Metro Nashville schools that used Discovery Education Assessment made greater improvements in AYP than Metro Nashville schools that didn't use Discovery Education Assessment. During the 2004-2005 school year, sixty-five elementary and middle schools in Metro Nashville, representing over 20,000 students, used Discovery Education Assessment assessments. Fifty-two elementary and middle schools, representing over 10,000 students, did not use Discovery Education Assessment 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 Education Assessment schools versus non-Discovery Education Assessment schools in Metro Nashville. Discovery Education Assessment 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.





(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 Education website at 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 peer-reviewed manuscripts and publications:

Jacqueline Shrago and **Dr. Michael Smith** of Discovery Education Assessment contributed a chapter on formative assessment to "Online Assessment and Measurement: Case Studies from Higher Education, K-12, and Corporate" by Scott Howell and Mary Hicko in 2006.

Dr. Elizabeth Vaughn-Neely, Associate Professor, Dept. of Leadership & Counselor Education, Ole Miss University (eiv@olemiss.edu) and **Dr. Marjorie Reed**, Associate Professor, Dept. of Psychology, Oregon State University presented their peer-reviewed findings based on their joint research and work with schools using Discovery Education Assessment at:

Society for Research & Child Development, Atlanta, GA. April 2005
Kappa Delta Pi Conference, Orlando, FL November 2005
Society on Scientific Study of Reading, July 2006

Two dissertations for Ed.S studies have also been published:
Dr. Juanita Johnson, Union University (johnsonj4@k12tn.net)
Dr. Monica Eversole, Richmond KY (meversol@madison.k12.ky.us)

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