



# Discovery Education Assessment RESEARCH

## What is Predictive Assessment?

### TENNESSEE EXECUTIVE SUMMARY

#### 1. Are Discovery Education Predictive Assessments reliable?

These benchmark assessments are highly reliable. For Grades 3 to 8 Reading tests over three time periods (Fall, Winter, Spring), the median reliability was .82 with a median sample size of 30,027. The median Mathematics reliability was .82 with a sample size of 30,390. The median Science reliability was .76 with a median sample size of 5,804.

#### 2. Do Discovery Education Predictive Assessments have content validity?

These benchmark assessments model the objectives and skills of the TCAP Assessment Standards for Reading, Mathematics, and Science.

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

The Grainger County school system participated in a criterion validity study for the 2006-2007 school year. Approximately 1500 students, Grades 3 through 8, completed the Discovery Education Predictive Assessments in Reading and Mathematics. For each grade and subject, the students' percent correct averages were broken down by reporting category and compared to their respective 2007 TCAP averages. The results indicated a close match between predicted and actual percent correct averages. Furthermore, the Memphis City School system, representing 40,000 students, took all three DEA benchmarks during the 2008-2009 school year. Predictive validity coefficients between DEA benchmark scale scores and TCAP 2009 test scores had a median value of .69. This result indicates that DEA benchmarks are highly predictive of TCAP scale scores.

#### 4. Can Discovery Education Predictive Assessments predict proficiency levels?

Yes, there is a greater than 90% accuracy rate for predicting combined state proficiency percentages. Due to our representativeness throughout the state of Tennessee, direct comparisons of Spring 2009 Test B and actual 2009 TCAP proficiency percentages ("Proficient" and "Advanced" combined) were made for Grades 3 through 8 in Reading and Mathematics. The median Proficiency Prediction Score for Reading was 98%, and the median Proficiency Prediction Score for Mathematics was 98%.

#### 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 TCAP to the 2007 TCAP was conducted for Grainger County and the state of Tennessee. Grainger County had significant improvement (exceeding the state's improvement during the same time period) in Grades 3, 4, 5, 6, and 8 Mathematics and in Grades 3 and 4 Reading.



# Discovery Education Assessment RESEARCH

## **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.



# Discovery Education Assessment RESEARCH

## What is Predictive Assessment?

### TENNESSEE

#### 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;*

*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 for 2008-2009.

The median Reading reliabilities were .82, .80, and .82 for the three time periods with median sample sizes of 23,589, 30,543, and 30,027. The median Mathematics reliabilities for the three time periods were .82, .80, and .82. The median Science reliabilities were .76 for the Winter and Spring time periods.

**Table 1: Test Reliabilities for Reading and Mathematics Fall 2008.**

Tennessee Test P – Fall 2008				
	Reading	N	Mathematics	N
Grade 3	.83	26,012	.81	26,265
Grade 4	.81	25,699	.83	25,704
Grade 5	.83	24,993	.82	24,965
Grade 6	.79	22,185	.81	21,938
Grade 7	.83	20,735	.78	21,389
Grade 8	.81	21,588	.83	21,457
<b>Median</b>	<b>.82</b>	<b>23,589</b>	<b>.82</b>	<b>23,452</b>

**Table 2: Test Reliabilities for Reading, Mathematics, and Science Winter 2008.**

Tennessee Test A – Winter 2008						
	Reading	N	Mathematics	N	Science	N
Grade 3	.79	33,242	.82	33,494	.86	127
Grade 4	.76	33,385	.79	33,471	.82	108
Grade 5	.81	31,626	.78	31,791	.77	373
Grade 6	.80	29,429	.77	29,949	.67	538
Grade 7	.81	29,190	.82	29,392	.66	411
Grade 8	.80	29,460	.81	29,636	.74	399
<b>Median</b>	<b>.80</b>	<b>30,543</b>	<b>.80</b>	<b>30,870</b>	<b>.76</b>	<b>386</b>



# Discovery Education Assessment RESEARCH

*Table 3: Test Reliabilities for Reading, Mathematics, and Science Spring 2009.*

	Tennessee Test B – Spring 2009					
	Reading	N	Mathematics	N	Science	N
Grade 3	.84	32,624	.82	32,796	.81	524
Grade 4	.82	32,520	.82	32,779	.76	524
Grade 5	.82	30,801	.82	31,021	.73	465
Grade 6	.81	29,253	.80	29,758	.77	635
Grade 7	.76	29,182	.82	29,258	.71	466
Grade 8	.84	29,043	.85	29,324	.77	775
<b>Median</b>	<b>.82</b>	<b>30,027</b>	<b>.82</b>	<b>30,390</b>	<b>.76</b>	<b>580</b>



# 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 on Discovery Education Predictive Assessments for Tennessee in Reading, Mathematics, and Science that were put into place during the 2009-2010 school year. They are based on TCAP Assessment Standards and thus mirror the TCAP reporting categories. We continually update our assessments to reflect the most current version of a state's standards.

#### TN 0910 Reading Reporting Categories

Language	Logic
Communications	Informational Text
Writing	Media
Research	Literature

#### TN 0910 HS End-Of-Course English Reporting Categories

Standard 1: Language	Standard 5: Logic
Standard 2: Communication	Standard 6: Informational Text
Standard 3: Writing	Standard 7: Media
Standard 4: Research	Standard 8: Literature



# Discovery Education Assessment RESEARCH

## TN 0910 Mathematics Reporting Categories

Mathematical Processes  
Number and Operations  
Algebra

Geometry and Measurement  
Data Analysis, Statistics, & Probability

## TN 0910 HS End-Of-Course Algebra Reporting Categories

Standard 1: Mathematical Processes  
Standard 2: Number and Operations  
Standard 3: Algebra

Standard 4: Geometry and Measurement  
Standard 5: Data Analysis, Statistics, & Probability

## TN 0910 Science Reporting Categories

Inquiry and Technology  
Life Science

Earth and Space Science  
Physical Science

## TN 0910 HS End-Of-Course Biology Reporting Categories

Embedded Inquiry  
Embedded Technology & Engineering  
Embedded Mathematics  
Standard 1: Cells

Standard 2: Interdependence  
Standard 3: Flow of Matter and Energy  
Standard 4: Heredity  
Standard 5: Biodiversity & Change





# Discovery Education Assessment RESEARCH

## CRITERION VALIDITY

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

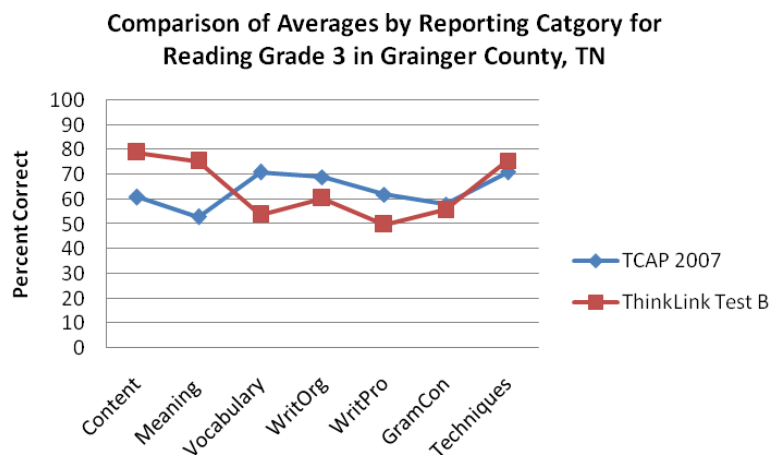
**Criterion validity** evidence demonstrates that test scores predict scores on an important criterion variable, such as a state's standardized large-scale assessment. Scientifically-based research provided evidence that Discovery Education Predictive Assessments matched TCAP difficulty levels across reporting categories, grades, and subjects.

#### Grainger County, Tennessee, Criterion Validity Study

The Grainger County school system participated in a criterion validity study for the 2006-2007 school year. Approximately 1500 students in grades 3 through 8 completed the Discovery Education Predictive Assessments. For each grade and subject, the students' percent correct averages were broken down by reporting category and compared to their respective 2007 TCAP averages.

Figure 1-18 feature comparisons made between Discovery Education Assessment Test B (Spring 2007) and actual 2007 TCAP reporting category averages. The results featured an average of 264 students per grade level for Reading and Mathematics and an average of 225 students per grade level for Science.

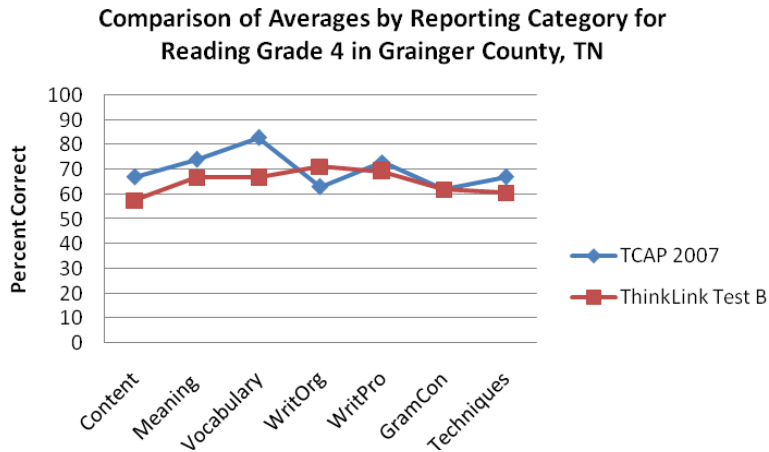
**Figure 1: Reading Reporting Category Averages for Discovery Education Assessment Test B and 2007 TCAP.**



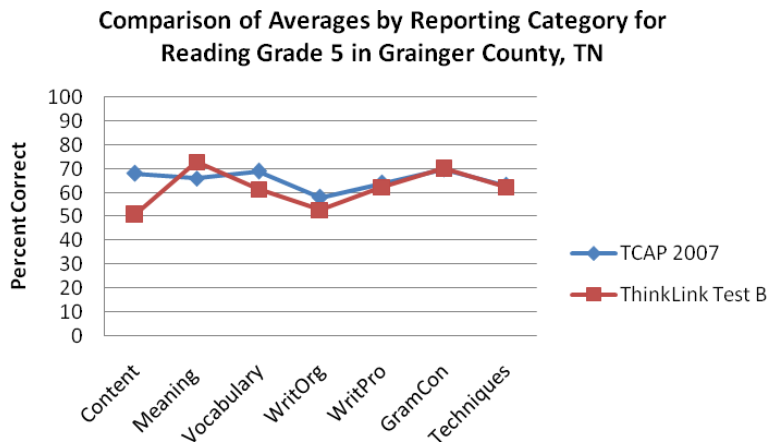


# Discovery Education Assessment RESEARCH

**Figure 2: Reading Reporting Category Averages for Discovery Education Assessment Test B and 2007 TCAP.**



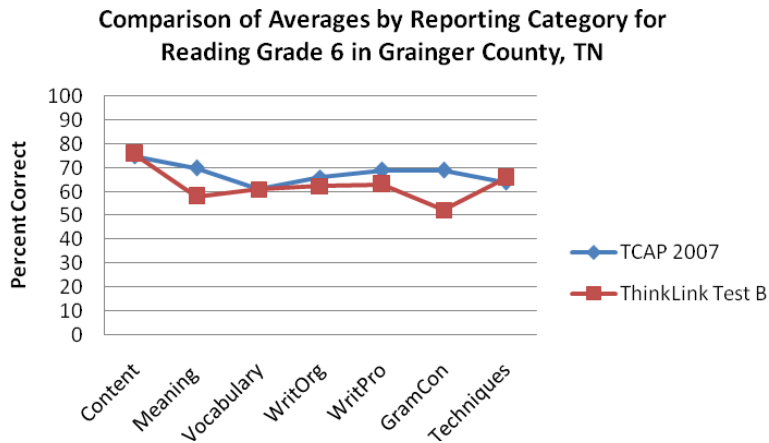
**Figure 3: Reading Reporting Category Averages for Discovery Education Assessment Test B and 2007 TCAP.**



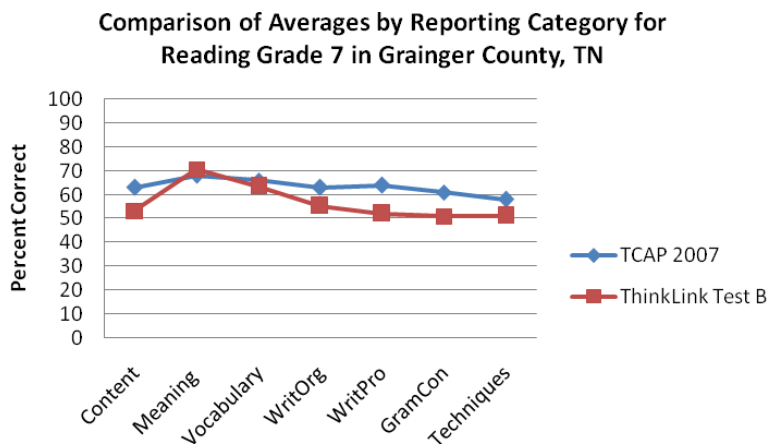


# Discovery Education Assessment RESEARCH

**Figure 4: Reading Reporting Category Averages for Discovery Education Assessment Test B and 2007 TCAP.**



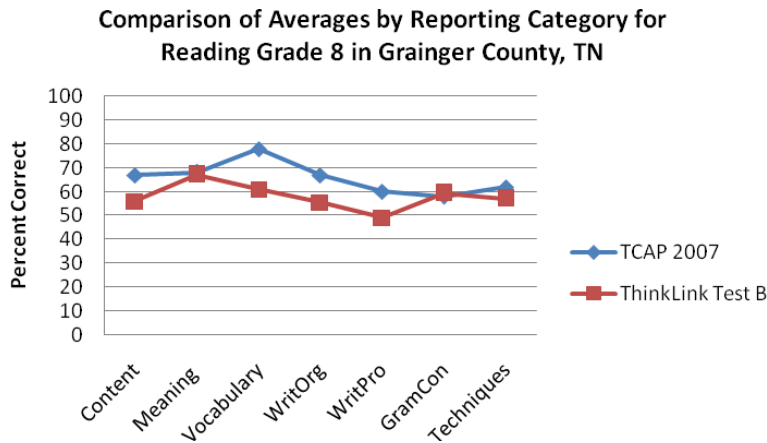
**Figure 5: Reading Reporting Category Averages for Discovery Education Assessment Test B and 2007 TCAP.**



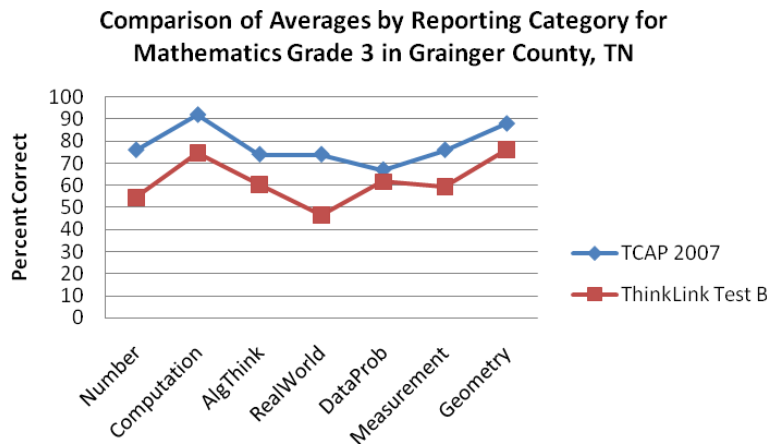


# Discovery Education Assessment RESEARCH

**Figure 6: Reading Reporting Category Averages for Discovery Education Assessment Test B and 2007 TCAP.**



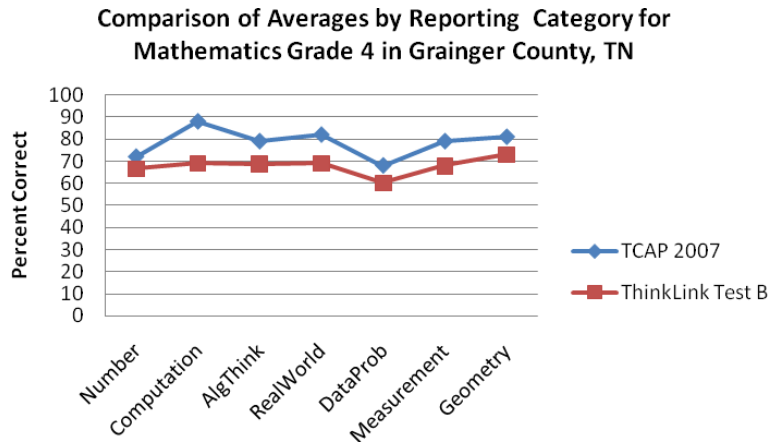
**Figure 7: Mathematics Reporting Category Averages for Discovery Education Assessment Test B and 2007 TCAP.**



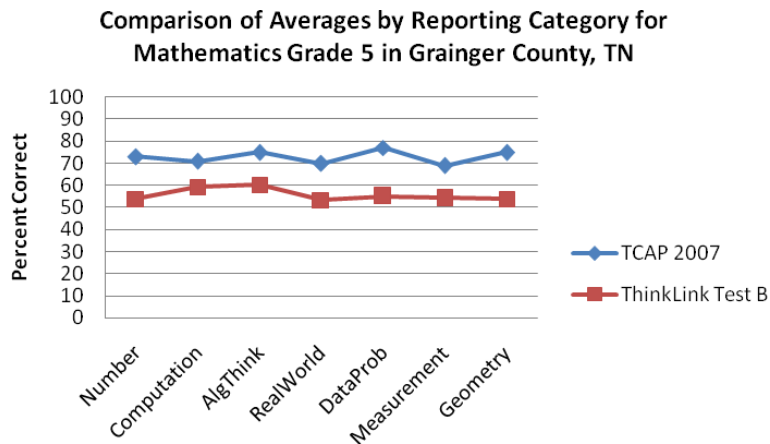


# Discovery Education Assessment RESEARCH

**Figure 8: Mathematics Reporting Category Averages for Discovery Education Assessment Test B and 2007 TCAP.**



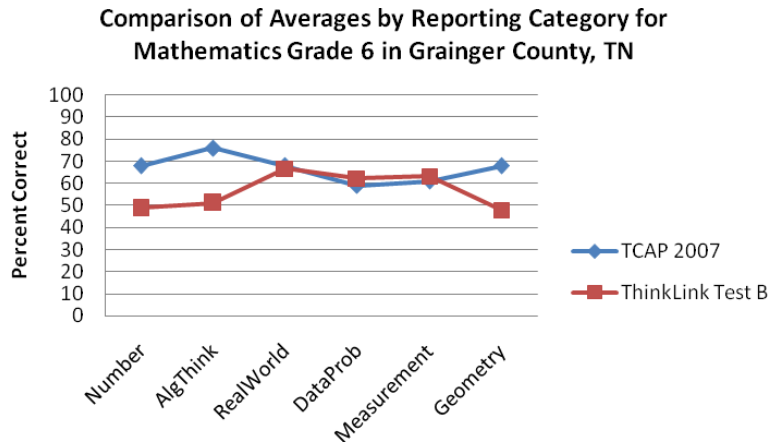
**Figure 9: Mathematics Reporting Category Averages for Discovery Education Assessment Test B and 2007 TCAP.**



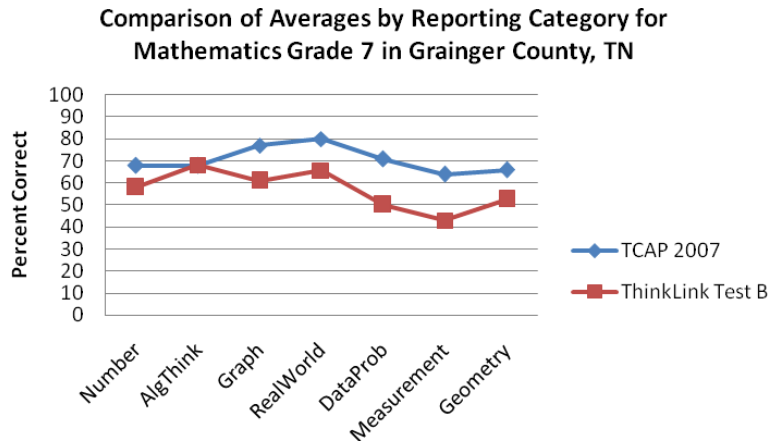


# Discovery Education Assessment RESEARCH

**Figure 10: Mathematics Reporting Category Averages for Discovery Education Assessment Test B and 2007 TCAP.**



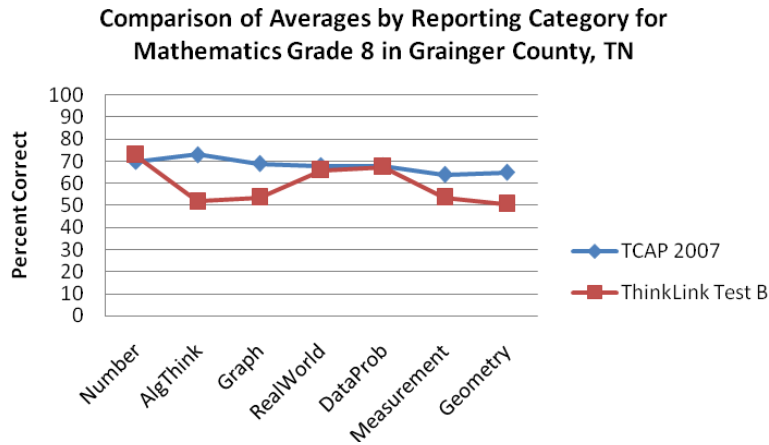
**Figure 11: Mathematics Reporting Category Averages for Discovery Education Assessment Test B and 2007 TCAP.**



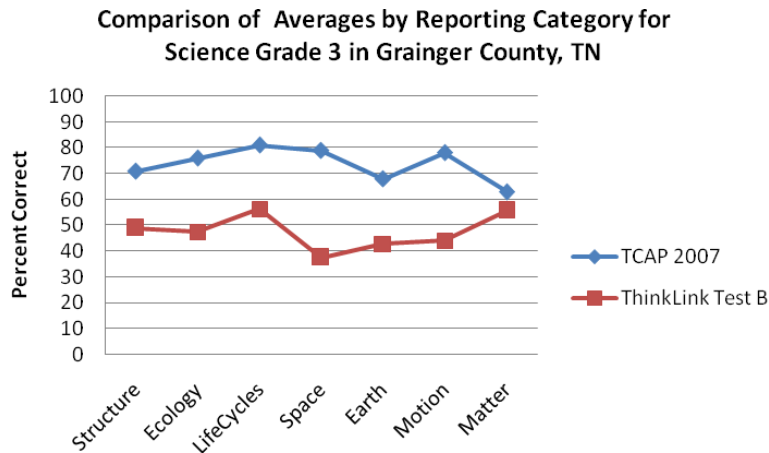


# Discovery Education Assessment RESEARCH

**Figure 12: Mathematics Reporting Category Averages for Discovery Education Assessment Test B and 2007 TCAP.**



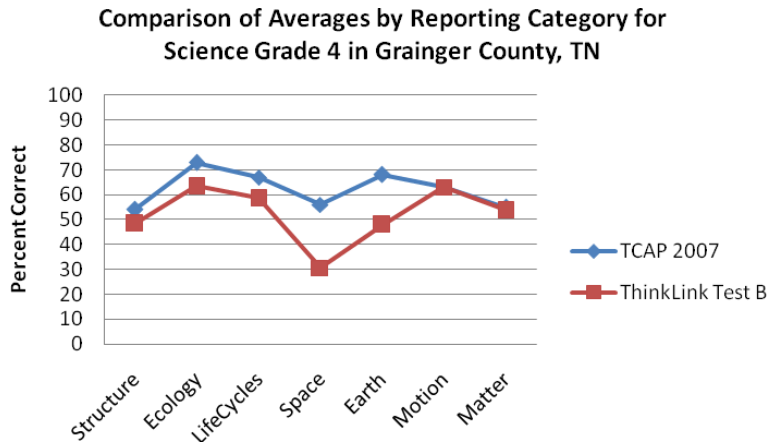
**Figure 13: Science Reporting Category Averages for Discovery Education Assessment Test B and 2007 TCAP.**



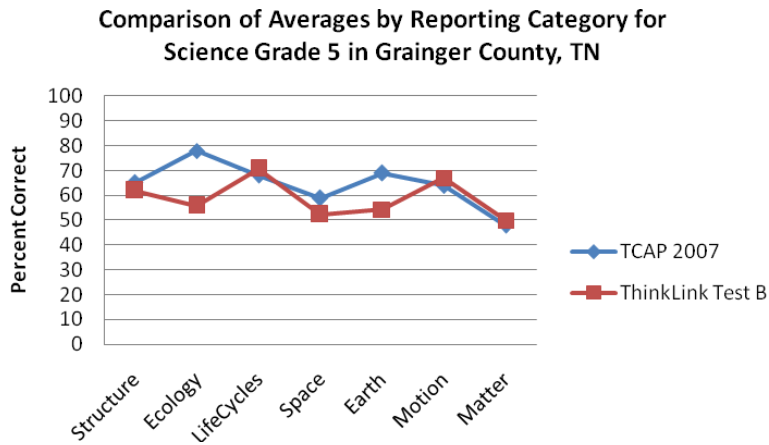


# Discovery Education Assessment RESEARCH

**Figure 14: Science Reporting Category Averages for Discovery Education Assessment Test B and 2007 TCAP.**



**Figure 15: Science Reporting Category Averages for Discovery Education Assessment Test B and 2007 TCAP.**

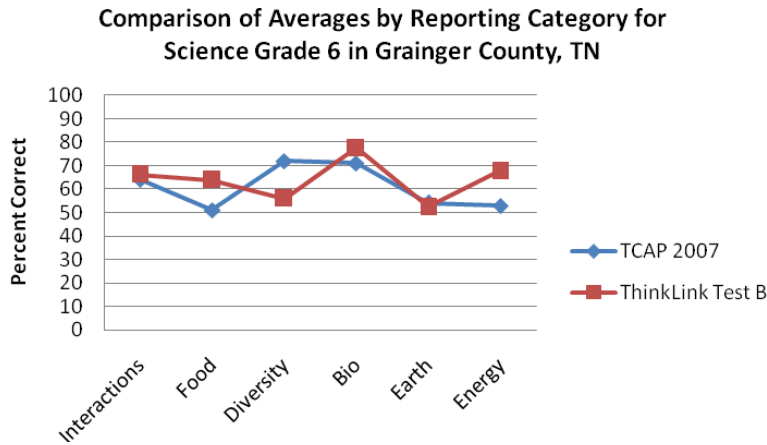




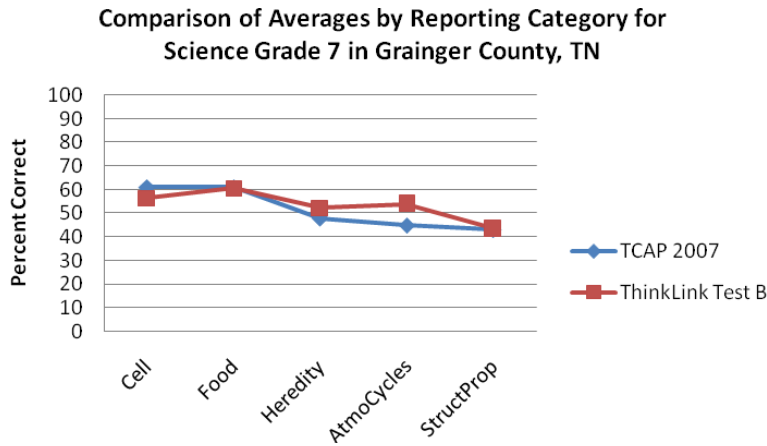


# Discovery Education Assessment RESEARCH

**Figure 16: Science Reporting Category Averages for Discovery Education Assessment Test B and 2007 TCAP.**



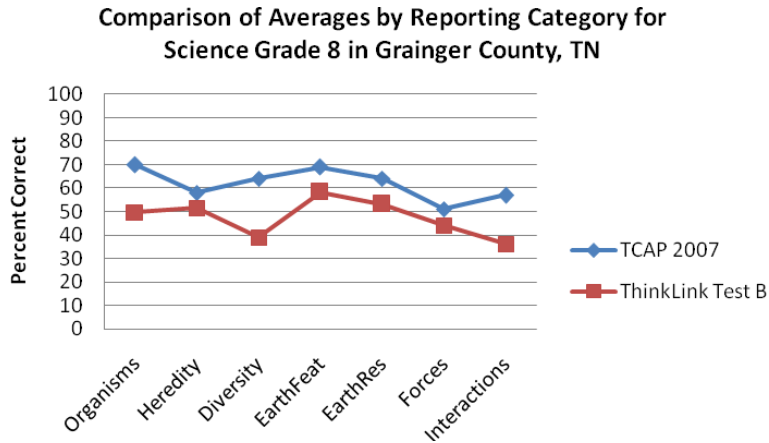
**Figure 17: Science Reporting Category Averages for Discovery Education Assessment Test B and 2007 TCAP.**





# Discovery Education Assessment RESEARCH

**Figure 18: Science Reporting Category Averages for Discovery Education Assessment Test B and 2007 TCAP.**





# Discovery Education Assessment RESEARCH

## Memphis City Criterion Validity Study for 2008-2009

Over 40,000 students in the Memphis City school system used Discovery Education Assessments during the 2008-2009 school year. Most students took three DEA assessments: Test A, Test B, and Test C. The Memphis City school provided DEA with TCAP scores for all students from the 2009 Spring TCAP assessment.

Tables 4 to 6 presents test averages on the three DEA assessments and the corresponding Spring 2009 test average for Grades 3 to 8 Reading and Mathematics. Correlations between DEA and TCAP test averages are also presented.

DEA assessments significantly predict TCAP test averages (all correlations are significant at  $p < .01$ ). These predictions hold for Grades 3 to 8 across the two subject areas of Reading and Mathematics.

Figure 19 present the averages for DEA Test C and TCAP for Grades 3 to 8 Reading. Figure 20 presents averages for DEA Test C and TCAP for Grades 3 to 8 Mathematics.

**Table 4: Memphis Discovery Education Assessment & TCAP Reading and Math averages for 0809.**

Grade	Test Period	Memphis Reading Test Averages			Memphis Math Test Averages		
		DEA	TCAP	Correlation	DEA	TCAP	Correlation
3	A	50.99	58.23	0.66	55.98	68.16	0.66
	B	56.22	58.24	0.76	53.97	68.15	0.66
	C	56.17	58.34	0.70	59.09	68.80	0.68
4	A	50.80	55.93	0.69	53.81	67.05	0.64
	B	54.40	55.94	0.71	53.44	67.09	0.68
	C	54.45	55.99	0.70	54.78	67.65	0.69
5	A	56.80	57.24	0.67	51.89	63.92	0.68
	B	53.19	57.32	0.70	55.96	63.88	0.70
	C	52.81	57.29	0.70	45.66	64.40	0.67
6	A	52.20	57.64	0.73	51.22	58.62	0.71
	B	47.91	57.50	0.72	45.29	58.55	0.68
	C	47.97	57.52	0.69	51.40	59.39	0.72
7	A	49.38	52.92	0.74	52.26	60.72	0.73
	B	45.35	52.78	0.71	46.20	60.40	0.74
	C	46.86	52.81	0.71	47.40	61.54	0.74
8	A	52.72	56.33	0.71	38.39	50.28	0.65
	B	51.31	56.35	0.72	43.20	50.20	0.69
	C	46.64	56.30	0.71	40.73	51.20	0.63
Average				0.71	Average		0.69



# Discovery Education Assessment RESEARCH

Figure 19: Memphis Discovery Education Assessment Test C & TCAP Reading averages for 0809.

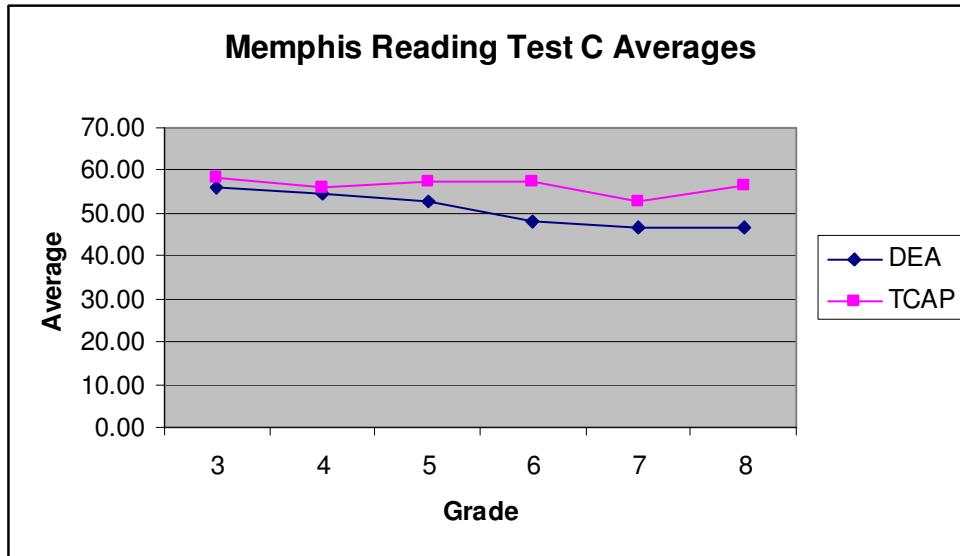
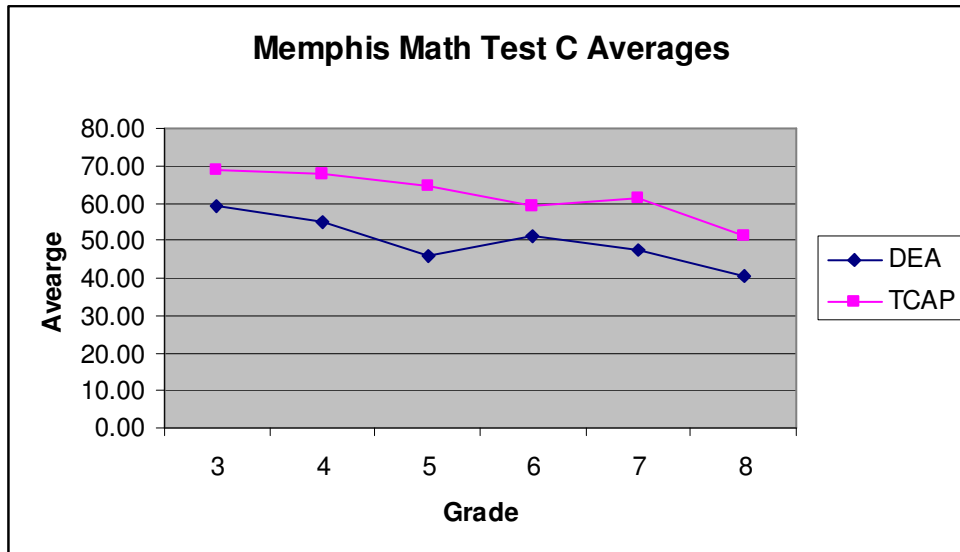


Figure 20: Memphis Discovery Education Assessment Test C & TCAP Math averages for 0809.





# 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 is 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 Tennessee 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%.***

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



# Discovery Education Assessment RESEARCH

correct classifications for every 1 misclassification. Discovery Education Assessment uses information from these scores to improve its benchmark assessments every year.

## Discovery Education Assessment Proficiency Predictions vs. TCAP Proficiency Predictions

Due to our representativeness throughout the state of Tennessee, direct comparisons of Spring 2009 (Test B) and actual 2009 TCAP 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 “Proficient” and “Advanced”. The results for all grades in Reading and Mathematics are presented in Table 4.

**Table 5: Proficiency Prediction Scores for Reading and Mathematics.**

Proficiency Prediction Scores for TN Statewide 2009								
	READING				MATHEMATICS			
	DEA Test B	TCAP 2009	Error Rate	Success Rate	DEA Test B	TCAP 2009	Error Rate	Success Rate
Grade 3	89.7	90	0.3	99.7	90.9	88	-2.9	97.1
Grade 4	89.5	90	0.5	99.5	90.8	90	-0.8	99.2
Grade 5	91.9	95	3.1	96.9	94.4	95	0.6	99.4
Grade 6	93.2	92	-1.2	98.8	91.9	90	-1.9	98.1
Grade 7	93.2	91	-2.2	97.8	92	90	-2.0	98
Grade 8	90.4	93	2.6	97.4	88.8	90	1.2	98.8
	Average		0.5	98.35	Average		-1.0	98.43

Figure 21 and 22 provide a direct comparison of statewide proficiency percentages for the combined category of “Proficient” and “Advanced” at each grade level. The average error statewide for Reading was .5% with a Proficiency Prediction Score of 99.5%. The average error statewide for Mathematics was 1.0% with a Proficiency Prediction Score of 99.0%.



# Discovery Education Assessment RESEARCH

Figure 21: Comparison of Discovery Education Assessment Test B vs. TCAP Proficiencies in Reading.

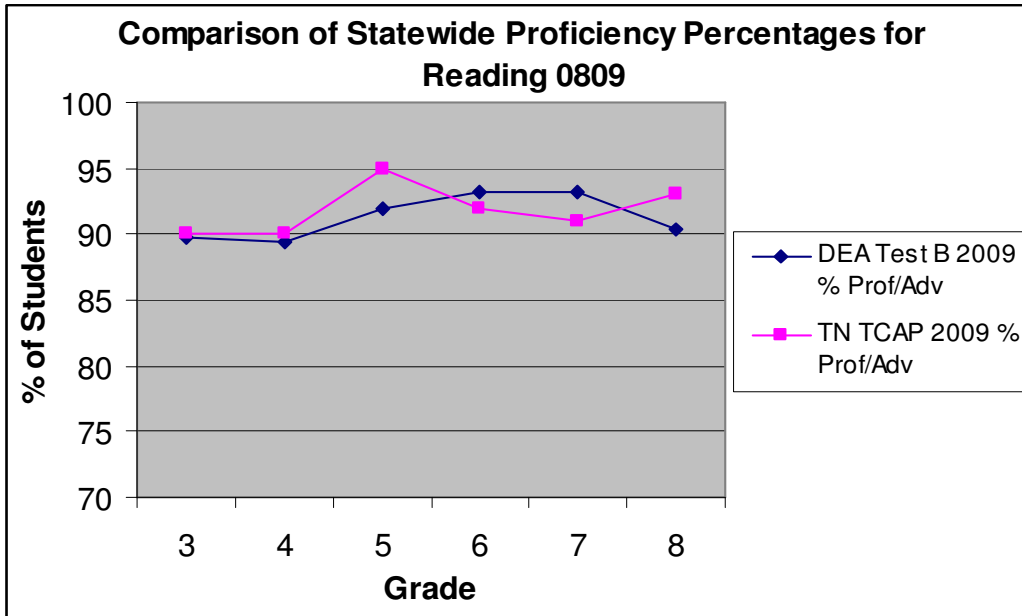
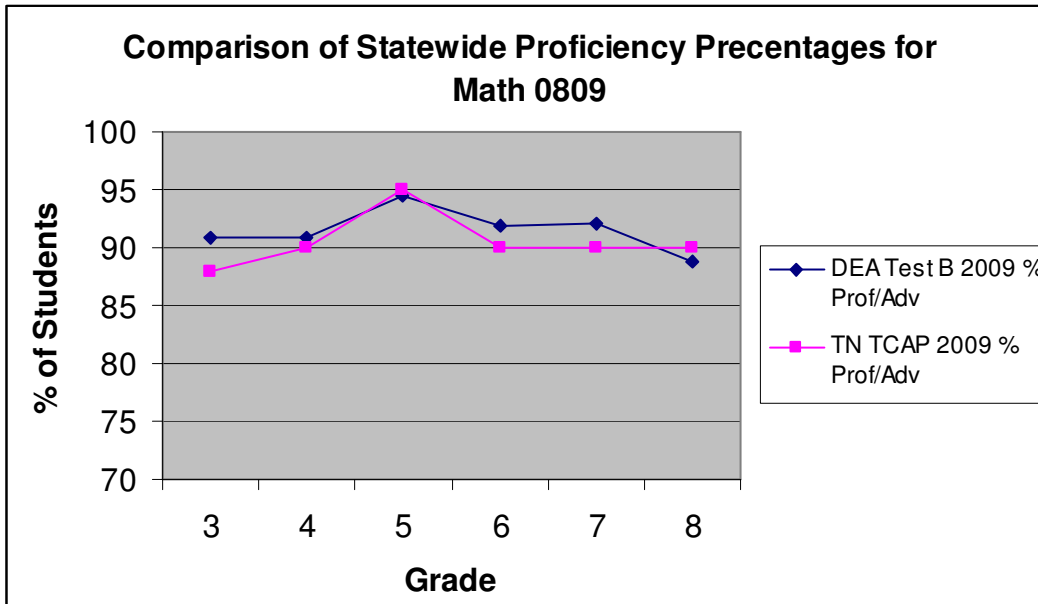


Figure 22: Comparison of Discovery Education Assessment Test B vs. TCAP Proficiencies in Mathematics.





# Discovery Education Assessment RESEARCH

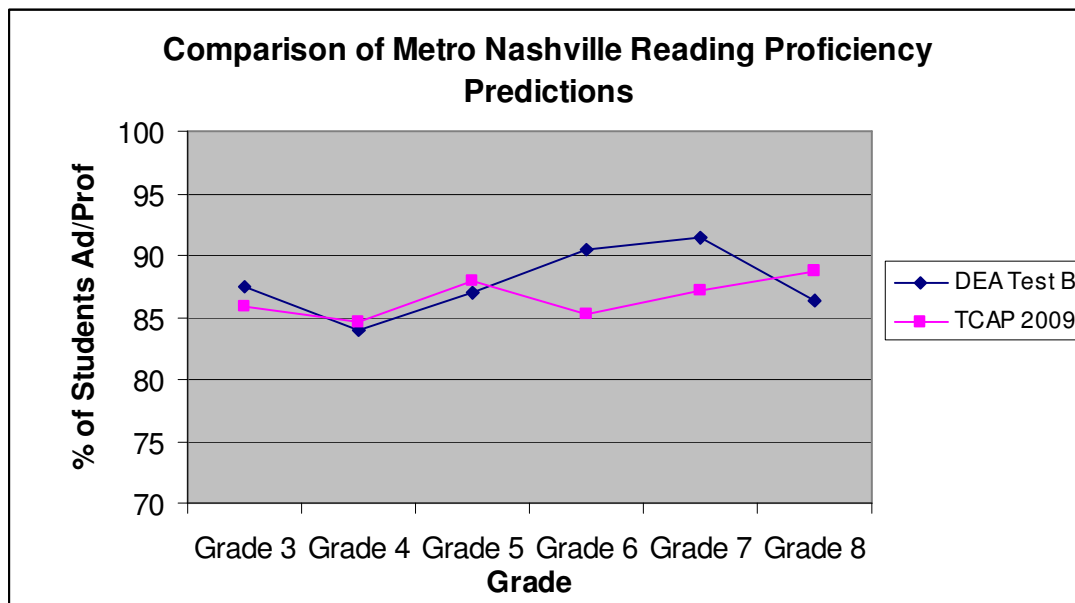
## Metro Nashville Proficiency Prediction Study

The same proficiency prediction scores were calculated for the Metro Nashville school system for 2008-2009. The results are presented in Table 6 and Figures 23 and 24. The average error rate for Reading was 1.2% with a Proficiency Prediction Score of 98.2%. The average error rate for Mathematics was 2.4% with a Proficiency Prediction Score of 97.60%.

**Table 6: Proficiency Prediction Scores for Reading and Mathematics in Metro Nashville 2009.**

Proficiency Prediction Scores for Metro Nashville 2009								
	READING				MATHEMATICS			
	DEA Test B	TCAP 2009	Error Rate	Success Rate	DEA Test B	TCAP 2009	Error Rate	Success Rate
Grade 3	87.4	85.8	1.6	98.4	89.2	86.3	2.9	97.1
Grade 4	84	84.6	-0.6	99.4	87.5	85.9	1.6	98.4
Grade 5	87	88	-1	99	90.8	89.7	1.1	98.9
Grade 6	90.4	85.2	5.2	94.8	88.6	84.9	3.7	96.3
Grade 7	91.5	87.1	4.4	95.6	88.1	82.5	5.6	94.4
Grade 8	86.4	88.7	-2.3	97.7	82.3	82.9	-0.6	99.4
	Average		1.2	97.48	Average		2.4	97.42

**Figure 23: Comparison of Discovery Education Assessment Test B vs. TCAP Reading Proficiencies in Metro Nashville.**

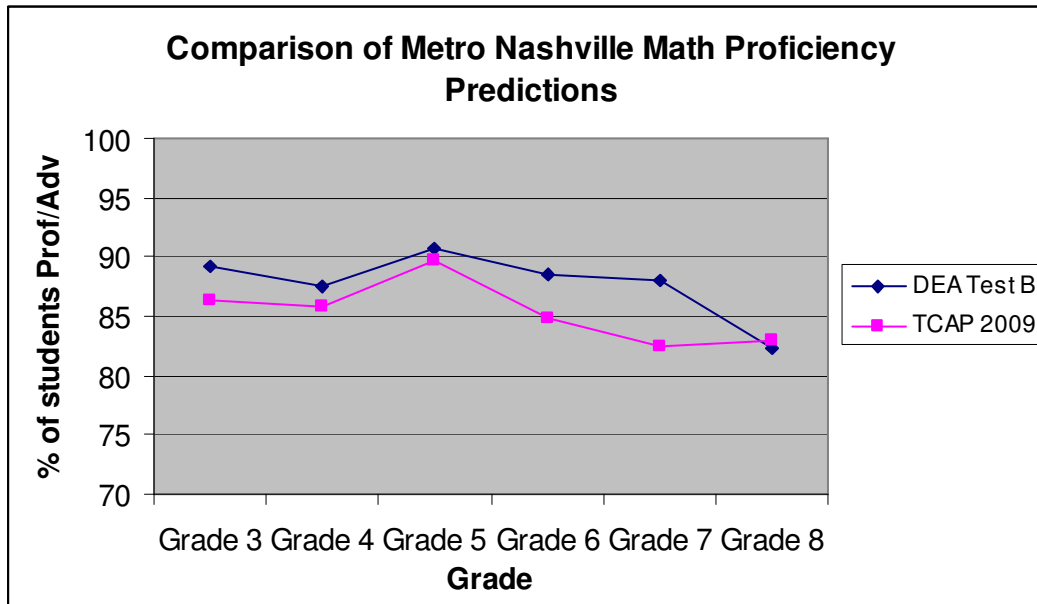






# Discovery Education Assessment RESEARCH

*Figure 24: Comparison of Discovery Education Assessment Test B vs. TCAP Math Proficiencies in Metro Nashville.*



### Shelby County Proficiency Prediction Study

The same proficiency prediction scores were calculated for the Shelby County school system for 2008-2009. The results are presented in Table 7 and Figures 25 and 26. The average error rate for Reading was 1.37% with a Proficiency Prediction Score of 98.63%. The average error rate for Mathematics was 4.82% with a Proficiency Prediction Score of 95.18.0%.

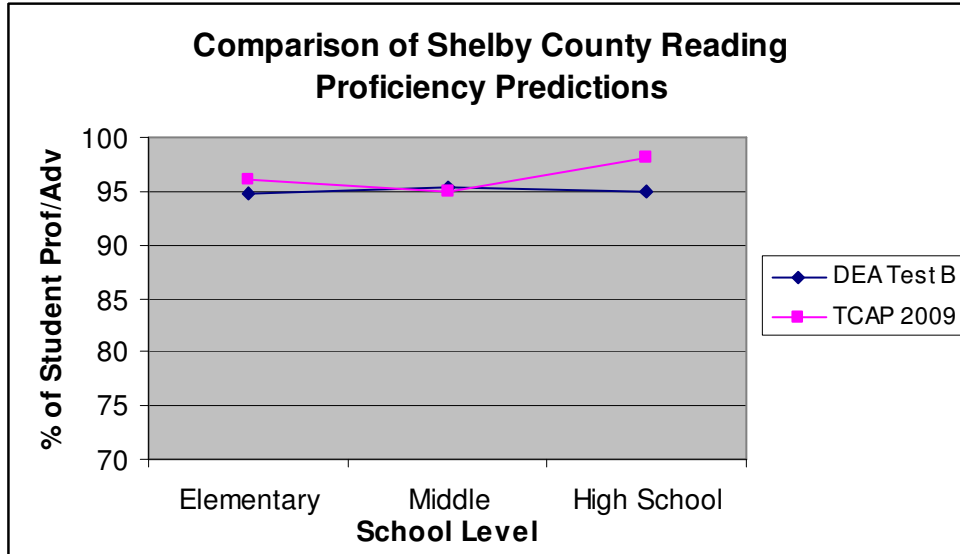
*Table 7: Proficiency Prediction Scores for Reading and Mathematics in Shelby County, TN 2009.*

	Proficiency Prediction Scores for Shelby County, TN 2009							
	READING				MATHEMATICS			
	DEA Test B	TCAP 2009	Error Rate	Success Rate	DEA Test B	TCAP 2009	Error Rate	Success Rate
Elementary	94.8	96.1	1.3	98.7	95.96	95.06	-0.9	99.1
Middle	95.27	94.95	-0.32	99.68	94.27	93.75	-0.52	99.48
High School	95	98.13	3.13	96.87	74.88	90.75	15.87	84.13
		Average	1.37	98.42		Average	4.82	94.24

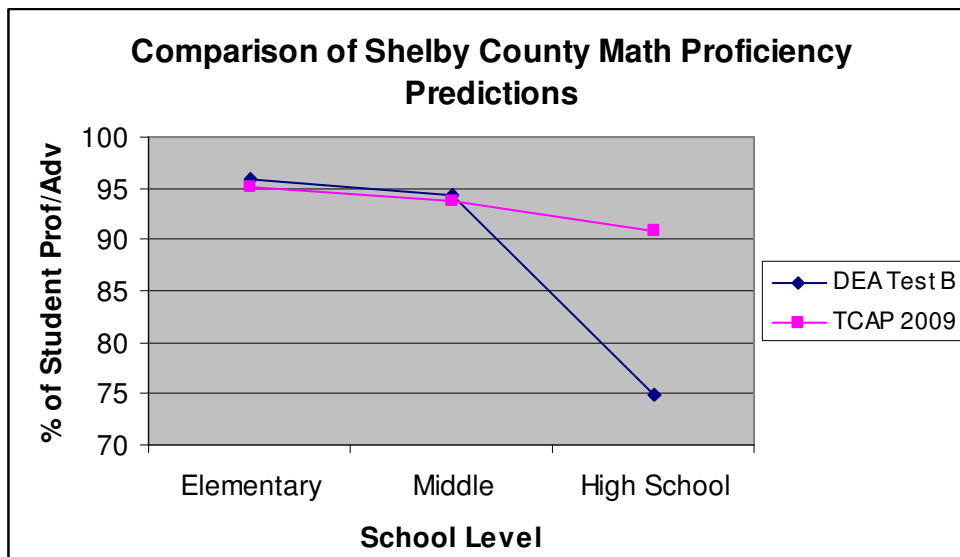


# Discovery Education Assessment RESEARCH

*Figure 25: Comparison of Discovery Education Assessment Test B vs. TCAP Reading Proficiencies in Shelby County, TN.*



*Figure 26: Comparison of Discovery Education Assessment Test B vs. TCAP Math Proficiencies in Shelby County, TN.*





# Discovery Education Assessment RESEARCH

## Memphis City Proficiency Prediction Study

The same proficiency prediction scores were calculated for the Memphis City County school system for 2008-2009. The results are presented in Tables 8 and 9 and Figures 27 to 29. The average error rate for Reading was 5.76% with a Proficiency Prediction Score of 94.24%. The average error rate for Mathematics was 3.23% with a Proficiency Prediction Score of 96.77.0%.

**Table 8: Proficiency Prediction Scores for Reading and Mathematics in Memphis, TN 0809.**

		Proficiency Prediction Scores for Memphis, TN 0809							
		Reading				Math			
Grade	Test Period	Discovery	TCAP	Error Rate	Success Rate	Discovery	TCAP	Error Rate	Success Rate
3	A	90.2	79.77	<b>10.44</b>	<b>89.56</b>	87.34	80.91	<b>6.43</b>	<b>93.57</b>
	B	90.44	79.85	<b>10.59</b>	<b>89.41</b>	85.22	80.89	<b>4.33</b>	<b>95.67</b>
	C	88.23	80.16	<b>8.07</b>	<b>91.93</b>	86.93	81.73	<b>5.2</b>	<b>94.8</b>
4	A	92	78.01	<b>13.99</b>	<b>86.01</b>	91.02	84.33	<b>6.7</b>	<b>93.3</b>
	B	91.05	77.99	<b>13.06</b>	<b>86.94</b>	88.63	84.31	<b>4.32</b>	<b>95.68</b>
	C	87.49	78.01	<b>9.48</b>	<b>90.52</b>	90.02	84.88	<b>5.14</b>	<b>94.86</b>
5	A	91.78	88.44	<b>3.34</b>	<b>96.66</b>	93.73	92.9	<b>0.82</b>	<b>99.18</b>
	B	92.38	88.49	<b>3.89</b>	<b>96.11</b>	91.88	92.81	<b>-0.93</b>	<b>99.07</b>
	C	91.94	88.5	<b>3.44</b>	<b>96.56</b>	89.61	93.16	<b>-3.55</b>	<b>96.45</b>
6	A	87.39	82.59	<b>4.8</b>	<b>95.2</b>	86.67	84.7	<b>1.97</b>	<b>98.03</b>
	B	92.01	82.34	<b>9.68</b>	<b>90.32</b>	86.11	84.57	<b>1.54</b>	<b>98.46</b>
	C	92.4	82.48	<b>9.91</b>	<b>90.09</b>	89.86	85.73	<b>4.14</b>	<b>95.86</b>
7	A	83.9	83.05	<b>0.84</b>	<b>99.16</b>	89.2	85.13	<b>4.07</b>	<b>95.93</b>
	B	80.79	82.68	<b>-1.89</b>	<b>98.11</b>	86.59	84.35	<b>2.24</b>	<b>97.76</b>
	C	81.56	82.96	<b>-1.4</b>	<b>98.6</b>	89.03	86.27	<b>2.76</b>	<b>97.24</b>
8	A	88.26	85.51	<b>2.75</b>	<b>97.25</b>	87.25	80.29	<b>6.96</b>	<b>93.04</b>
	B	87.84	85.62	<b>2.22</b>	<b>97.78</b>	83.81	80.61	<b>3.2</b>	<b>96.8</b>
	C	86.33	85.8	<b>0.52</b>	<b>99.48</b>	85.2	82.47	<b>2.72</b>	<b>97.28</b>
			Average	5.76	93.87		Average	3.23	96.28

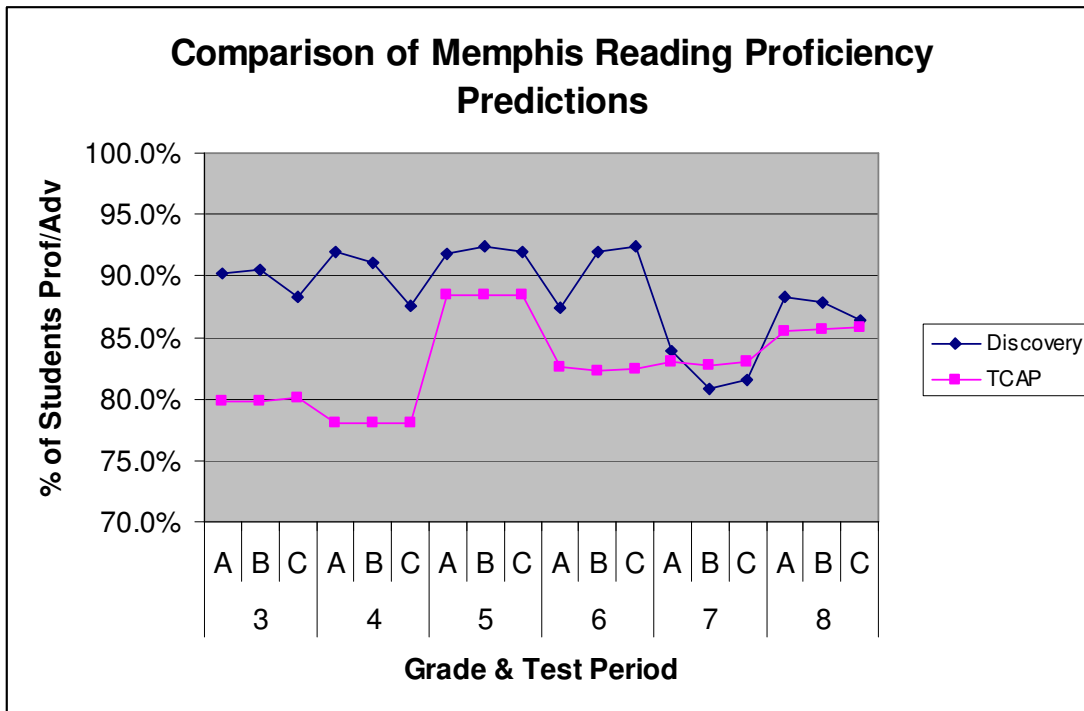


# Discovery Education Assessment RESEARCH

**Table 9: Proficiency Prediction Scores for High School Algebra & English in Memphis, TN 0809.**

Proficiency Prediction Scores for HS Memphis, TN 0809					
Subject	Test Period	Discovery	TCAP	Error Rate	Success Rate
Algebra	A	79.2	78.1	1.1	98.9
	B	61.8	77.6	-15.8	84.2
	C	63.9	77.4	-13.5	86.5
English II	A	90.6	96.1	-5.5	94.5
	B	95	96	-1	99
	C	96.1	96.3	-0.2	99.8
			Average	-5.8	93.82

**Figure 27: Comparison of Discovery Education Assessment vs. TCAP Reading Proficiencies in Memphis.**





# Discovery Education Assessment RESEARCH

Figure 28: Comparison of Discovery Education Assessment vs. TCAP Math Proficiencies in Memphis.

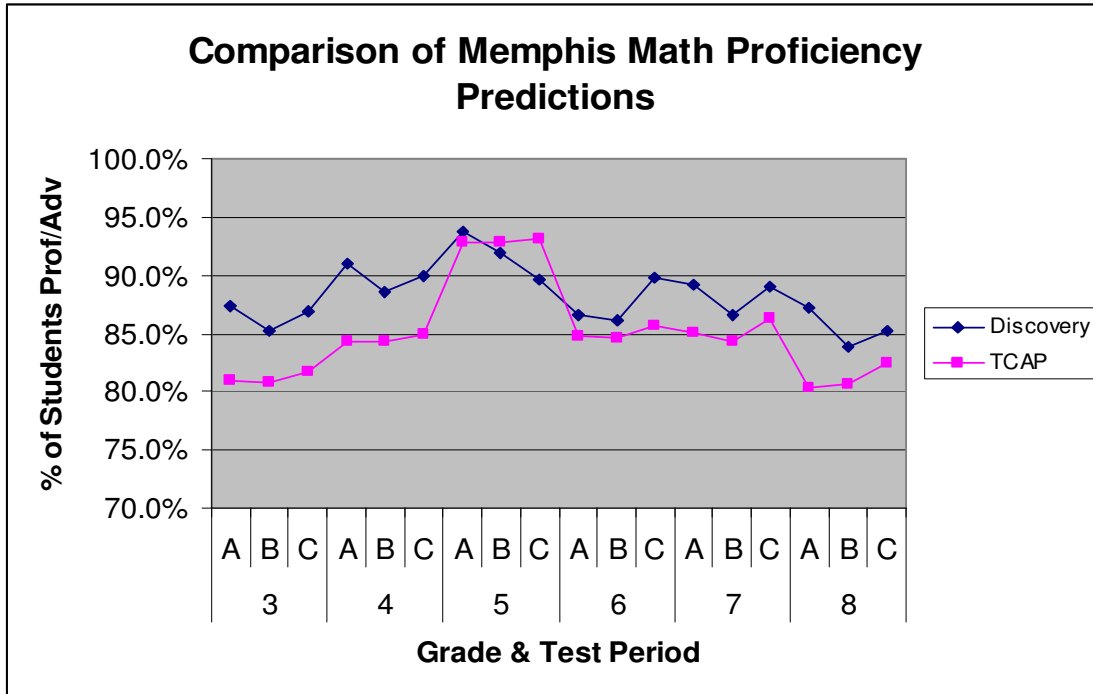
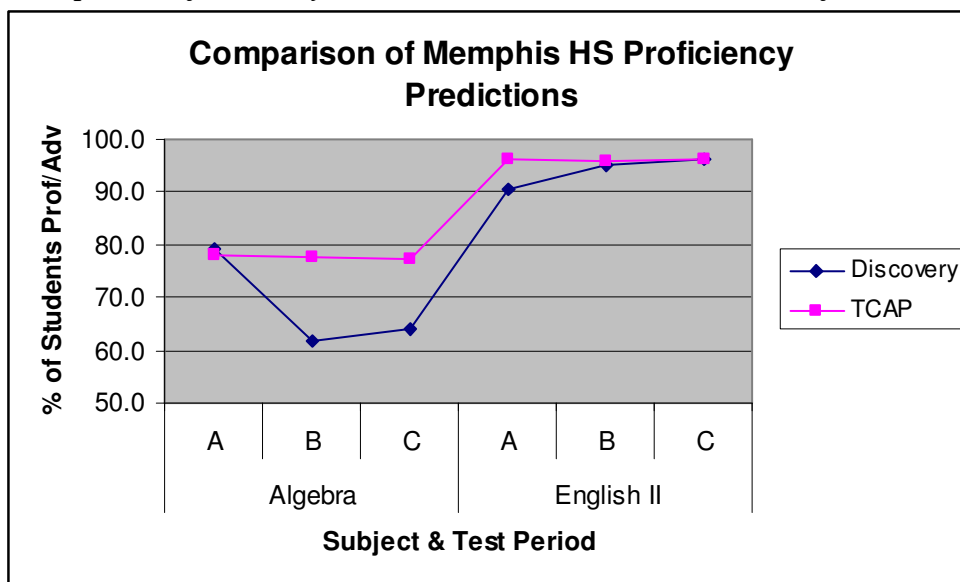


Figure 29: Comparison of Discovery Education Assessment vs. TCAP HS Proficiencies in Memphis.





# Discovery Education Assessment RESEARCH

## Proficiency Success Rate

When individual student data are available, an additional measure, the *Proficiency Success Rate*, can also be calculated. After taking a Discovery Education Benchmark Assessment, a student receives a prediction of his or her proficiency status: *Proficient* (Meets or Exceeds) or *Not Proficient* (Below or Warning). When ISAT and PSAE results are received, a comparison of this prediction with actual ISAT or PSAE status can be made. The percentage of students predicted as proficient by Discovery Education Assessment that actually are proficient on the ISAT or PSAE is called the Proficiency Success Rate. For instance, a Proficiency Success Rate of 90% indicates that ninety percent of the students that Discovery Education predicted as proficient actually achieved this result on the ISAT or PSAE.

## Case Study: Memphis, TN

The Memphis School District also participated in proficiency success rate study during the 2008-2009 school year. Individual student proficiency scores were obtained for Reading and Mathematics in Grades 3 to 8 and compared with proficiency predictions on Discovery Education Predictive Assessments. Table 10 present the Proficiency Success Rates for Reading and Mathematics. The median Proficiency Success Rate for Reading was 87%, and the median Proficiency Success Rate for Mathematics was 89.9%.

**Table 10: Proficiency Success Rate for Memphis, TN**

Proficiency Success Rate		
Memphis DEA Test C with TCAP		
	Reading	Math
Grade 3	85.6	87.7
Grade 4	84.0	89.7
Grade 5	91.3	95.1
Grade 6	85.5	90.0
Grade 7	88.3	90.4
Grade 8	89.8	86.6
Median	87.0	89.9



# 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.

Once again, the Grainger County school system 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 “Proficient” and “Advanced” on the 2007 TCAP was tabulated and compared with the percent of students that were classified as “Proficient” and “Advanced” on the 2006 TCAP. 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 2007 and 2006 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 “Proficient” and “Advanced” in the state of Tennessee was compared to this Difference score.

The percentages are to be understood as follows. Take a look below at Grade 3 Mathematics. The percent of students proficient in 2006 was 87, and the percent proficient in 2007 was 93, a difference or improvement of 5% (using exact not rounded percentages). However, Grade 3 Mathematics in the state of Tennessee improved by only 1% during the same time. Therefore, the “Grainger ↔ TN State” calculation is actually 4%. That is, the Grainger County Grade 3 Reading classes improved 4% in the percent of students proficient compared to the state of Tennessee.

**Table 11: Results of Consequential Validity Study for Grainger County in Mathematics.**

Grainger County, TN Mathematics				
Grade	2006	2007	Difference*	Grainger ↔ TN State*
3	87%	93%	5%	4%
4	90%	96%	6%	4%
5	94%	97%	2%	2%
6	87%	92%	4%	4%
7	92%	91%	-1%	-2%
8	89%	92%	3%	0%

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



# Discovery Education Assessment RESEARCH

**Table 12: Results of Consequential Validity Study for Grainger County in Reading.**

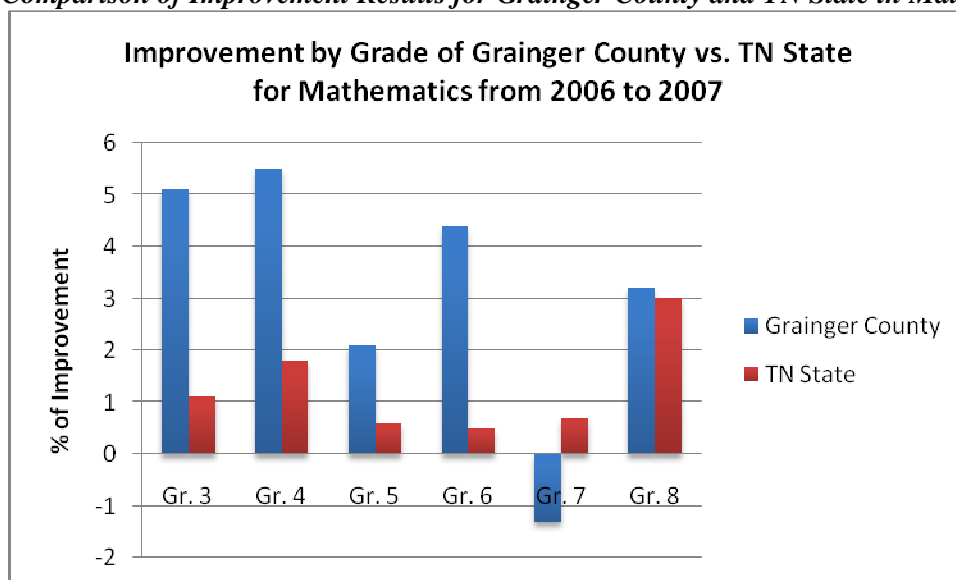
Grainger County, TN Reading				
Grade	2006	2007	Difference*	Grainger ↔ TN State*
3	87%	93%	5%	2%
4	92%	88%	-4%	-4%
5	92%	97%	5%	2%
6	91%	94%	3%	-2%
7	90%	93%	2%	0%
8	92%	94%	2%	0%

\*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. Grainger County had significant improvement (exceeding the state's improvement during the same time period) in Grades 3, 4, 5, 6, and 8 Mathematics and in Grades 3 and 4 Reading (see Figure 30 and 31).

**Figure 30: Comparison of Improvement Results for Grainger County and TN State in Mathematics.**

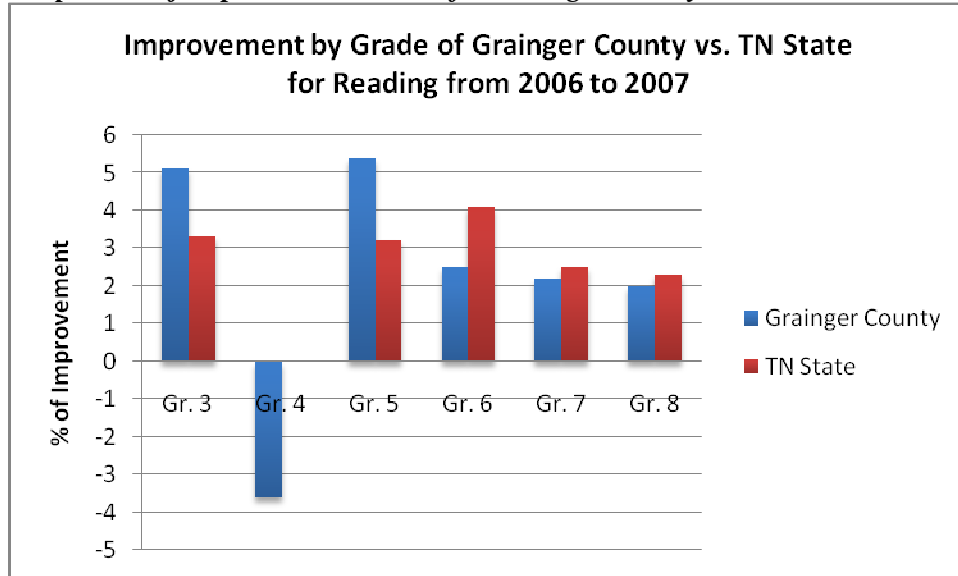






# Discovery Education Assessment RESEARCH

*Figure 31: Comparison of Improvement Results for Grainger County and TN State in Reading.*





# 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).

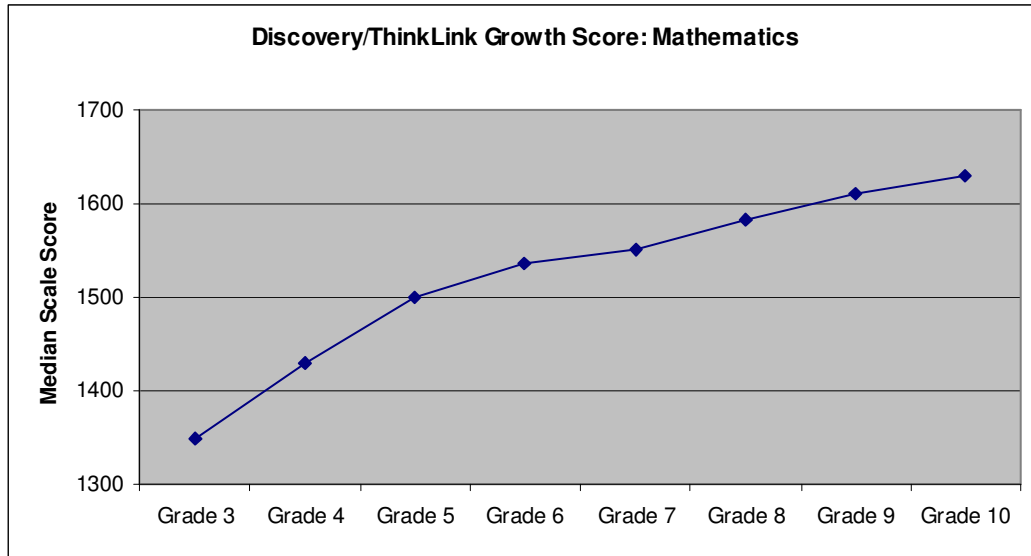
#### Why use such a rigorous vertical scale?

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 32 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.



# Discovery Education Assessment RESEARCH

*Figure 32: Vertically Scaled Growth Results for Discovery Education Assessment Mathematics Tests.*



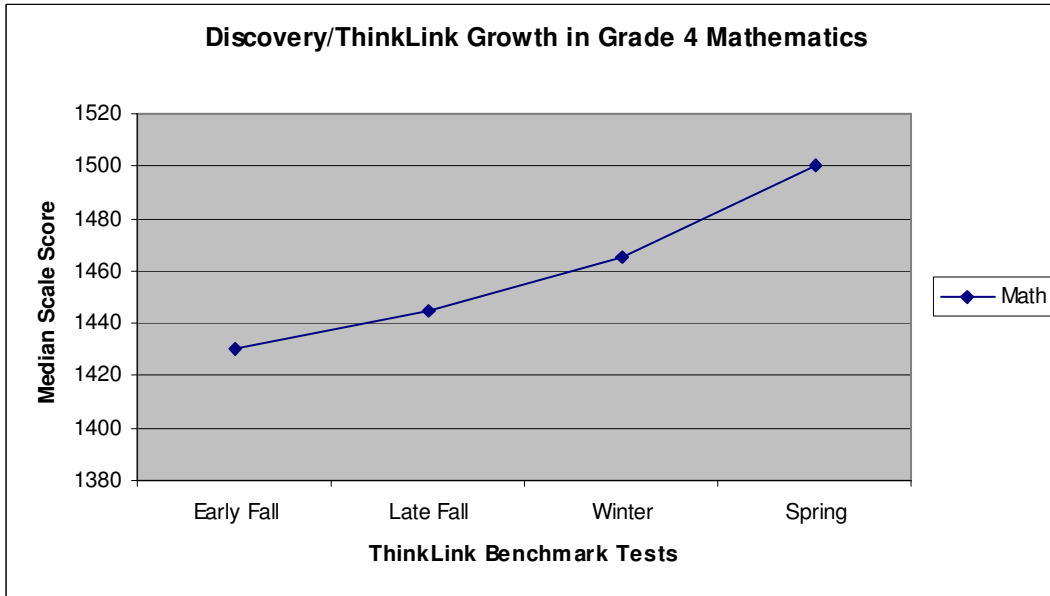
### **What is unique about the Discovery Education Assessment vertical growth scores?**

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 33 illustrates this growth for Grade 4 Mathematics using our benchmark assessments.



# Discovery Education Assessment RESEARCH

**Figure 33: Within-Year Growth Results for Discovery Education Assessment Mathematics Tests.**



### Tennessee Growth Scale

Tables 13 and 14 and Figures 34 and 35 illustrate student averages Discovery Education Assessment vertical growth scale for Grades 3 to 8 Reading and Mathematics tests between three time periods, Fall, Winter, and Spring 2008-2009.

**Table 13: Vertical Growth Score Comparisons for Fall & Winter 2008 and Spring 2009 in Reading.**

Tennessee 0809 Growth Score Comparisons									
Reading									
		Gr. 3	Gr. 4	Gr. 5	Gr. 6	Gr. 7	Gr. 8	High School	
Test P (Fall)		1452	1468	1529	1538	1591	1585		
Test A (Winter)		1436	1499	1559	1568	1596	1618	1595	
Test B (Spring)		1479	1533	1563	1587	1590	1628	1659	

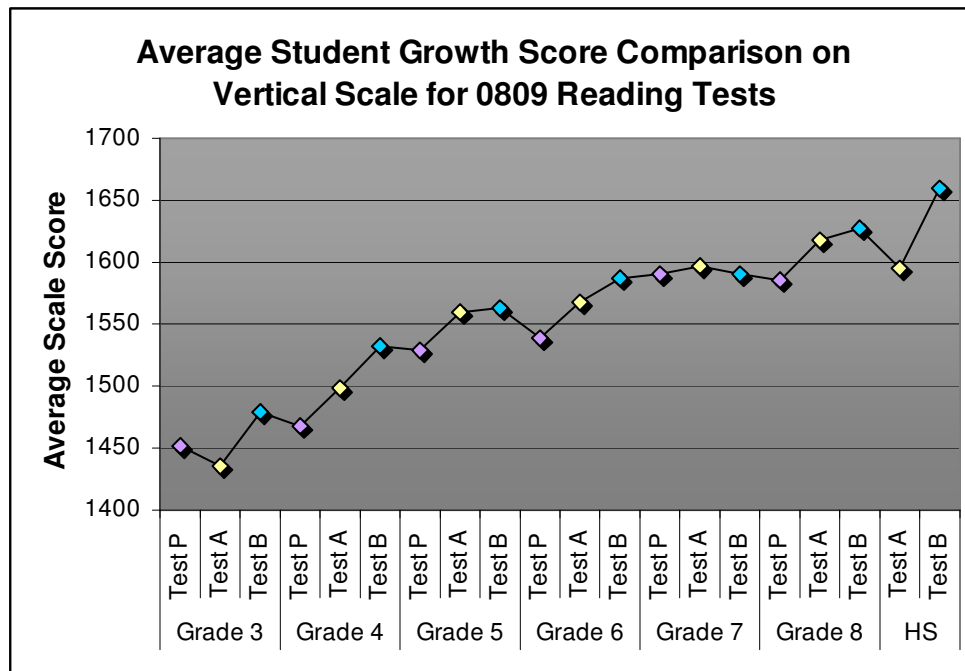


# Discovery Education Assessment RESEARCH

**Table 14: Vertical Growth Score Comparisons for Fall & Winter 2008 and Spring 2009 in Mathematics.**

Tennessee 0809 Growth Score Comparisons Mathematics									
		Gr. 3	Gr. 4	Gr. 5	Gr. 6	Gr. 7	Gr. 8	High School	
Test P (Fall)		1409	1437	1500	1561	1576	1613		
Test A (Winter)		1412	1477	1520	1584	1606	1606	1618	
Test B (Spring)		1413	1497	1541	1596	1614	1637	1623	

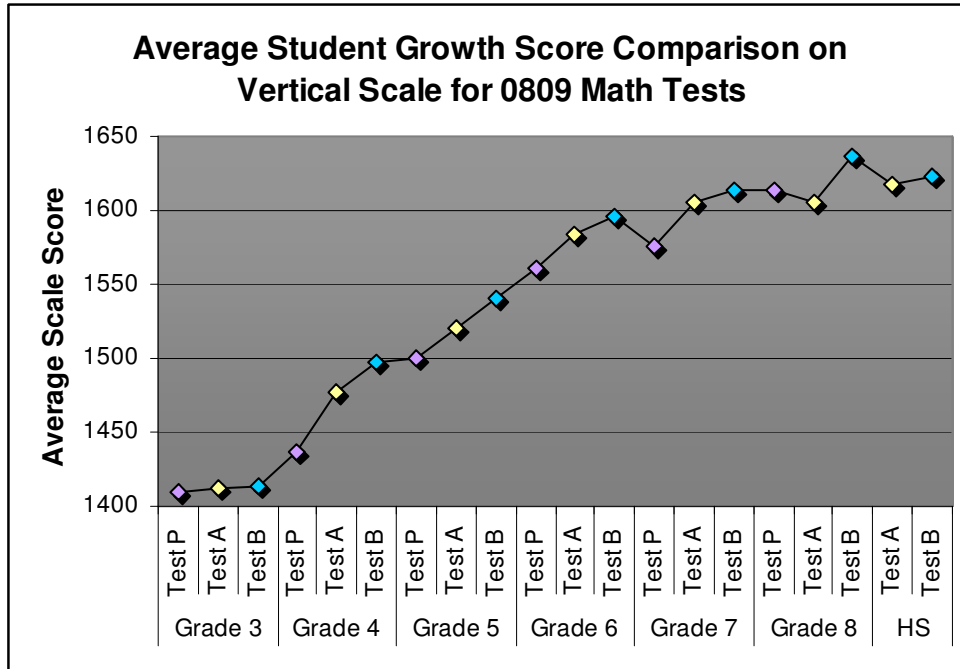
**Figure 34: Average Student Growth Score Comparison on Vertical Scale for 0809 Reading Tests.**





# Discovery Education Assessment RESEARCH

Figure 35: Average Student Growth Scores on Vertical Scale for 0809 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, 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

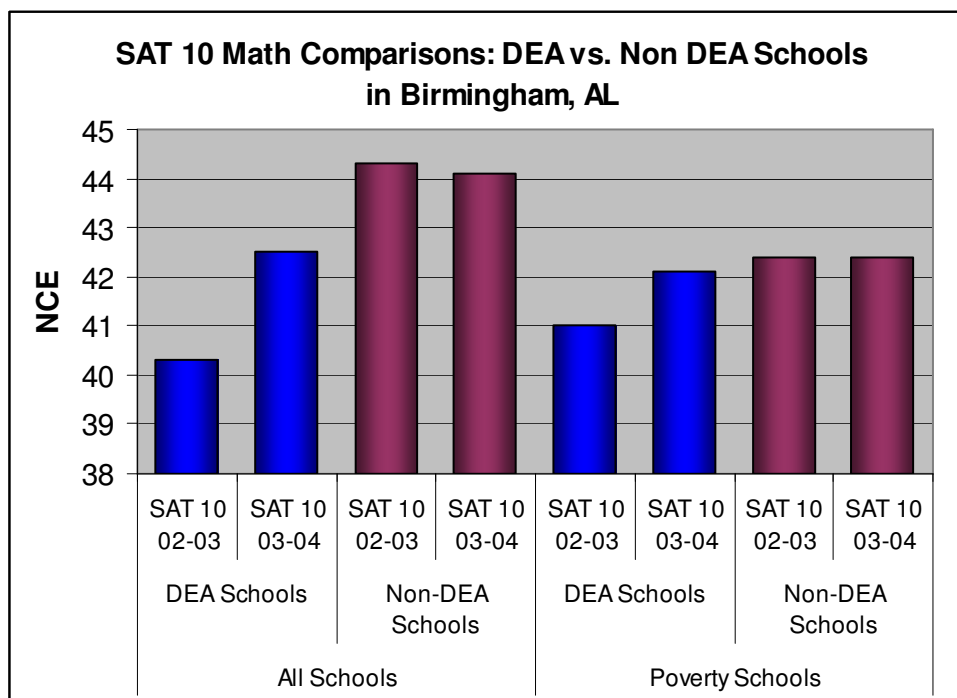


# Discovery Education Assessment RESEARCH

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 ThinkLink Schools vs. Non-ThinkLink 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-ThinkLink 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.

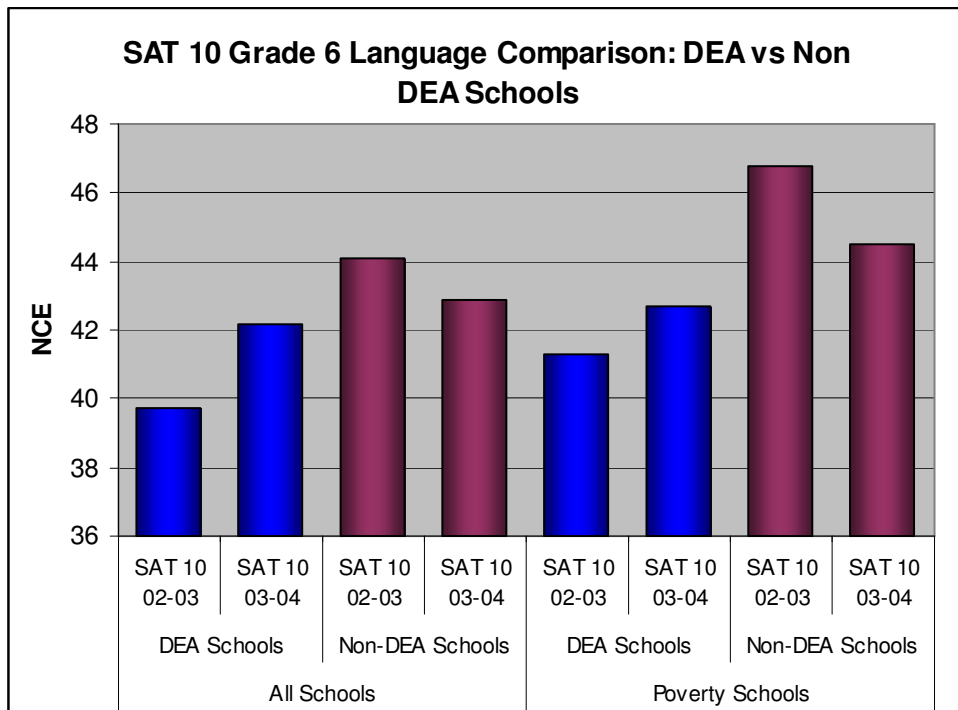
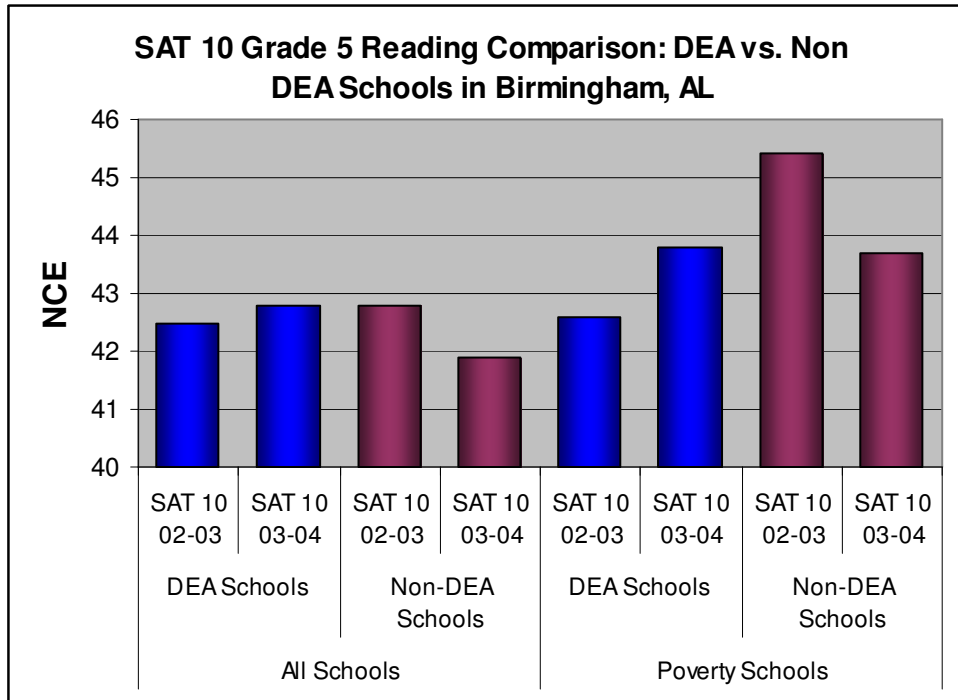
As a result of the improvement that many of these schools made in school year 2003-04, the Birmingham City Schools selected Discovery Education Assessment to be used with *all* of the schools in school year 2004-05. The Birmingham City Schools also chose to provide professional development in each school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







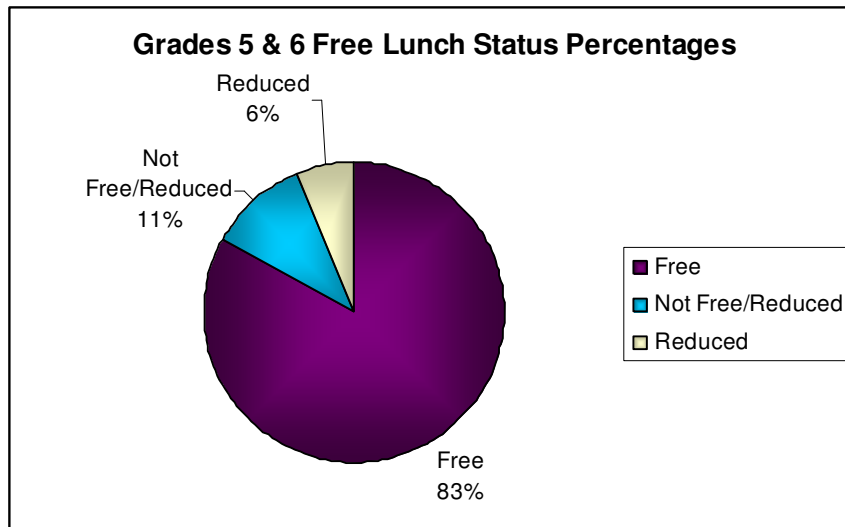
# Discovery Education Assessment RESEARCH





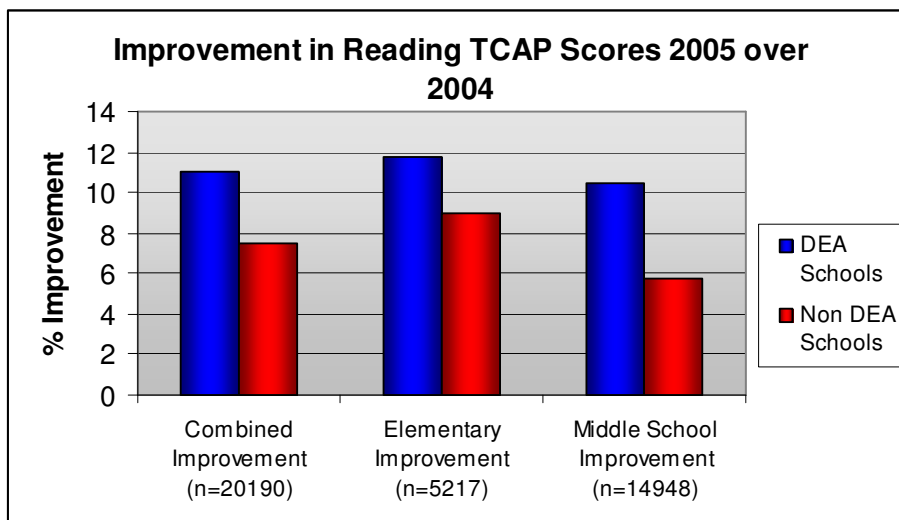
# Discovery Education Assessment RESEARCH

The following pie graph shows the Lunch Status percentages provided by Birmingham, AL school system for grades 5<sup>th</sup> and 6<sup>th</sup>.



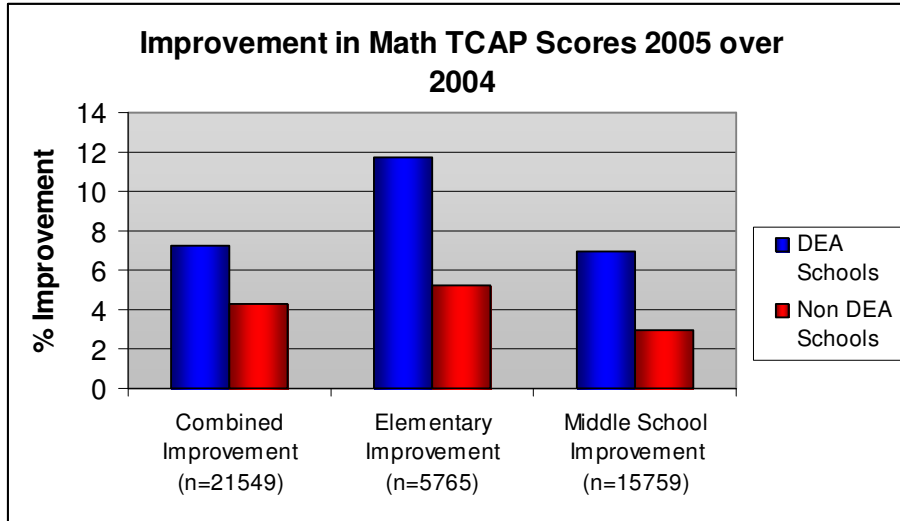
## Case Study Two: Metro Nashville, Tennessee City Schools

*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 ThinkLink schools versus non-ThinkLink 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.

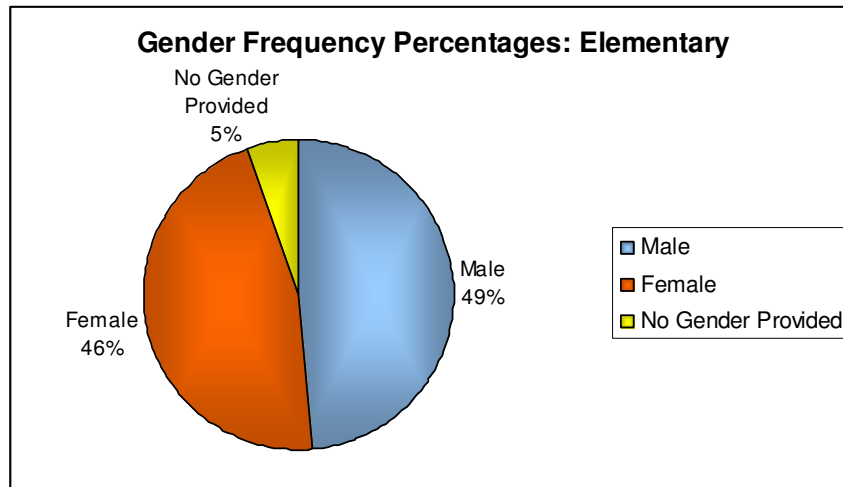




# Discovery Education Assessment RESEARCH

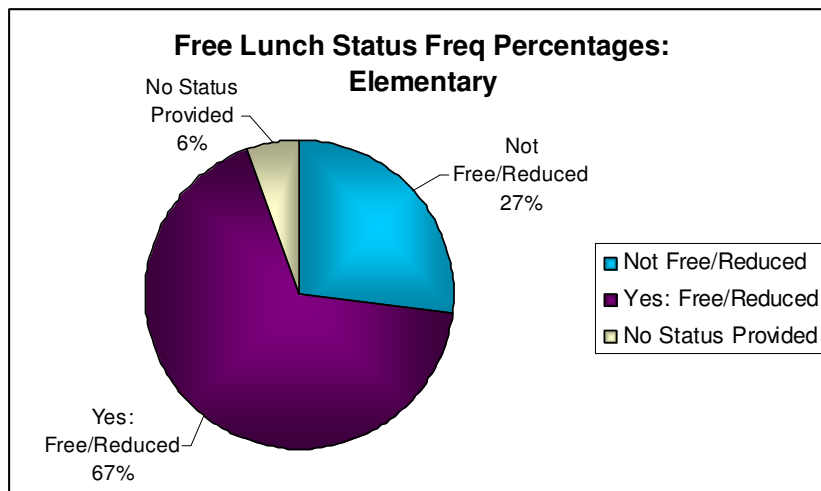
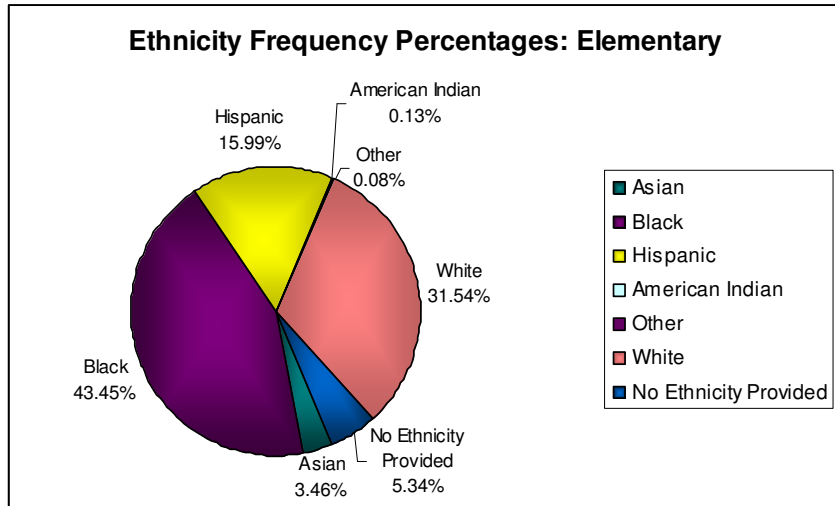


The following pie charts display the frequency percents of the NCLB data provided to DEA from the Metro Nashville Public School System for the Elementary Schools.



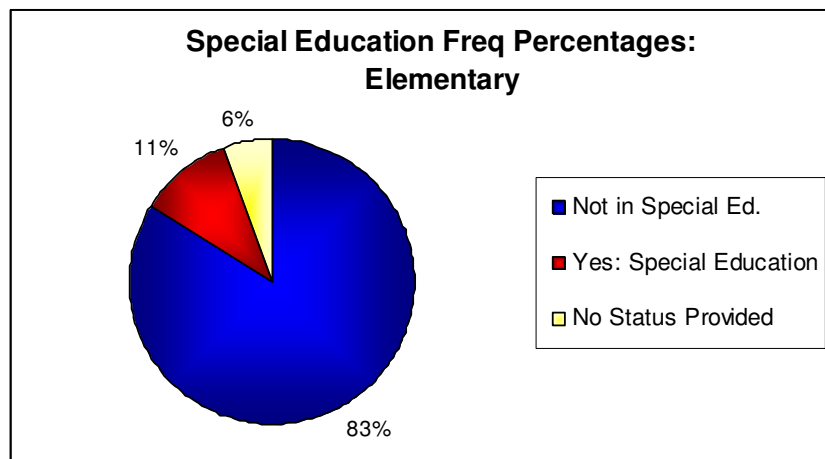
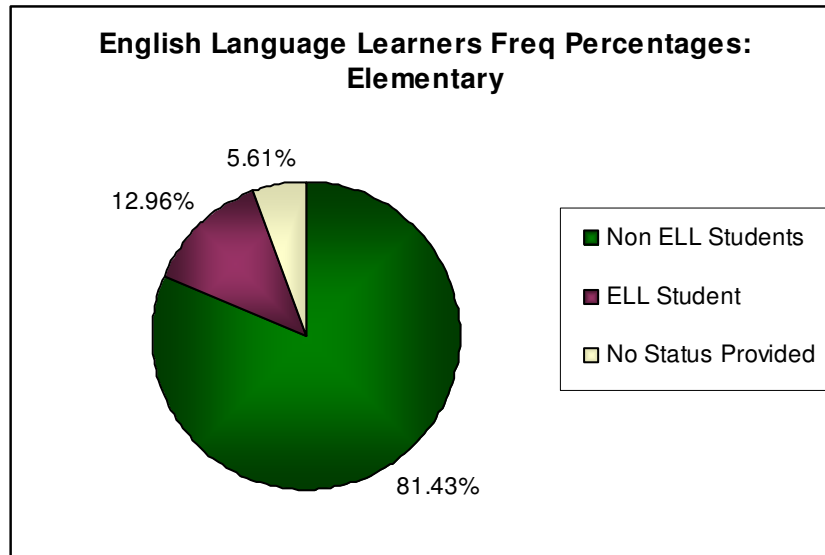


# Discovery Education Assessment RESEARCH





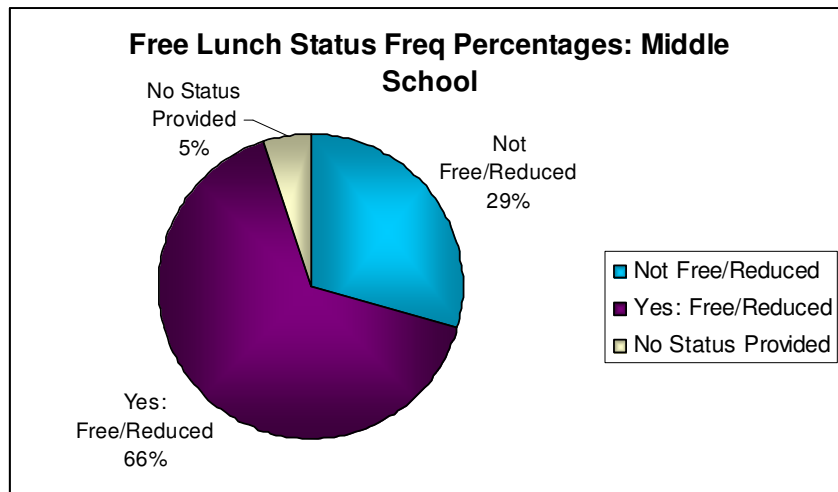
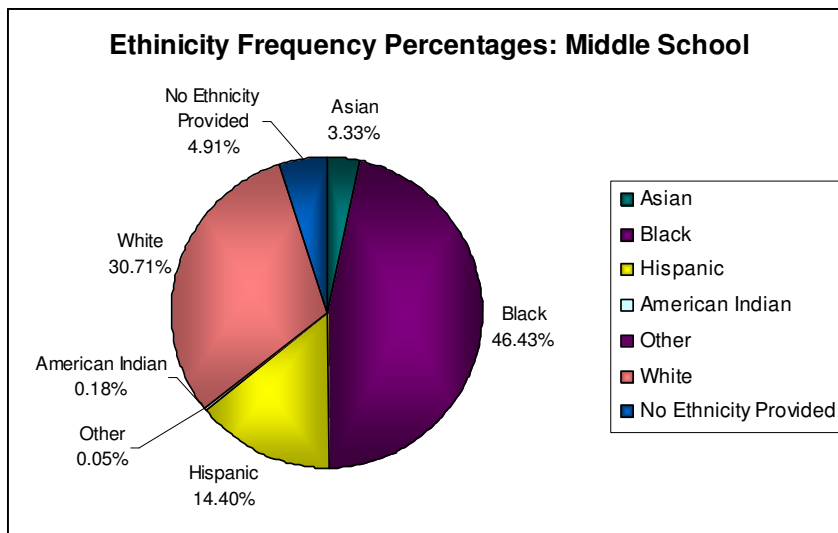
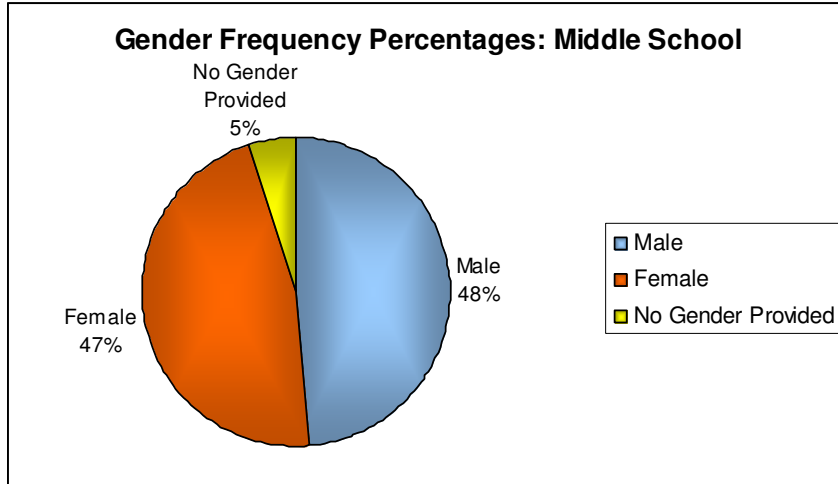
# Discovery Education Assessment RESEARCH



The following pie charts display the frequency percents of the NCLB data provided to DEA from the Metro Nashville Public School System for the Middle Schools.

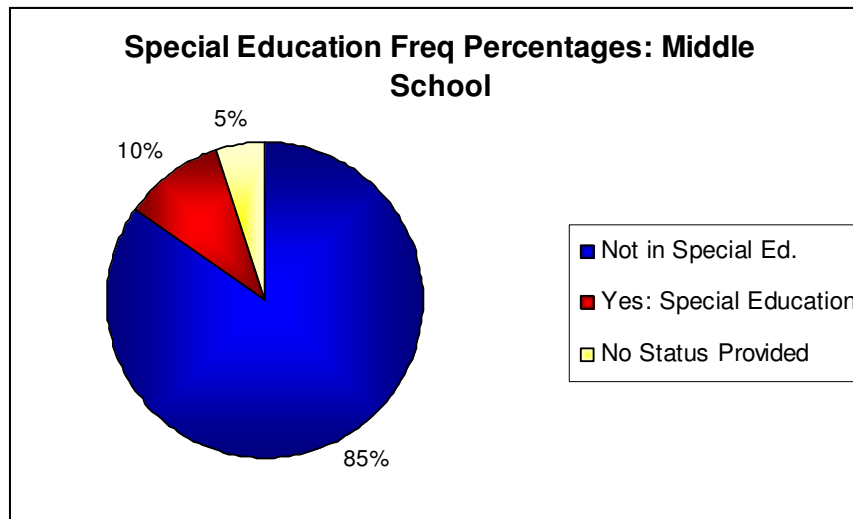
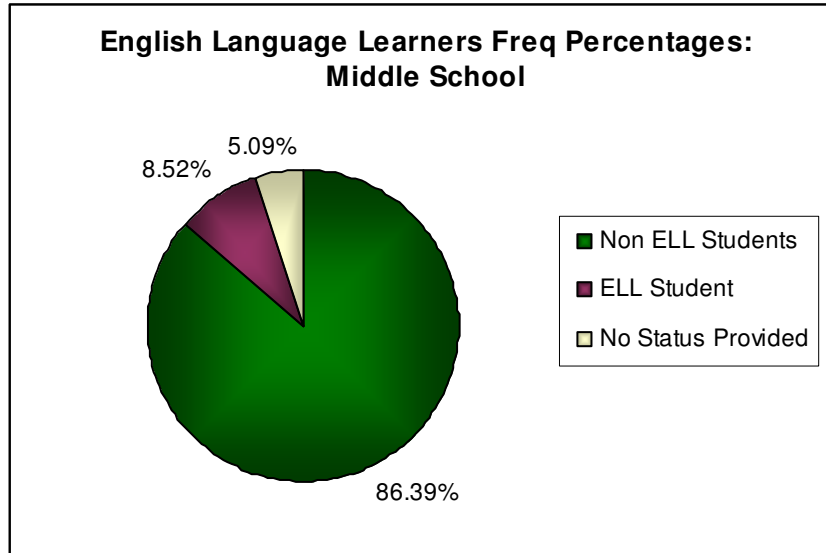


# Discovery Education Assessment RESEARCH





# 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



# Discovery Education Assessment RESEARCH

“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 ThinkLink website [www.thinklinklearning.com](http://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.

## 2. Conference Proceedings

Shrago, J.B. chair. (2006, June). *Perspectives on large-scale formative assessment*. Presented at 36<sup>th</sup> 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 36<sup>th</sup> 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 36<sup>th</sup> 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 36<sup>th</sup> 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 student*. Unpublished doctoral dissertation, Union University, Jackson, TN.





# Discovery Education Assessment RESEARCH

## 4. Research Documents

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

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

Discovery Education. (2008). *What is predictive assessment: Alabama?* Nashville, TN: Author.

Discovery Education. (2008). *What is predictive assessment: Florida?* Nashville, TN: Author.

Discovery Education. (2008). *What is predictive assessment: Illinois?* Nashville, TN: Author.

Discovery Education. (2008). *What is predictive assessment: Kentucky?* Nashville, TN: Author.

Discovery Education. (2008). *What is predictive assessment: New York?* Nashville, TN: Author.

Discovery Education. (2008). *What is predictive assessment: Tennessee?* Nashville, TN: Author.

## 5. Tests

Discovery Education Benchmark Assessments. *Tennessee 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.**



# Discovery Education Assessment RESEARCH

## Procedures for Item and Test Review

Discovery Education Assessment has established policies to review items in each benchmark assessment for appropriate item statistics and for evidence of bias. Furthermore, the collection of items that form a specific test is reviewed for alignment to state or national standards and for appropriate test difficulty. This document outlines in more detail how these processes are implemented.

### Item Statistics

P-values and item discrimination indices are calculated for each item based on the number of students that completed a benchmark assessment.

P-values represent the percentage of students tested who answered the item correctly. Based on p-values alone, the following item revision procedures are followed:

Items with p-values of .90 or above are considered too “easy” and are revised or replaced.

Items with p-values of .20 or below are considered too “hard” and are revised or replaced.

Item discrimination indices (biserial correlations) are calculated for each item. Items with low biserial correlations (less than .20) are revised or replaced.

### Test Content Validity and Statistics

The blueprints of state tests are monitored for changes in state standards. When state blueprints have changed, Discovery Education Assessment revises the blueprints for benchmark tests used in that state. Items may be revised or replaced where necessary to match any revised blueprint and to maintain appropriate test content validity.

P, A, B, and C tests within the same year are compared to verify that all tests are of comparable difficulty. When necessary, items are revised to maintain the necessary level of difficulty within each test.

The predictability of Discovery Education benchmark assessments is examined closely. When necessary, items are revised or replaced to improve the predictability of benchmark assessments. In order to maintain high levels of predictability, Discovery Education Assessment strives to revise or replace less than 20% of the items in any benchmark assessment each year.



# Discovery Education Assessment RESEARCH

## Differential Item Functioning

Differential Item Functioning (DIF) analyses are performed on items from tests where the overall sample size is large ( $n = 1000$  or more per test) and the sample size for each subgroup meets minimum standards (usually 300 or more students per subgroup). DIF analyses for Gender (males and females) and Ethnicity (Caucasian vs. African-American, or Caucasian vs. Hispanic) are routinely performed when sample size minimums are met.

DIF analyses are conducted using Rasch modeling through the computer program WINSTEPS. This program calculates item difficulty parameters for each DIF subgroup and compares these logit estimates. DIF Size is calculated using industry standard criteria (see p.1070 “An Adjustment for Sample Size in DIF Analysis”, Rasch Measurement Transactions, 20:3, Winter 2006).

The following criteria are used to determine DIF Size:

Negligible:	0 logits to .42 logits (absolute value)
Moderate:	.43 logits to .63 logits (absolute value)
Large:	.64 logits and up (absolute value)

Items with Large DIF Size are marked for the content team to review. Subject matter content experts analyze an item to determine the cause of Gender or Ethnic DIF. If the content experts can determine this cause, the item is revised to remove the gender or ethnicity bias. If the cause cannot be determined, the item is replaced with a different item that measures the same sub-skill.



# Discovery Education Assessment RESEARCH

## TEST AND QUESTION STATISTICS, RELIABILITY, AND PERCENTILES

The following section reports test and question statistics, reliability, and percentiles for the benchmark tests, for grades 3-8, Reading and Mathematics. These benchmark tests were administered in Tennessee in Spring of 2008. These two benchmark tests are representative samples of over 1000 benchmark tests developed by Discovery Education Assessment. Benchmark tests are revised each year based on test and question statistics, particularly low item discrimination indices and significant DIF.

The following statistics are reported:

<b><i>Number of Students:</i></b>	Number of students used for calculation of test statistics.
<b><i>Number of Items:</i></b>	Number of items in each benchmark test (including common items used for scaling purposes).
<b><i>Mean:</i></b>	Test mean in terms of number correct.
<b><i>Standard Deviation:</i></b>	Test standard deviation.
<b><i>Reliability:</i></b>	Cronbach's alpha.
<b><i>SEM:</i></b>	Standard Error of Measurement (SEM) for the test.
<b><i>Scale Score:</i></b>	Discovery Education Assessment Scale Score for each number correct (Scale scores are vertically scaled using Rasch measurement. Scale scores from grades K-12 range from 1000 to 2000).
<b><i>Percentiles:</i></b>	Percentage of students below each number correct score.
<b><i>Stanines:</i></b>	Scale scores that range from 1 to 9.
<b><i>Question P-values:</i></b>	The proportion correct for each item.
<b><i>Biserial:</i></b>	Item discrimination using biserial correlation.
<b><i>Rasch Item Difficulty:</i></b>	Rasch item difficulty parameter calculated using WINSTEPS.
<b><i>DIF Gender:</i></b>	Rasch item difficulty difference (Male vs. Female).
<b><i>DIF Ethnicity:</i></b>	Rasch item difficulty difference (White vs. Black).
<b><i>DIF Size</i></b>	
<b><i>Negligible:</i></b>	0 logits to .42 logits (absolute value).
<b><i>Moderate:</i></b>	.43 logits to .63 logits (absolute value).
<b><i>Large:</i></b>	.64 logits and up (absolute value).

(see p.1070 "An Adjustment for Sample Size in DIF Analysis", Rasch Measurement Transactions, 20:3, Winter 2006)



# Discovery Education Assessment RESEARCH

Technical Data Tennessee Spring Test 2009 Reading Grade 3	
Test Statistics	
Number of Students	32624
Number of Items	28
Mean	15.09
Std. Deviation	4.72
Reliability	0.84
Std. Error of Measurement	1.89

Question Statistics					
Item No.	P-Value	Biserial	Rasch Item Difficulty	DIF Gender	DIF Ethnicity
1	0.73	0.37	-0.51	-0.01	0.11
2	0.57	0.44	0.37	-0.16	-0.13
3	0.31	0.21	1.72	-0.12	0.50
4	0.71	0.46	-0.42	0.16	0.13
5	0.74	0.46	-0.60	-0.02	-0.18
6	0.83	0.43	-1.24	-0.07	-0.75
7	0.64	0.52	0.01	0.03	-0.28
8	0.64	0.46	0.00	-0.20	-0.26
9	0.72	0.47	-0.47	0.11	-0.26
10	0.80	0.48	-1.04	0.06	-0.42
11	0.72	0.44	-0.48	0.19	0.17
12	0.39	0.45	1.31	0.07	-0.24
13	0.71	0.46	-0.44	0.07	0.16
14	0.43	0.26	1.06	-0.34	0.52
15	0.60	0.53	0.21	0.06	-0.08
16	0.78	0.50	-0.85	0.20	-0.23
17	0.60	0.53	0.24	-0.06	-0.11
18	0.74	0.42	-0.60	0.12	-0.11
19	0.22	0.22	2.34	-0.28	0.53
20	0.82	0.44	-1.18	0.45	0.09
21	0.41	0.33	1.19	-0.22	0.27
22	0.62	0.49	0.10	0.18	0.04
23	0.77	0.50	-0.81	0.04	-0.18
24	0.70	0.47	-0.33	0.07	0.03
25	0.50	0.41	0.73	-0.07	0.24
26	0.70	0.47	-0.34	-0.08	-0.24
27	0.72	0.39	-0.47	-0.04	0.07
28	0.54	0.39	0.50	0.02	0.42

Scale Scores & Percentiles			
No. Correct	Scale Score	Percentile	Stanine
0	1057	1	1
1	1150	1	1
2	1206	1	1
3	1241	1	1
4	1267	1	1
5	1288	1	1
6	1306	2	2
7	1323	4	3
8	1338	6	3
9	1352	9	3
10	1365	12	4
11	1378	15	4
12	1391	19	4
13	1403	22	4
14	1415	26	4
15	1428	31	4
16	1440	36	5
17	1453	41	5
18	1467	46	5
19	1480	53	5
20	1495	60	5
21	1511	67	5
22	1528	74	6
23	1548	81	6
24	1571	88	6
25	1598	93	7
26	1635	97	8
27	1694	99	9
28	1789	99	9



# Discovery Education Assessment RESEARCH

Technical Data Tennessee Spring Test 2009 Math Grade 3	
Test Statistics	
Number of Students	32796
Number of Items	28
Mean	15.54
Std. Deviation	4.99
Reliability	0.82
Std. Error of Measurement	2.12

Question Statistics					
Item No.	P-Value	Biserial	Rasch Item Difficulty	DIF Gender	DIF Ethnicity
1	0.84	0.33	-1.39	-0.26	-0.04
2	0.37	0.29	1.28	0.02	0.23
3	0.81	0.36	-1.12	0.25	0.05
4	0.47	0.46	0.75	-0.08	0.00
5	0.58	0.39	0.2	0.04	0.08
6	0.62	0.38	-0.02	0.03	0.17
7	0.71	0.44	-0.47	0.19	-0.06
8	0.51	0.35	0.56	0.16	0.18
9	0.57	0.38	0.23	-0.24	-0.30
10	0.64	0.4	-0.13	0.15	0.06
11	0.42	0.26	1.02	-0.01	0.53
12	0.87	0.36	-1.62	0.17	0.07
13	0.46	0.52	0.81	0.20	-0.23
14	0.66	0.48	-0.22	-0.02	-0.38
15	0.52	0.47	0.51	-0.21	-0.32
16	0.58	0.47	0.21	-0.06	-0.11
17	0.73	0.45	-0.59	-0.14	-0.09
18	0.67	0.43	-0.28	-0.14	-0.09
19	0.71	0.47	-0.49	-0.03	-0.24
20	0.66	0.4	-0.21	0.00	0.07
21	0.40	0.33	1.12	0.03	0.53
22	0.68	0.43	-0.33	0.18	0.12
23	0.50	0.5	0.61	0.09	-0.53
24	0.28	0.32	1.77	-0.15	0.39
25	0.66	0.36	-0.19	-0.18	-0.06
26	0.67	0.46	-0.27	-0.19	-0.19
27	0.85	0.39	-1.46	0.23	-0.09
28	0.68	0.33	-0.3	0.08	0.35

Scale Scores & Percentiles			
No. Correct	Scale Score	Percentile	Stanine
0	1006	1	1
1	1100	1	1
2	1157	1	1
3	1192	1	1
4	1218	1	1
5	1240	1	1
6	1259	2	2
7	1275	3	2
8	1291	5	3
9	1305	8	3
10	1319	11	4
11	1332	15	4
12	1345	19	4
13	1357	24	4
14	1369	29	4
15	1382	35	5
16	1394	41	5
17	1407	47	5
18	1420	54	5
19	1434	60	5
20	1449	67	5
21	1464	73	6
22	1481	79	6
23	1500	85	6
24	1522	90	6
25	1549	94	7
26	1584	97	8
27	1642	99	9
28	1735	99	9



# Discovery Education Assessment RESEARCH

Technical Data Tennessee Spring Test 2009 Reading Grade 4	
Test Statistics	
Number of Students	32520
Number of Items	28
Mean	17.06
Std. Deviation	5.38
Reliability	0.82
Std. Error of Measurement	2.28

Question Statistics					
Item No.	P-Value	Biserial	Rasch Item Difficulty	DIF Gender	DIF Ethnicity
1	0.42	0.24	0.97	-0.19	0.34
2	0.70	0.45	-0.45	-0.12	0.00
3	0.68	0.43	-0.31	0.05	-0.11
4	0.85	0.39	-1.46	0.00	-0.31
5	0.37	0.31	1.26	-0.12	0.47
6	0.72	0.44	-0.54	-0.02	-0.44
7	0.52	0.18	0.52	0.05	0.22
8	0.72	0.53	-0.54	0.13	-0.17
9	0.39	0.37	1.12	-0.09	0.26
10	0.51	0.37	0.56	-0.14	0.25
11	0.85	0.41	-1.50	0.16	-0.18
12	0.58	0.46	0.21	0.15	-0.29
13	0.33	0.32	1.47	-0.28	-0.07
14	0.52	0.45	0.49	0.26	0.06
15	0.63	0.45	-0.05	0.20	0.02
16	0.74	0.36	-0.67	0.06	0.00
17	0.62	0.41	-0.01	-0.12	0.05
18	0.57	0.45	0.26	0.26	0.09
19	0.47	0.40	0.74	-0.20	0.10
20	0.50	0.42	0.57	-0.08	-0.04
21	0.75	0.53	-0.73	0.11	-0.58
22	0.78	0.51	-0.92	-0.07	-0.18
23	0.65	0.41	-0.16	-0.11	0.35
24	0.72	0.47	-0.57	0.15	-0.14
25	0.46	0.43	0.78	0.04	0.04
26	0.69	0.46	-0.39	-0.05	-0.12
27	0.78	0.44	-0.97	-0.13	-0.14
28	0.56	0.36	0.31	0.05	0.17

Scale Scores & Percentiles			
No. Correct	Scale Score	Percentile	Stanine
0	1123	1	1
1	1217	1	1
2	1274	1	1
3	1309	1	1
4	1335	1	1
5	1357	1	1
6	1376	2	2
7	1393	4	3
8	1408	6	3
9	1422	9	3
10	1436	12	4
11	1449	16	4
12	1462	20	4
13	1474	24	4
14	1486	29	4
15	1499	34	5
16	1511	40	5
17	1524	46	5
18	1537	52	5
19	1550	59	5
20	1564	66	5
21	1579	73	6
22	1596	80	6
23	1615	86	6
24	1636	91	7
25	1662	95	7
26	1697	98	8
27	1753	99	9
28	1847	99	9



# Discovery Education Assessment RESEARCH

Technical Data Tennessee Spring Test 2009 Math Grade 4	
Test Statistics	
Number of Students	32779
Number of Items	28
Mean	17.06
Std. Deviation	5.40
Reliability	0.82
Std. Error of Measurement	2.29

Question Statistics					
Item No.	P-Value	Biserial	Rasch Item Difficulty	DIF Gender	DIF Ethnicity
1	0.82	0.36	-1.23	0.07	-0.54
2	0.58	0.44	0.18	-0.28	0.11
3	0.71	0.39	-0.50	0.24	0.39
4	0.84	0.33	-1.43	-0.09	-0.35
5	0.57	0.45	0.25	0.01	0.18
6	0.59	0.34	0.15	-0.03	-0.05
7	0.63	0.35	-0.04	-0.19	0.33
8	0.86	0.35	-1.61	0.17	-0.26
9	0.70	0.30	-0.44	-0.30	0.22
10	0.54	0.31	0.41	0.06	0.03
11	0.85	0.37	-1.50	0.15	0.09
12	0.56	0.46	0.30	0.01	0.22
13	0.70	0.43	-0.46	0.29	0.33
14	0.74	0.48	-0.68	0.14	-0.01
15	0.37	0.46	1.25	-0.34	-0.11
16	0.61	0.44	0.06	0.32	-0.39
17	0.48	0.50	0.72	0.03	-0.19
18	0.77	0.45	-0.86	0.36	-0.28
19	0.55	0.38	0.36	0.12	0.22
20	0.70	0.42	-0.44	-0.26	-0.33
21	0.43	0.39	0.97	-0.07	0.23
22	0.66	0.45	-0.21	0.32	-0.30
23	0.42	0.50	0.99	-0.02	-0.53
24	0.26	0.36	1.91	-0.33	0.12
25	0.51	0.45	0.55	0.19	0.16
26	0.62	0.37	-0.03	-0.12	0.25
27	0.47	0.36	0.74	-0.21	0.09
28	0.51	0.43	0.55	-0.15	-0.07

Scale Scores & Percentiles			
No. Correct	Scale Score	Percentile	Stanine
0	1085	1	1
1	1178	1	1
2	1235	1	1
3	1271	1	1
4	1297	1	1
5	1319	1	1
6	1338	2	2
7	1355	3	2
8	1370	5	3
9	1385	8	3
10	1398	11	4
11	1412	15	4
12	1424	20	4
13	1437	25	4
14	1449	30	4
15	1461	36	5
16	1474	42	5
17	1487	49	5
18	1500	55	5
19	1513	61	5
20	1527	68	5
21	1543	74	6
22	1559	79	6
23	1578	84	6
24	1600	89	6
25	1626	93	7
26	1661	96	7
27	1718	98	8
28	1811	99	9





# Discovery Education Assessment RESEARCH

Technical Data	
Tennessee Spring Test 2009	
Reading Grade 5	
Test Statistics	
Number of Students	30801
Number of Items	28
Mean	16.41
Std. Deviation	5.38
Reliability	0.82
Std. Error of Measurement	2.28

Question Statistics						Scale Scores & Percentiles			
Item No.	P-Value	Biserial	Rasch Item Difficulty	DIF Gender	DIF Ethnicity	No. Correct	Scale Score	Percentile	Stanine
1	0.71	0.43	-0.61	0.00	0.06	0	1164	1	1
2	0.69	0.52	-0.52	0.25	-0.30	1	1258	1	1
3	0.55	0.36	0.20	-0.08	0.08	2	1315	1	1
4	0.66	0.38	-0.35	0.12	-0.12	3	1351	1	1
5	0.67	0.50	-0.39	-0.11	-0.35	4	1377	1	1
6	0.66	0.48	-0.35	0.17	-0.06	5	1399	1	1
7	0.78	0.28	-1.04	0.47	0.35	6	1418	2	2
8	0.50	0.45	0.45	-0.14	-0.03	7	1435	4	3
9	0.54	0.46	0.25	-0.16	0.08	8	1450	7	3
10	0.38	0.37	1.08	-0.18	-0.04	9	1465	10	3
11	0.39	0.35	1.00	-0.18	0.22	10	1478	14	4
12	0.37	0.32	1.13	-0.09	0.44	11	1491	19	4
13	0.47	0.45	0.61	-0.03	-0.60	12	1504	24	4
14	0.29	0.27	1.58	-0.30	0.31	13	1516	29	4
15	0.68	0.45	-0.44	-0.04	-0.04	14	1529	34	5
16	0.77	0.46	-0.97	0.02	-0.04	15	1541	40	5
17	0.43	0.36	0.79	0.00	0.10	16	1553	46	5
18	0.40	0.32	0.98	-0.07	0.41	17	1566	52	5
19	0.87	0.40	-1.77	0.02	-0.27	18	1579	59	5
20	0.48	0.44	0.57	0.23	-0.04	19	1592	65	5
21	0.85	0.31	-1.61	0.07	-0.07	20	1606	71	6
22	0.60	0.37	-0.01	0.02	0.29	21	1622	77	6
23	0.59	0.46	0.04	0.07	-0.02	22	1638	83	6
24	0.68	0.43	-0.43	-0.03	-0.43	23	1657	88	6
25	0.61	0.48	-0.09	-0.11	-0.41	24	1678	92	7
26	0.59	0.41	0.03	-0.06	0.03	25	1704	96	7
27	0.62	0.44	-0.15	0.21	0.23	26	1739	98	8
28	0.59	0.36	0.03	0.05	0.18	27	1796	99	9
						28	1889	99	9



# Discovery Education Assessment RESEARCH

Technical Data Tennessee Spring Test 2009 Math Grade 5	
Test Statistics	
Number of Students	31021
Number of Items	28
Mean	15.82
Std. Deviation	4.75
Reliability	0.82
Std. Error of Measurement	2.01

Question Statistics					
Item No.	P-Value	Biserial	Rasch Item Difficulty	DIF Gender	DIF Ethnicity
1	0.77	0.40	-0.82	0.04	0.20
2	0.63	0.36	-0.04	-0.24	-0.06
3	0.55	0.40	0.37	0.16	0.40
4	0.68	0.44	-0.30	0.18	-0.02
5	0.36	0.48	1.36	0.07	0.12
6	0.56	0.36	0.34	0.16	0.37
7	0.40	0.36	1.12	-0.29	-0.32
8	0.62	0.43	0.02	0.02	-0.11
9	0.70	0.39	-0.43	-0.01	-0.23
10	0.65	0.51	-0.16	-0.37	-0.53
11	0.49	0.42	0.66	0.06	-0.40
12	0.62	0.48	0.03	-0.21	-0.13
13	0.81	0.26	-1.13	0.21	0.27
14	0.49	0.34	0.67	-0.09	-0.10
15	0.86	0.39	-1.56	0.36	-0.26
16	0.54	0.44	0.43	-0.15	-0.15
17	0.55	0.52	0.38	0.10	-0.04
18	0.57	0.30	0.28	0.05	0.29
19	0.69	0.47	-0.39	0.06	-0.01
20	0.83	0.37	-1.29	0.24	-0.33
21	0.62	0.37	0.00	0.06	0.13
22	0.86	0.36	-1.56	-0.02	-0.03
23	0.63	0.39	-0.03	-0.09	0.10
24	0.20	0.32	2.35	-0.14	0.55
25	0.60	0.35	0.12	-0.21	0.02
26	0.82	0.38	-1.22	0.15	0.36
27	0.63	0.44	-0.02	0.07	-0.11
28	0.46	0.42	0.81	0.10	0.14

Scale Scores & Percentiles			
No. Correct	Scale Score	Percentile	Stanine
0	1133	1	1
1	1226	1	1
2	1283	1	1
3	1318	1	1
4	1344	1	1
5	1366	1	1
6	1385	2	2
7	1401	3	2
8	1417	5	3
9	1431	7	3
10	1445	10	3
11	1458	14	4
12	1470	18	4
13	1483	23	4
14	1495	29	4
15	1507	35	5
16	1519	41	5
17	1532	47	5
18	1545	54	5
19	1559	60	5
20	1573	67	5
21	1588	73	6
22	1605	79	6
23	1624	85	6
24	1646	90	6
25	1672	94	7
26	1708	97	8
27	1765	99	9
28	1859	99	9



# Discovery Education Assessment RESEARCH

Technical Data Tennessee Spring Test 2009 Reading Grade 6	
Test Statistics	
Number of Students	29253
Number of Items	28
Mean	14.34
Std. Deviation	4.82
Reliability	0.81
Std. Error of Measurement	2.10

Question Statistics					
Item No.	P-Value	Biserial	Rasch Item Difficulty	DIF Gender	DIF Ethnicity
1	0.72	0.43	-0.82	-0.38	-0.15
2	0.72	0.34	-0.83	-0.03	0.22
3	0.77	0.42	-1.15	0.07	-0.38
4	0.51	0.43	0.25	-0.13	0.00
5	0.78	0.43	-1.19	0.04	-0.30
6	0.80	0.44	-1.36	-0.23	-0.51
7	0.84	0.45	-1.67	0.05	-0.27
8	0.42	0.48	0.70	-0.11	0.07
9	0.44	0.39	0.57	-0.02	-0.17
10	0.18	0.21	2.14	-0.24	0.62
11	0.79	0.35	-1.27	0.31	0.14
12	0.68	0.48	-0.61	0.27	-0.24
13	0.55	0.43	0.06	0.00	-0.10
14	0.80	0.47	-1.32	0.16	-0.30
15	0.52	0.41	0.22	-0.03	-0.08
16	0.60	0.52	-0.18	0.15	-0.32
17	0.42	0.45	0.69	-0.48	-0.21
18	0.66	0.42	-0.49	-0.07	0.01
19	0.41	0.37	0.73	-0.03	0.00
20	0.48	0.35	0.40	-0.04	0.22
21	0.40	0.40	0.81	-0.12	0.17
22	0.29	0.24	1.37	-0.09	0.49
23	0.75	0.49	-1.03	0.11	-0.38
24	0.44	0.39	0.58	0.19	0.27
25	0.30	0.27	1.32	-0.04	0.51
26	0.48	0.38	0.41	0.34	0.46
27	0.45	0.37	0.55	0.29	0.06
28	0.34	0.35	1.12	0.08	0.18

Scale Scores & Percentiles			
No. Correct	Scale Score	Percentile	Stanine
0	1193	1	1
1	1286	1	1
2	1343	1	1
3	1379	1	1
4	1406	1	1
5	1428	2	2
6	1447	3	2
7	1464	6	3
8	1480	8	3
9	1495	12	4
10	1509	17	4
11	1523	21	4
12	1536	27	4
13	1550	33	5
14	1562	39	5
15	1575	45	5
16	1588	52	5
17	1601	59	5
18	1615	66	5
19	1629	72	6
20	1644	78	6
21	1659	84	6
22	1677	88	6
23	1696	92	7
24	1718	95	7
25	1744	98	8
26	1780	99	9
27	1837	99	9
28	1930	99	9



# Discovery Education Assessment RESEARCH

Technical Data Tennessee Spring Test 2009 Math Grade 6	
Test Statistics	
Number of Students	29758
Number of Items	24
Mean	13.22
Std. Deviation	4.80
Reliability	0.8
Std. Error of Measurement	2.15

Question Statistics					
Item No.	P-Value	Biserial	Rasch Item Difficulty	DIF Gender	DIF Ethnicity
1	0.84	0.32	-1.72	0.17	-0.10
2	0.39	0.46	0.80	0.04	0.14
3	0.65	0.41	-0.49	-0.16	0.02
4	0.43	0.33	0.63	-0.10	0.17
5	0.57	0.49	-0.10	-0.25	-0.27
6	0.74	0.32	-1.01	-0.14	0.13
7	0.47	0.27	0.41	-0.04	0.37
8	0.64	0.47	-0.42	0.12	-0.02
9	0.61	0.32	-0.30	0.02	0.14
10	0.45	0.44	0.52	-0.04	-0.11
11	0.46	0.42	0.48	0.25	0.11
12	0.46	0.49	0.48	-0.36	-0.18
13	0.62	0.41	-0.32	0.07	-0.09
14	0.36	0.41	0.99	-0.20	-0.19
15	0.66	0.47	-0.53	0.12	-0.12
16	0.51	0.40	0.21	0.22	0.12
17	0.63	0.50	-0.36	0.07	-0.32
18	0.55	0.39	0.00	-0.08	-0.12
19	0.23	0.35	1.78	-0.04	0.34
20	0.74	0.38	-0.97	0.18	-0.09
21	0.54	0.52	0.05	0.02	-0.04
22	0.39	0.38	0.80	0.38	0.32
23	0.61	0.45	-0.30	-0.04	0.07
24	0.68	0.42	-0.63	-0.11	-0.26

Scale Scores & Percentiles			
No. Correct	Scale Score	Percentile	Stanine
0	1226	1	1
1	1320	1	1
2	1378	1	1
3	1414	1	1
4	1441	2	2
5	1463	4	3
6	1482	6	3
7	1500	10	3
8	1516	16	4
9	1531	22	4
10	1545	28	4
11	1560	35	5
12	1573	42	5
13	1587	49	5
14	1602	56	5
15	1616	63	5
16	1631	70	6
17	1647	76	6
18	1665	81	6
19	1684	86	6
20	1707	91	7
21	1734	94	7
22	1770	97	8
23	1827	99	9
24	1921	99	9



# Discovery Education Assessment RESEARCH

Technical Data Tennessee Spring Test 2009 Reading Grade 7	
Test Statistics	
Number of Students	29182
Number of Items	28
Mean	14.03
Std. Deviation	4.63
Reliability	0.76
Std. Error of Measurement	2.27

Question Statistics						Scale Scores & Percentiles			
Item No.	P-Value	Biserial	Rasch Item Difficulty	DIF Gender	DIF Ethnicity	No. Correct	Scale Score	Percentile	Stanine
1	0.78	0.29	-1.34	0.06	-0.10	0	1218	1	1
2	0.56	0.33	-0.13	-0.40	-0.34	1	1311	1	1
3	0.58	0.45	-0.21	0.10	-0.16	2	1368	1	1
4	0.32	0.23	1.00	-0.22	0.18	3	1403	1	1
5	0.64	0.39	-0.51	0.02	0.08	4	1429	1	1
6	0.47	0.32	0.28	-0.12	0.12	5	1451	2	2
7	0.41	0.41	0.57	0.07	0.13	6	1470	3	2
8	0.45	0.29	0.36	-0.15	0.25	7	1487	5	3
9	0.68	0.52	-0.73	0.29	-0.33	8	1502	9	3
10	0.58	0.41	-0.21	0.05	-0.16	9	1516	13	4
11	0.79	0.40	-1.35	0.18	-0.17	10	1530	18	4
12	0.74	0.43	-1.04	0.05	-0.01	11	1543	24	4
13	0.43	0.35	0.49	-0.13	0.05	12	1555	30	4
14	0.61	0.36	-0.38	0.04	-0.06	13	1568	37	5
15	0.84	0.41	-1.73	0.01	-0.20	14	1580	44	5
16	0.39	0.31	0.66	-0.13	0.03	15	1592	52	5
17	0.21	0.26	1.70	-0.51	-0.16	16	1605	59	5
18	0.54	0.35	-0.05	0.09	0.27	17	1617	66	5
19	0.39	0.37	0.65	0.18	-0.02	18	1630	73	6
20	0.41	0.34	0.59	-0.38	-0.06	19	1644	79	6
21	0.28	0.27	1.25	-0.51	-0.05	20	1658	85	6
22	0.65	0.29	-0.60	0.37	0.37	21	1673	89	6
23	0.55	0.52	-0.10	0.28	-0.13	22	1690	93	7
24	0.63	0.37	-0.49	0.20	0.14	23	1708	96	7
25	0.64	0.31	-0.51	0.29	0.16	24	1730	98	8
26	0.42	0.42	0.51	0.20	-0.12	25	1756	99	9
27	0.30	0.38	1.12	-0.06	-0.14	26	1791	99	9
28	0.49	0.32	0.18	0.02	0.12	27	1848	99	9
						28	1941	99	9



# Discovery Education Assessment RESEARCH

Technical Data Tennessee Spring Test 2009 Math Grade 7	
Test Statistics	
Number of Students	29258
Number of Items	28
Mean	15.30
Std. Deviation	5.48
Reliability	0.82
Std. Error of Measurement	2.32

Question Statistics					
Item No.	P-Value	Biserial	Rasch Item Difficulty	DIF Gender	DIF Ethnicity
1	0.67	0.45	-0.61	-0.05	-0.11
2	0.62	0.46	-0.37	-0.36	0.10
3	0.84	0.36	-1.77	0.09	-0.07
4	0.56	0.32	-0.02	0.21	0.14
5	0.65	0.41	-0.50	-0.07	0.21
6	0.34	0.28	1.05	0.16	0.17
7	0.60	0.45	-0.23	0.10	-0.03
8	0.68	0.45	-0.66	0.07	0.00
9	0.63	0.46	-0.42	-0.09	-0.16
10	0.60	0.27	-0.26	0.18	0.55
11	0.54	0.39	0.07	0.01	0.08
12	0.66	0.56	-0.57	0.01	-0.63
13	0.55	0.44	-0.01	-0.33	-0.35
14	0.26	0.30	1.56	-0.13	0.26
15	0.29	0.46	1.37	-0.07	-0.06
16	0.59	0.48	-0.18	-0.03	-0.60
17	0.31	0.41	1.26	-0.05	0.12
18	0.60	0.53	-0.26	0.32	0.02
19	0.34	0.40	1.04	-0.35	-0.30
20	0.36	0.26	0.97	-0.79	0.24
21	0.80	0.46	-1.41	0.54	0.06
22	0.45	0.38	0.50	0.03	0.32
23	0.69	0.51	-0.75	0.27	-0.17
24	0.40	0.37	0.77	-0.25	-0.02
25	0.46	0.36	0.44	0.20	0.27
26	0.58	0.40	-0.16	0.08	0.16
27	0.81	0.40	-1.47	0.12	-0.20
28	0.42	0.46	0.63	0.21	0.12

Scale Scores & Percentiles			
No. Correct	Scale Score	Percentile	Stanine
0	1221	1	1
1	1315	1	1
2	1372	1	1
3	1407	1	1
4	1434	1	1
5	1456	2	2
6	1475	4	3
7	1492	7	3
8	1508	10	3
9	1522	14	4
10	1536	19	4
11	1549	25	4
12	1562	30	4
13	1575	36	5
14	1587	42	5
15	1600	48	5
16	1612	54	5
17	1625	60	5
18	1638	66	5
19	1652	72	6
20	1666	78	6
21	1681	83	6
22	1698	87	6
23	1717	91	7
24	1738	94	7
25	1764	97	8
26	1799	98	8
27	1856	99	9
28	1949	99	9



# Discovery Education Assessment RESEARCH

Technical Data Tennessee Spring Test 2009 Reading Grade 8	
Test Statistics	
Number of Students	29043
Number of Items	28
Mean	14.96
Std. Deviation	5.21
Reliability	0.84
Std. Error of Measurement	2.08

Question Statistics					
Item No.	P-Value	Biserial	Rasch Item Difficulty	DIF Gender	DIF Ethnicity
1	0.39	0.29	1.10	-0.10	0.23
2	0.67	0.38	-0.35	0.00	-0.26
3	0.57	0.46	0.19	-0.07	-0.12
4	0.76	0.43	-0.87	0.09	-0.18
5	0.57	0.31	0.16	0.17	0.43
6	0.77	0.43	-0.94	0.06	-0.13
7	0.60	0.46	-0.01	0.09	-0.26
8	0.36	0.34	1.23	-0.14	0.33
9	0.63	0.45	-0.14	-0.10	0.03
10	0.60	0.42	0.03	-0.09	-0.18
11	0.64	0.42	-0.22	0.18	0.12
12	0.68	0.42	-0.43	0.02	-0.03
13	0.58	0.49	0.11	-0.11	-0.15
14	0.46	0.50	0.70	-0.11	-0.18
15	0.72	0.52	-0.66	0.10	-0.18
16	0.78	0.54	-1.02	0.22	-0.24
17	0.51	0.33	0.48	-0.02	0.27
18	0.63	0.49	-0.12	0.01	0.08
19	0.63	0.35	-0.14	-0.09	0.37
20	0.55	0.52	0.29	-0.11	-0.46
21	0.63	0.41	-0.15	-0.39	-0.04
22	0.54	0.37	0.32	-0.20	-0.04
23	0.66	0.44	-0.29	0.06	0.03
24	0.37	0.28	1.21	0.00	0.61
25	0.36	0.36	1.23	-0.05	0.17
26	0.62	0.51	-0.07	0.22	-0.28
27	0.71	0.42	-0.61	0.28	0.05
28	0.78	0.48	-1.01	0.25	-0.12

Scale Scores & Percentiles			
No. Correct	Scale Score	Percentile	Stanine
0	1225	1	1
1	1318	1	1
2	1374	1	1
3	1408	1	1
4	1434	1	1
5	1454	2	2
6	1472	3	2
7	1488	5	3
8	1503	8	3
9	1516	11	4
10	1529	15	4
11	1541	19	4
12	1553	23	4
13	1565	27	4
14	1577	32	5
15	1588	37	5
16	1600	43	5
17	1612	48	5
18	1624	54	5
19	1637	60	5
20	1651	67	5
21	1666	73	6
22	1682	79	6
23	1700	85	6
24	1722	90	6
25	1747	94	7
26	1782	97	8
27	1838	99	9
28	1932	99	9



# Discovery Education Assessment RESEARCH

Technical Data Tennessee Spring Test 2009 Math Grade 8	
Test Statistics	
Number of Students	29324
Number of Items	28
Mean	16.24
Std. Deviation	5.93
Reliability	0.85
Std. Error of Measurement	2.30

Question Statistics					
Item No.	P-Value	Biserial	Rasch Item Difficulty	DIF Gender	DIF Ethnicity
1	0.68	0.39	-0.54	-0.06	-0.33
2	0.50	0.31	0.45	-0.07	0.19
3	0.62	0.48	-0.17	-0.06	-0.38
4	0.35	0.43	1.27	0.06	0.31
5	0.41	0.38	0.90	0.21	0.60
6	0.46	0.36	0.63	0.05	0.44
7	0.85	0.35	-1.69	0.41	0.06
8	0.83	0.43	-1.55	-0.04	-0.15
9	0.62	0.41	-0.17	0.04	-0.29
10	0.81	0.38	-1.33	0.12	0.37
11	0.70	0.48	-0.65	0.27	0.23
12	0.44	0.36	0.74	-0.19	0.08
13	0.59	0.56	-0.04	-0.18	-0.42
14	0.61	0.43	-0.11	0.00	0.04
15	0.69	0.46	-0.57	0.10	0.01
16	0.64	0.47	-0.30	0.03	-0.07
17	0.53	0.45	0.26	-0.11	-0.09
18	0.48	0.34	0.57	-0.07	0.40
19	0.48	0.37	0.52	-0.21	0.00
20	0.69	0.54	-0.58	0.14	-0.17
21	0.66	0.48	-0.39	-0.10	-0.17
22	0.40	0.39	0.97	-0.08	0.23
23	0.74	0.45	-0.87	0.26	0.27
24	0.54	0.59	0.25	0.30	-0.27
25	0.24	0.40	1.96	-0.15	0.02
26	0.62	0.50	-0.19	-0.18	-0.06
27	0.47	0.55	0.62	-0.09	-0.32
28	0.58	0.52	0.01	-0.26	-0.31

Scale Scores & Percentiles			
No. Correct	Scale Score	Percentile	Stanine
0	1237	1	1
1	1331	1	1
2	1388	1	1
3	1424	1	1
4	1450	1	1
5	1472	2	2
6	1491	4	3
7	1508	6	3
8	1523	9	3
9	1538	13	4
10	1551	18	4
11	1565	22	4
12	1577	27	4
13	1590	32	5
14	1602	37	5
15	1615	43	5
16	1627	48	5
17	1640	53	5
18	1653	59	5
19	1667	64	5
20	1681	70	6
21	1696	75	6
22	1713	80	6
23	1732	85	6
24	1754	89	6
25	1781	93	7
26	1816	96	7
27	1873	98	8
28	1967	99	9