Maryland School Assessment
2006

Technical Report 2006

for

Maryland School Assessment

Submitted to Maryland State Department of Education

September 06

Contents

Tables

Figures

OVERVIEW

Historical Overview

The Maryland School Assessment (MSA) program replaces the Maryland Student Performance Assessment Program (MSPAP), which had been administered from 1992 to 2002. In 2003, the MSA Reading and Mathematics Assessments were introduced in Grades 3, 5, 8, and 10. In 2004, Grades 4, 6, and 7 were added to the program. CTB/McGraw-Hill was responsible for the Mathematics assessments in Grades 3 through 8 and the Reading assessment in Grade 10. In 2006, CTB/McGraw-Hill was only responsible for the Mathematics assessments in Grades 3 through 8. This technical report addresses only those assessments for which CTB/McGraw-Hill was responsible.

The MSA Mathematics assessments include CTB/McGraw-Hill's *TerraNova* survey (TN) as well as custom selected-response (SR), student- produced-response (SPR), and constructed-response (CR) items written to measure performance on the Maryland content standards. *TerraNova* survey Form C was administered at Grades 3, 4, 5, 7, and 8; *TerraNova* survey Form D was administered at Grade 6.

In 2003 and 2004, two types of scores were reported for the Reading and Mathematics assessments: Norm Referenced Test (NRT) scores and Criterion Referenced Test (CRT) scores. The NRT scores were computed using *TerraNova* items only. The CRT scores were calculated using the custom items written to the Maryland content standards plus a subset of *TerraNova* items that align with the state content standards. In 2005, both NRT and CRT scores were reported for Mathematics, but only CRT scores were reported for Reading. In 2006, both NRT and CRT scores were reported for Mathematics.

A Bookmark standard setting was conducted in 2003 to set proficiency level cut scores for the Mathematics tests in Grades 3, 5 and 8 and the Reading tests in Grade 10. Because 2004 was the first testing year for Grades 4, 6, and 7, a second Bookmark standard setting was held in summer 2004 to set cut scores for these additional grades. The performance level cut scores obtained from the standard setting are used to assign students to three proficiency levels (basic, proficient, and advanced) for AYP reporting under the "No Child Left Behind" act. Information about the Bookmark procedures and results can be found in separate standard-setting technical reports, submitted to the Maryland Department of Education in August 2003 and August 2004.

Development of Items and Tests to Meet the MSA Statewide Academic Learning Standards

The MSA Mathematics assessments are designed and constructed to meet the Maryland Statewide Academic Learning Standards. (For purposes of item development and review, these standards are referred to as the "Content Standards and Assessment Limits.")

The item development process used for MSA is an iterative process, involving multiple rounds of item review and revision. The processes used for developing items for the 2005 test administration are described below. Item writing began in early February, 2003, and the item content review meeting was held July 14-16, 2004.

1. MSDE and CTB staff attended item writer training sessions in Tacoma, Washington. MSDE staff trained the item writers on the Maryland content standards and assessment limits. CTB staff provided training on the item specifications documents.

2. Items were edited by CTB staff. MSDE staff came to Monterey and reviewed the items with CTB staff during a nine day "side-by-side" review in April 2004 to prepare for item content review.

3. Separate committees comprised of Maryland educators were convened for content and sensitivity. The content review committee members recommended edits, and then the sensitivity committee reviewed items. MSDE and CTB staff reviewed and reconciled all recommended edits during "side-by-side" reviews for three days. Form selection also occurred at this time.

4. Following the item content review meeting, test book manuscripts were prepared and the items were reviewed for style at the time manuscripts were processed. During the page production cycles, items underwent further content and style refinements.

Test Design and Specifications

Table 1 shows the test designs for Mathematics Grades 3 through 8. The test designs presented in this table represent the targeted test design for each grade, and show the targeted distribution of score points by content standard. The final operational forms may deviate slightly from these targets.

For Mathematics, some standards are combined for reporting purposes. Table 2 presents the actual distribution of score points by reporting category for Mathematics.

Table 1
Test Designs by Grade / Content

Grade 3 Mathematics

	Content Standard	TerraNova Items that Contribute to CRT Score	Number of CRT SR Items	Number of CRT BCR Items	Points	Percent
1	Algebra, Patterns, and Functions	12	11	1	13	18%
2	Geometry	16, 17	5	1	8	11%
3	Measurement	10, 14	4	1	7	10%
4	Statistics	24	10	1	12	17%
5	Probability		2		2	3%
6	Number Relationships and Computation	1, 2, 4, 13, 18	8	3	16	22%
7	Process of Mathematics			7	14	19%
	Total Score Points	11	40	21	72	100%

Grade 4 Mathematics

	Content Standard	TerraNova Items that Contribute to CRT Score	Number of CRT SR Items	Number of CRT BCR Items	Number of CRT ECR Items	Points	Percent
1	Algebra, Patterns, and Functions		13	1		14	20%
2	Geometry	20	5	1		7	10%
3	Measurement	31	5	1		7	10%
4	Statistics		7	1		8	11%
5	Probability		6	1		7	10%
6	Number Relationships and Computation	1,2,3,4,10, 17,18,27	4	2		14	20%
7	Process of Mathematics			7		14	20%
	Total Score Points	10	40	21		71	100%

Grade 5 Mathematics

	Content Standard	TerraNova Items that Contribute to CRT Score	Number of CRT SR Items	Number of CRT BCR Items	Number of CRT ECR Items	Points	Percent
1	Algebra, Patterns, and Functions	27, 28	11	1	1	15	20%
2	Geometry	13	4	1		6	8%
3	Measurement	17, 23, 26	4	1		8	11%
4	Statistics	12	7	1		9	12%
5	Probability	32	2	1		4	5%
6	Number Relationships and Computation	2, 4, 6, 8, 31	8	2		15	20%
7	Process of Mathematics			7	1	17	23%
	Total Score Points	13	36	21	4	74	100%

Table 1 (cont.)
Test Designs by Grade / Content

Grade 6 Mathematics

	Content Standard	TerraNova Items that Contribute to CRT Score	Number of CRT SR Items	Number of CRT BCR Items	Number of CRT ECR Items	Points	Percent
1	Algebra, Patterns, and Functions	13	11	1	1	14	20%
2	Geometry	17	6	1		8	11%
3	Measurement		5	1		6	9%
4	Statistics		8	1		9	13%
5	Probability		4			4	6%
6	Number Relationships and Computation	6, 18, 20	9	2		14	20%
7	Process of Mathematics			6	1	15	21%
	Total Score Points	5	43	18	4	70	100%

Grade 7 Mathematics

	Content Standard	TerraNova Items that Contribute to CRT Score	Number of CRT SR Items	Number of CRT SPR Items	Number of CRT BCR Items	Number of CRT ECR Items	Points	Percent
1	Algebra, Patterns, and Functions		9	3	1	1	14	20%
2	Geometry		4	2		1	7	10%
3	Measurement	24	3	1	1		6	8%
4	Statistics		5	1	1	1	8	11%
5	Probability		3	2	1		6	8%
6	Number Relationships and Computation	3, 9, 13, 15, 32	6	3			14	19%
7	Process of Mathematics				4	3	17	24%
	Total Score Points	6	30	12	12	12	72	100%

Grade 8 Mathematics

	Content Standard	TerraNova Items that Contribute to CRT Score	Number of CRT SR Items	Number of CRT SPR Items	Number of CRT BCR Items	Number of CRT ECR Items	Points	Percent
1	Algebra, Patterns, and Functions	21, 29	6	4	2	1	15	20%
2	Geometry	27	4	2	1		8	11%
3	Measurement	16	2	1		1	5	7%
4	Statistics	13	5	1	1	1	9	12%
5	Probability		2	2	1		5	7%
6	Number Relationships and Computation	2, 3, 6, 7, 20, 26	6	2			14	19%
7	Process of Mathematics				5	3	19	25%
	Total Score Points	11	25	12	15	12	75	100%

Table 2
Summary of Score Points

Content Standard Reporting Category	Grade 3		Grade 4		Grade 5	
	Score Points	Percentage	Score Points	Percentage	Score Points	Percentage
1	13	18.1%	14	19.7%	15	20.3%
2&3	15	20.8%	14	19.7%	14	18.9%
4&5	14	19.4%	15	21.1%	13	17.5%
6	16	22.2%	14	19.7%	15	20.3%
7	14	19.4%	14	19.7%	17	23.0%
Total	72	100%	71	100%	74	100%

Content Standard Reporting Category	Grade 6		Grade 7		Grade 8	
	Score Points	Percentage	Score Points	Percentage	Score Points	Percentage
1	14	20.0%	14	19.4%	15	20.0%
2&3	14	20.0%	13	18.1%	13	17.3%
4&5	13	18.6%	14	19.4%	14	18.7%
6	14	20.0%	14	19.4%	14	18.7%
7	15	21.4%	17	23.6%	19	25.3%
Total	70	100%	72	100%	75	100%

Processing and Scoring of Test Materials

CTB's primary goal in the scoring and processing of test documents is to deliver quality results to MSA according to established timelines. The accuracy and timeliness of reports are the primary concerns of the team devoted to providing scoring services.

CTB's MSA scoring team is based in Monterey, California and Delran, New Jersey. This team of trained technical specialists has been responsible for coordinating all scoring and reporting activities related to the processing of MSA test documents. Document preparation, interdepartmental coordination and communication, processing specifications, and problem resolution are functions to be performed by a designated Scoring Project Manager from this team. The scoring team works closely with all CTB departments to ensure successful scoring and reporting of MSA.

Scoring Process Overview

CTB's scoring process includes many quality assurance steps that are integrated into each step. Presented below, in order of occurrence, are quality assurance procedures applicable to the Scoring and Reporting process.

Prework

Prior to document arrival at CTB, the scoring team utilizes available customer data to prepare materials to expedite the document-handling process. Team members verify the accuracy of the following materials:

- Expected number of students by grade and school
- Test date
- Precoded headers generated from school/district enrollment files
- Return Shipping Labels
- Report services specifications
- Sample reports
- Report collation examples
- Report packing schematics
- Document type (i.e., selected response/constructed-response)
- Packing lists generated for report shipments
- Other requirements to meet MSA specifications

Prior to receipt of answer documents, detailed scoring specifications for MSA are distributed to the various workstations involved in the scoring and editing process.

Receiving

Shipments are tracked electronically, from the time of pickup at the sites, until delivery at CTB. After receipt, documents are organized by LAC. For each LAC the following steps were performed:

1. The box count is verified against the carrier's bill of lading and/or box count indicators as printed on the outside of the box. If a discrepancy is encountered, boxes are placed in a problem resolution area and discrepancy procedures are enforced. If missing boxes are not located within 24 hours, the Scoring Team is notified and they contact the LAC for resolution.

2. The shipment is checked for damaged materials. If the integrity of the documents is affected by any kind of damage, the Scoring Team is notified. Depending on the severity of the problem, the team member contacts the LAC for resolution. A record of all damaged materials is maintained.

3. Before documents leave the Receiving area they are logged into the computerized tracking system which provides real-time information regarding the status of the documents throughout the scoring and editing process. The electronic profile for each LAC is updated with at least the following information:

   1. LEA name

   2. Date of receipt

   3. Box count

   4. Shipping carrier

CTB follows-up with each LAC whose test materials are not received by the date agreed upon by CTB and MSDE.

Login

Documents released by Receiving are transferred to Log-In, where the following activities are performed:

1. The headers (Group Information Sheets) are checked against School Group Lists (SGLs) to verify the number of students tested within each group (class).

2. The documents are grouped in manageable stacks and document alignment is checked to ensure proper scanning.

3. A scannable header is placed on top of each stack and a number is assigned to identify each unique stack of documents within a group.

Scanning

After login verifies all of the information has been received and has prepared the documents for scanning, the documents are moved to the scanning area. Here they are cut into single sheets and electronically scanned. Scanners are calibrated periodically.

The scanners used by CTB have built-in checks for miscalibration. Hardware bias checking is used in real-time to verify that the scanner calibration is maintained during the scanning process. Additional checks are implemented by CTB to reinforce the built-in hardware checks and to ensure optimal scanner setup.

CTB's scanning software utilizes the speed of the NCS 5000I optical scanners to capture document images and bubbled data without requiring specific document editing and resolution rules. Scanners are thus able to run at rated speed with no interruptions except for problems with the physical documents. All editing of the scanned documents is performed, in a subsequent step, in the raw scoring/editing system.

The scanning program evaluates every detectable mark on both sides of each page, and records the intensity and coordinates of solid marks for resolution in the subsequent raw scoring step. The form identification (i.e., "skunk marks") determines the type of document, and the headers determine customer identification and district, school, and class.

Editing/Updates

Raw scoring and editing of scanned data is performed in a client/server system (WinScore), where a sophisticated system of edits are invoked to review the integrity of each batch scanned and to produce a list of error suspects. While the editors can view data from any document on-line, the error suspect list concentrates on the most likely problems based on pre-defined guidelines. This system reduces editing time and provides a high degree of quality control.

CTB continues to enhance the capability of editing software to simplify the detection and correction of errors. On-line editing screens focus an editor on potential problems and then provide related information. The actual scanned documents are always available to the editor, and the software supports the review and correction of any field in the scanned record. Entry and verification of the necessary corrections are enhanced to ensure each error is actually corrected.

As batches are extracted for scoring, a final edit is performed to ensure all requirements for scoring are met. This automated final edit flags a batch for further editing if any error is still detected. A batch containing errors cannot be extracted for reporting. This ensures a high level of accuracy of the scored data.

CTB has maintained a professional staff of specialized data processing technicians to lead the verification process to ensure the integrity of the student response data at both group and individual levels. This process includes the following error checks**:**

1. **Reliability.** This check ensures that the raw scores for each subtest are above chance levels. Scores not passing this edit are checked by a trained specialist to ensure that responses are being read correctly and that the correct form and level of the test is being used.

2. **Biographical data.** Electronic edits are performed on such elements as student name to ensure leading or embedded blanks are corrected when possible.

3. **Student counts.** Actual counts based on scanned records are electronically compared with expected counts, and discrepancies are flagged.

4. **School name/number.** Pre-assigned school numbers and names are verified against an electronic file.

5. **Custom edits.** Special edits can be performed using custom software that works in conjunction with our standard scoring process.

Document retention

When the editing process is completed, documents are moved to a staging area to be prepared for retention. Bundles are caged, warehoused in a recoverable location, and retained for possible retrieval during the specified retention period. Once this period is over, documents are destroyed according to procedures that ensure security is maintained.

Scoring/Reporting Software

The primary set of products utilizing CTB's mainframe scoring software (EISS) is *TerraNova* Survey and MSA.

- **Shelf software** supports each test available in the CTB annual catalog. When a customer's scoring request is entered on a scoring order screen, the software activates the scoring and reporting requested by the customer. Parameters from the scoring order screen control which scoring and reporting programs are executed, as well as the content and sequence of the printed output.

- **Custom software** is necessary to support contracts with unique requirements. CTB has developed many modules to meet customized scoring and reporting requirements. In addition, our large programming staff can develop new software to meet the needs of a new customization. CTB has the resources to develop custom software for very large and complex contracts.

EISS receives data from WinScore. The data is scored, summarized, sorted/selected, and reported according to the contract requirements. This system is optimized for efficient high volume processing, and providing for maximum flexibility to fulfill the contract's specific needs.

Advanced Function Printing (AFP)

The IBM Advanced Function Printing (AFP) system is a key factor in CTB's ability to print large volumes of reports with varied content and sequences. CTB provides the functionality to print reports in the actual shipping sequence, with no manual sorting or collation required. In addition, each page may contain complex graphics and the visual aids necessary to clearly convey the information to the wide variety of people who read the reports. CTB converted all mainframe systems to AFP and developing all new reports in this environment.

AFP operates on high-speed laser printers using large roll feeders for several hours of uninterrupted printing at a rate of over 200 pages per minute. The printers' output processors then separate packages, or sets, of reports.

AFP supports report collation. Reports can be printed in any desired sequence, since the contents of each set of reports can be predefined. The sequence in which these packages are printed is also predefined. A "break page" of control and routing information precedes each package of reports. For example, for a district-wide school package, the break page may contain test, type of report, report level/grade, school name, principal's name and school address information. Packages are produced in the final order for quality checks and packaging for shipment.

With AFP graphic capabilities, CTB can design more meaningful reports. Form and content can be varied at any time while printing, fonts can be mixed on a page, graphics can be added, and complex graphics can be inserted to represent variable data.

CTB adopts procedures to provide unprecedented flexibility in the reporting software. In many cases, an application program need not be changed to modify or enhance a report; the much simpler AFP page definition can be changed, leaving the application program intact. Thus, programming, testing, and quality assurance are all simplified.

Scoring Quality Assurance

The Technology and Scoring Departments at CTB both have quality assurance sections specifically charged with reviewing scoring data and reports during all stages of the process. The Technology quality assurance team verifies the accuracy of all reporting programs before they become operational. The Scoring quality assurance team verifies the accuracy of report information during the scoring process. After all data is entered into the scoring system and all reporting programs are completed, a sample of reports are printed and submitted to the Scoring quality assurance group. They review the sample reports extensively to verify the accuracy and correct presentation of all data.

Red Team Review

During the scoring process, numerous quality assurance checks are in place to ensure the complete accuracy of reports. Prior to delivering any electronic files or hard-copy score reports, all reports underwent one final, extensive quality check, known as a "Red Team

Review." Red Teams are comprised of individuals from every CTB department coming together to form an interdisciplinary team. Samples of each type of report are printed from the active scoring system, and the Red Team carefully reviews these samples for accuracy and correct format, as well as a number of other issues including:

- Verify contents of reports against scoring specifications, report schematics and the Department approve format

- Reports print on correct form/color

- Reports collate correctly

- Data reported is reasonable (A complete data reasonableness check done by Research is completed prior to Red Team Review)

- Student-level data is accurate, compared by hand with student rosters and other documentation

- Required footnotes are in place

- Proficiency ranges reported match with scaled score ranges

- Cut scores are correct

- Reports are not sent out until all necessary corrections determined by the Red Team are resolved and samples of all reports sent to the Department are approved for distribution.

<u>Handscoring Process</u>

For MSA, the electronic handscoring system is used to score constructed response (CR) items. The imaging handscoring system presents images of scanned test books to trained readers, who assign scores for constructed response items. Scanned output is viewed on high quality 19″ workstation monitors. Images of each student's responses are automatically routed to two or more readers when required, and images of specific subsets of test items are routed to designated groups of readers trained to score these items. In addition to increased reader reliability, significant gains in reader productivity are noticed following the implementation of this technology.
CTB is committed to using the finest imaging equipment, software presentation system, data management system, and quality control to provide valid, reliable, cost-efficient scoring.

Constructed-Response Scorers

Scoring Personnel

CTB recruits, trains, and manages a sufficient number of staff to complete all handscoring operations within the time lines of this contract. CTB's experience involves extensive consultations between CTB Scoring, Publishing, and the customer to review scoring rubrics, develop anchor papers and other reader training materials, and provide analyses of student responses to tryout forms.

Readers

Many CTB readers have a great deal of classroom teaching experience. Our reader pool includes editors, published authors, and a number of individuals with advanced degrees. The minimum qualification for all Scoring Center readers is a Bachelor's degree.

All MSA CR items are scored in Delran, NJ. Handscoring readers were recruited from the southern New Jersey and Philadelphia areas. In order to work as a Handscoring reader at CTB, one must possess, and show evidence, of having either a BA or BS degree. The evaluator staff is comprised of individuals from many walks of life -- from retired or current educators to engineers, all possessing BAs to PhDs.

Team Leaders

Team leaders are selected on the basis of having demonstrated a high degree of scoring accuracy and consistency, often across multiple subjects and grades. They must also possess good interpersonal and leadership skills in order to be effective when training and counseling readers. The ratio of readers to team leaders is no more than 10 to 1. While it is possible to conduct handscoring with more readers per team leader, it has been CTB's experience that inter-rater reliability and production goals are jeopardized unless a trained leader can frequently monitor all readers.

Scoring Supervisors

Scoring Supervisors are the core group at CTB scoring centers. They direct and organize the assessment process, and train team leaders and readers. Scoring Supervisors have extensive experience as Team Leaders prior to their qualification and selection. The Scoring Supervisors are subject area experts in the content(s) that they supervise and train.

Anchor and Training Papers

Prior to the actual scoring, the CTB Scoring Center creates training materials. CR items for the MSA are assessed using MSDE holistic rubric with an X-point score scale. CTB randomly samples student answer documents to ensure that we are looking at a representative sample of the possible responses. A Rangefinder meeting is held with

MSDE staff and representatives to select sample papers of each score point. These samples are used to construct scoring guides and training papers. CTB's scoring team collaborates with MSDE to make any revisions to the rubrics and selection of scoring guide and training papers.

The process includes several presorting steps and subsequent iterative/consensus processes in order to achieve ever-increasing agreement and precision through a kind of "round robin" scoring, followed by discussion and selection.

When all papers for a form are selected and assigned status as good anchors training, qualifying, or check-set papers, they are consolidated into training formats.
Once approved by MSDE, the Scoring Guides (consisting of rubrics, anchors, and annotations) serves as a constant, setting the course for all subsequent training and scoring.

Training

Validation is a critical task in the assessment training process. It is the final determinant in reader readiness. All readers, including team leaders, must achieve 80 percent exact agreement on the qualifying round following training. Those readers not validating on the first attempt receive further training prior to taking an additional qualifying round. Only those training who successfully validate are qualified as readers and could score tests. Team leaders are required to complete two validation rounds with 80 percent exact agreement in each round.

Intra-rater Reliability

Throughout the course of the handscoring process, calibration sets of pre-scored papers (check-sets) are administered daily to the team leaders as well as to the readers, to monitor scoring accuracy and to maintain a consistent focus on the established rubric and guidelines. Imaging permits this monitoring without reader knowledge of when a check-set is administered. Readers whose check-set scores fall below the qualifying level are removed from live scoring and are given additional training and another qualifying (validation) round. Readers unable to qualify are dismissed.

The "read-behind" is another valuable intra-rater reliability monitoring technique. On a daily basis, each team leader reads a random selection of each reader's scored items. The scores are compared, and if they agree, the team leader is able to offer feedback, which enhances the reader's confidence and ability to score quickly and accurately. However, if an individual is straying from the standard established in the training and validation samples, the aberrant scoring is detected, and the team leader is able to offer the guidance necessary to refocus the reader's effort. Readers whose scoring is inconsistent are read behind more frequently by their team leaders. Thus, any scoring variation is corrected.

Inter-rater Reliability

Each constructed response is scored by at least two readers, and inter-rater reliability is monitored throughout the scoring process. If the scores of the two assigned readers differ by one point, the student will receive the higher of the two scores. If the scores of the two readers differ by more than one point, a third rating is provided by an expert rater, who will resolve the discrepancy and assign a final score.

<u>Characteristics of the Test Population</u>

Table 3 shows the ethnic characteristics of the students who took the 2006 MSA. Because percentages are rounded up to whole numbers, the percentages in this table do not always sum to 100. Among the Mathematics examinees, 47 to 50 percent were White, 37 to 40 percent were African American, and 7 to 9 percent were Hispanic. As expected, these percentages were similar across all test forms within a grade, because the test forms were spiraled within the classrooms. As shown in Table 4, there were slightly more male students than female students. The 2006 distributions of ethnicity and gender for the Mathematics tests are essentially the same as the 2003, 2004, and 2005 distributions.

Table 3
2006 MSA Ethnic Composition by Grade Level and Test Form*

Grade	Test Form	Number of Students**	Percent White	Percent African American	Percent Hispanic	Percent Others
3	A	12388	48	37	9	6
	B	12213	48	38	9	6
	C	12087	48	38	8	6
	D	11907	48	37	9	6
	E	11793	48	38	8	6
	Total	60388	48	38	8	6
4	A	12622	49	37	9	6
	B	12517	48	38	8	6
	C	12337	49	37	8	6
	D	12257	48	37	8	6
	E	12052	49	38	8	6
	Total	61785	49	37	8	6
5	A	12960	48	38	8	6
	B	12818	49	39	7	5
	C	12715	49	38	8	6
	D	12554	48	38	8	6
	E	12426	49	38	8	6
	Total	63473	49	38	8	6
6	A	13242	47	39	8	5
	B	13000	48	39	8	5
	C	12916	48	40	7	6
	D	12828	47	40	7	6
	E	12764	47	40	7	6
	Total	64750	48	40	7	5
7	A	13429	48	40	7	5
	B	13249	48	40	7	5
	C	13113	48	39	7	5
	D	13047	48	40	7	5
	E	12991	48	39	7	5
	Total	65829	48	40	7	5
8	A	13802	48	40	7	5
	B	13593	50	39	7	5
	C	13533	49	39	7	5
	D	13440	49	39	7	5
	E	13372	49	39	7	6
	Total	67740	49	39	7	5

* Because percentages are rounded to whole numbers, some rows may not sum to 100.
**Students of unspecified ethnicity are not included in this table.

Table 4
2006 MSA Student Gender by Grade Level and Test Form*

Grade	Test Form	Number of Students**	Percent Male	Percent Female
3	A	12388	53	47
	B	12213	52	48
	C	12087	51	49
	D	11907	51	49
	E	11793	50	50
	Total	60388	51	49
4	A	12622	52	48
	B	12517	51	49
	C	12337	51	49
	D	12257	50	50
	E	12052	51	49
	Total	61785	51	49
5	A	12960	52	48
	B	12818	51	49
	C	12715	51	49
	D	12554	51	49
	E	12426	51	49
	Total	63473	51	49
6	A	13242	53	47
	B	13000	52	48
	C	12916	51	49
	D	12828	51	49
	E	12764	52	48
	Total	64750	52	48
7	A	13429	52	48
	B	13249	51	49
	C	13113	51	49
	D	13047	51	49
	E	12991	51	49
	Total	65829	51	49
8	A	13802	52	48
	B	13593	52	48
	C	13533	51	49
	D	13440	51	49
	E	13372	52	48
	Total	67740	52	48

*Students who did not specify gender are not included in this table.

Norm Referenced Test (NRT)

NRT Test Design

In 2006, the MSA Mathematics tests included the *TerraNova* Mathematics Survey (TN) Form C at Grades 3, 4, 5, 7, and 8 and Form D at Grade 6. CTB's *TerraNova* is an assessment system designed to measure concepts, processes, and skills taught throughout the nation. *TerraNova* Survey consists of SR items only. The number of items and scale score ranges can be found in Table 5. *TerraNova* Mathematics scale scores based on IRT pattern scoring were reported. Scores on the *TerraNova* English Language Arts Survey were not reported in 2005.

Table 5
The Number of Items and Scale Score Range

Content Grade	SR Items	Scale Score Range
MA3	30	385-740
MA4	32	403-770
MA5	32	430-797
MA6	31	477-820
MA7	32	487-850
MA8	31	502-872

MA: Mathematics

Distributions of NRT Scores

NRT summary statistics for raw score (NCS), scale score (SS), national percentile rank (NP), and performance level are presented in Tables 6-9[1]. The NP shows that Maryland students' performance on the NRT was higher (55th through 63rd percentile) than the national average. As can be seen from NCS, SS, and NP in Tables 8 and 9, students' 2006 performance at most grade levels was somewhat higher than in 2005. Note that performance cuts in Table 9 were obtained from the *TerraNova* standard setting, not the Maryland standard setting.

Tables 10 and 11 show the scale score statistics (including the mean, standard deviation, minimum and maximum) for ethnicity and gender subgroups on each form. Overall, White students performed better than the other ethnic groups. There was almost one standard deviation (40 points) difference between the scores of white students and African American students in many grades. Note that *TerraNova* scores are vertically scaled so that scale scores across grades can be compared. On average across grades,

---

[1] Note that case counts for the NRT are lower than for the CRT because NRT scores were not computed for students who attempted fewer than 5 *TerraNova* items.

standard deviations were larger for White and Hispanic students than for African-American students and were larger for males than for females.

Figure 1 shows the 2006 Mathematics mean scale scores by grade level for each ethnic group. The increases in NRT score means from year to year appear to be similar for African-American and Hispanic groups. The year-to-year score increases for African-American and Hispanic students were similar to those for White students overall, but the African-American and White students showed small score increases between Grade 6 and Grade 7, while the Hispanic students actually showed a slight score drop between Grade 6 and Grade 7.

Tables 12 to 17 show the proportion of students answering each NRT item correctly (i.e., item p-values) in 2005 and 2006 on the *TerraNova* Mathematics items in Grades 3 through 8. For most items in Grades 3 through 7, p-values were higher in 2006 than in 2005, indicating an improvement in student performance. At Grade 8, the 2005 and 2006 average p-values were approximately the same, with some items showing increases and others showing decreases in p-values between the two years.

Table 6
NRT Summary Statistics based on Number-Correct Scores

Grade	N	Mean	SD	Skewness	Kurtosis	KR20	SEM
3	60269	24.17	4.387	-0.951	0.735	0.818	1.874
4	61666	23.65	5.772	-0.575	-0.379	0.862	2.147
5	63369	24.91	5.547	-0.846	0.068	0.860	2.078
6	64355	21.84	6.424	-0.576	-0.613	0.877	2.249
7	65253	21.77	6.943	-0.439	-0.810	0.891	2.292
8	67126	20.88	5.981	-0.276	-0.729	0.855	2.277

Table 7
NRT Summary Statistics based on Scale Scores and National Percentile Rank (NP)

Grade	N	Mean	SD	Skewness	Kurtosis	NP
3	60269	621	47.4	0.451	1.307	61
4	61666	641	45.7	0.505	2.332	58
5	63369	668	53.6	0.399	0.728	63
6	64355	676	51.1	-0.018	1.838	59
7	65253	681	54.3	0.415	2.450	55
8	67126	701	52.5	0.358	1.840	58

Table 8
NRT Summary Statistics: 2005 and 2006 State Means

Grade	2005			2006		
	NCS	SS	NP	NCS	SS	NP
3	23.966	619	61	24.169	621	61
4	23.017	636	57	23.648	641	58
5	24.281	662	61	24.913	668	63
6	21.340	673	58	21.840	676	59
7	21.239	677	53	21.774	681	55
8	20.745	700	58	20.878	701	58

NCS = number-correct score (i.e., raw score)
SS = scale score
NP = national percentile rank

Table 9
Percentages of Students in Each NRT Performance Level in 2005 and 2006

Grade	2005 Performance Level					2006 Performance Level				
	1	2	3	4	5	1	2	3	4	5
3	31.9	35.3	22.4	4.9	5.5	30.8	34.1	23.3	5.4	6.4
4	16.6	33.6	30.9	11.9	7.0	14.0	32.4	32.3	12.9	8.5
5	9.7	20.3	27.1	22.1	20.9	7.9	18.0	26.7	23.4	24.0
6	28.3	24.2	27.3	14.0	6.1	26.0	23.4	28.3	15.2	7.2
7	27.8	22.7	26.1	16.0	7.4	25.2	22.4	26.8	17.2	8.4
8	13.7	17.2	27.7	24.6	16.7	13.9	17.4	26.3	24.4	18.0

Table 10
NRT Scale Score Descriptive Statistics by Ethnicity

Grade Content	Test Form	White					African American					Hispanic				
		N	Mean	SD	Min	Max	N	Mean	SD	Min	Max	N	Mean	SD	Min	Max
MA3	A	5911	633.44	47.61	385	740	4615	600.99	41.36	385	740	1089	605.50	41.21	385	740
	B	5800	636.20	46.38	385	740	4610	603.87	41.23	385	740	1041	608.84	42.59	385	740
	C	5761	636.16	47.04	385	740	4538	604.54	41.25	385	740	992	610.77	41.99	385	740
	D	5717	635.84	47.24	385	740	4456	604.02	41.92	385	740	1025	608.98	38.94	458	740
	E	5650	635.56	46.50	385	740	4419	602.59	41.86	385	740	956	605.91	40.26	437	740
	Total	28839	635.43	46.97	385	740	22638	603.20	41.54	385	740	5103	607.98	41.06	385	740
MA4	A	6123	651.45	45.22	403	770	4674	620.75	38.98	403	770	1079	625.40	40.27	403	770
	B	6036	653.80	45.04	403	770	4711	622.97	37.60	403	770	1041	628.93	40.90	403	770
	C	6038	654.19	43.78	403	770	4543	624.20	38.29	403	770	1014	629.71	39.56	403	770
	D	5929	654.51	45.21	403	770	4536	623.42	38.59	403	770	1022	627.89	43.17	403	770
	E	5861	654.78	45.70	403	770	4518	622.70	37.84	403	770	972	628.30	41.53	403	770
	Total	29987	653.73	45.00	403	770	22982	622.80	38.28	403	770	5128	628.02	41.10	403	770
MA5	A	6268	680.57	53.66	430	797	4893	643.76	46.16	430	797	1058	650.60	50.71	430	797
	B	6219	684.68	52.05	430	797	4972	646.44	46.40	430	797	937	654.35	50.98	430	797
	C	6175	684.91	51.88	430	797	4789	646.29	44.95	430	797	990	653.26	47.18	486	797
	D	6077	684.41	51.38	430	797	4760	646.36	45.22	430	797	946	656.81	49.95	430	797
	E	6027	685.49	51.56	519	797	4674	645.18	46.33	430	797	987	655.07	49.61	430	797
	Total	30766	684.00	52.15	430	797	24088	645.61	45.83	430	797	4918	653.94	49.73	430	797
MA6	A	6248	689.51	49.22	477	820	5178	652.70	46.58	477	820	1020	659.99	50.27	477	820
	B	6239	693.12	48.03	477	820	5049	656.75	45.55	477	820	987	664.77	46.18	477	820
	C	6143	693.09	47.31	477	820	5058	655.63	46.32	477	820	906	664.85	46.98	477	820
	D	6075	693.42	46.76	477	820	5053	655.98	46.18	477	820	912	664.27	51.33	477	820
	E	5990	693.68	47.20	477	820	4994	655.00	46.03	477	820	950	662.78	49.74	477	820
	Total	30695	692.55	47.74	477	820	25332	655.20	46.15	477	820	4775	663.27	48.95	477	820

Table 10 (cont.)
NRT Scale Score Descriptive Statistics by Ethnicity

Grade Content	Test Form	White					African American					Hispanic				
		N	Mean	SD	Min	Max	N	Mean	SD	Min	Max	N	Mean	SD	Min	Max
MA7	A	6428	695.88	52.90	487	850	5259	657.00	46.82	487	850	953	658.79	47.88	487	850
	B	6349	697.38	52.49	487	850	5156	659.04	46.11	487	850	898	658.06	51.16	487	850
	C	6293	699.06	51.81	487	850	5081	660.32	45.95	487	850	942	661.55	47.20	487	850
	D	6209	699.14	52.60	487	850	5096	658.16	47.04	487	850	902	665.34	50.21	487	850
	E	6260	698.03	52.24	487	850	5033	658.66	46.55	487	850	899	660.30	50.93	487	850
	Total	31539	697.89	52.42	487	850	25625	658.62	46.51	487	850	4594	660.79	49.51	487	850
MA8	A	6652	715.40	51.93	502	872	5394	674.97	45.23	502	872	940	685.66	47.70	502	872
	B	6712	718.98	50.70	502	872	5166	676.37	42.60	502	872	879	682.67	49.40	502	872
	C	6595	718.39	49.02	502	872	5226	678.39	43.23	502	872	880	681.97	49.99	502	872
	D	6557	718.88	50.30	502	872	5172	677.55	43.11	502	872	872	688.62	45.01	502	872
	E	6488	719.30	50.32	502	872	5133	677.50	42.63	502	872	887	682.27	45.89	502	872
	Total	33004	718.18	50.48	502	872	26091	676.94	43.40	502	872	4458	684.25	47.69	502	872

Table 11
NRT Scale Score Descriptive Statistics by Gender

Grade Content	Test Form	Male					Female				
		N	Mean	SD	MIN	MAX	N	Mean	SD	MIN	MAX
MA3	A	6514	620.99	49.95	385	740	5845	617.31	45.40	385	740
	B	6292	624.30	48.94	385	740	5896	619.49	45.12	385	740
	C	6166	624.52	48.75	385	740	5890	620.50	45.25	385	740
	D	6038	624.34	49.29	385	740	5850	619.96	45.44	385	740
	E	5933	622.45	48.82	385	740	5840	620.05	45.95	385	740
	Total	30943	623.30	49.18	385	740	29321	619.46	45.44	385	740
MA4	A	6560	639.20	47.82	403	770	6034	638.04	43.92	403	770
	B	6363	641.42	46.94	403	770	6128	640.52	43.84	403	770
	C	6260	643.57	45.72	403	770	6052	640.55	43.92	403	770
	D	6127	642.82	47.64	403	770	6100	640.18	44.16	403	770
	E	6089	642.60	47.30	403	770	5947	640.24	44.57	403	770
	Total	31399	641.89	47.11	403	770	30261	639.91	44.09	403	770
MA5	A	6687	665.76	55.28	430	797	6246	664.43	53.16	430	797
	B	6571	669.14	54.83	430	797	6232	667.59	52.31	430	797
	C	6456	668.90	53.79	430	797	6245	668.36	51.87	430	797
	D	6327	670.06	54.06	430	797	6200	667.81	52.35	430	797
	E	6374	670.99	54.67	430	797	6025	666.57	52.74	430	797
	Total	32415	668.94	54.56	430	797	30948	666.95	52.50	430	797
MA6	A	6941	673.14	53.25	477	820	6217	674.10	50.37	477	820
	B	6679	677.09	51.51	477	820	6228	677.87	49.65	477	820
	C	6576	677.35	52.72	477	820	6258	676.98	49.05	477	820
	D	6481	677.23	51.02	477	820	6279	677.28	50.36	477	820
	E	6583	677.16	52.86	477	820	6102	676.75	49.42	477	820
	Total	33260	676.36	52.31	477	820	31084	676.60	49.79	477	820
MA7	A	6935	678.62	56.13	487	850	6383	678.99	52.38	487	850
	B	6726	680.81	55.88	487	850	6404	679.96	52.97	487	850
	C	6590	682.98	55.11	487	850	6417	680.91	51.58	487	850
	D	6597	681.23	56.67	487	850	6326	681.55	52.93	487	850
	E	6553	681.66	55.89	487	850	6313	679.81	52.67	487	850
	Total	33401	681.03	55.95	487	850	31843	680.24	52.51	487	850
MA8	A	7027	701.14	56.48	502	872	6640	695.07	49.59	502	872
	B	6988	703.34	55.80	502	872	6486	698.87	49.62	502	872
	C	6856	704.05	54.82	502	872	6551	698.26	48.17	502	872
	D	6809	705.08	55.48	502	872	6509	698.44	48.62	502	872
	E	6826	704.79	54.93	502	872	6422	697.69	48.49	502	872
	Total	34506	703.66	55.53	502	872	32608	697.65	48.92	502	872

Figure 1
NRT Mathematics Mean Scale Scores by Grade and Ethnicity
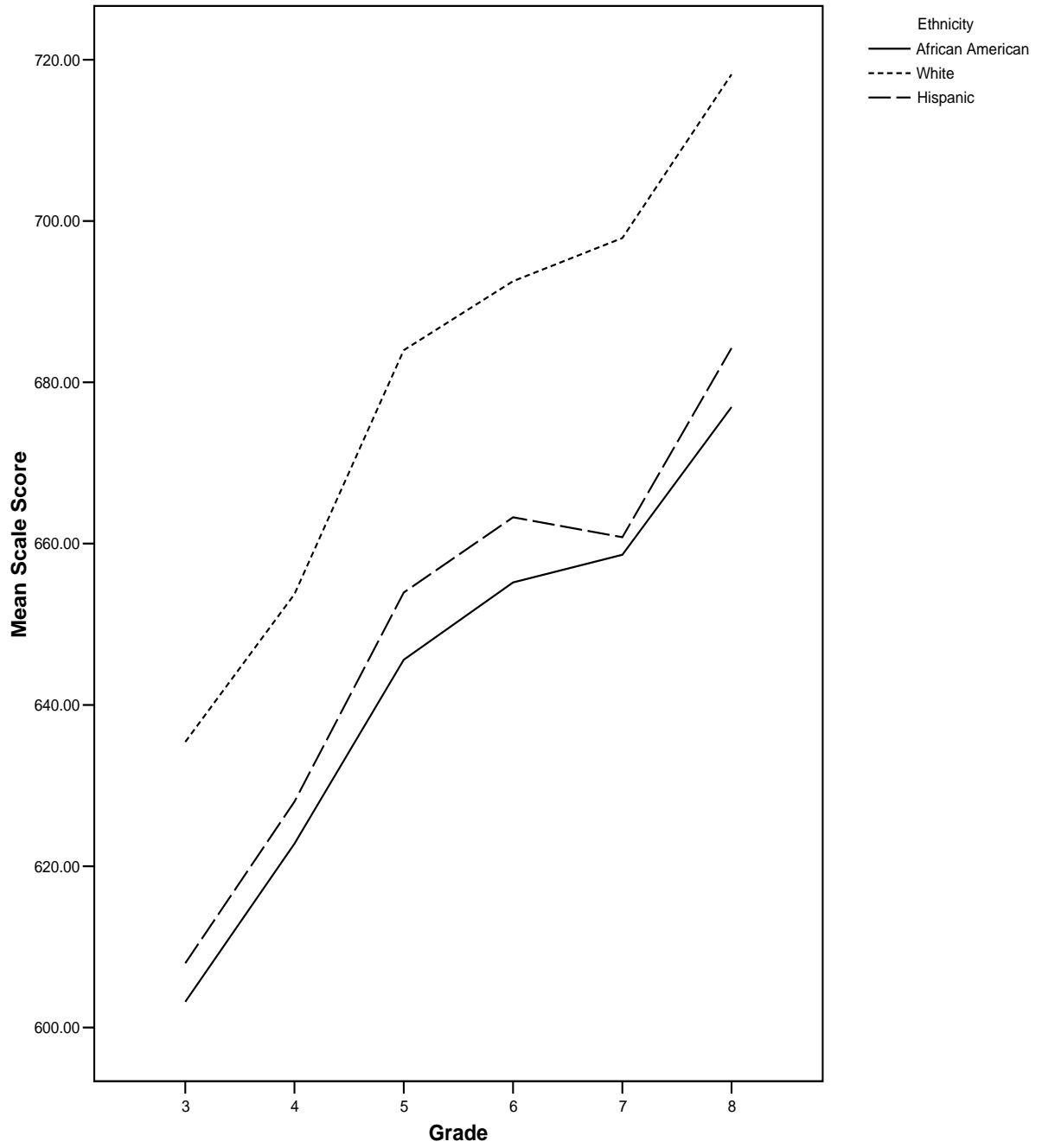
Table 12
Grade 3 Mathematics NRT Item p-values

Item	2005 (N=61509)	2006 (N=60388)	Difference
1	0.86	0.87	0.01
2	0.92	0.93	0.01
3	0.91	0.92	0.01
4	0.89	0.90	0.01
5	0.76	0.77	0.01
6	0.96	0.96	0.00
7	0.69	0.70	0.01
8	0.72	0.72	0.00
9	0.83	0.79	-0.04
10	0.90	0.91	0.01
11	0.91	0.91	0.00
12	0.90	0.91	0.01
13	0.88	0.88	0.00
14	0.88	0.90	0.02
15	0.86	0.87	0.01
16	0.96	0.97	0.01
17	0.91	0.92	0.01
18	0.97	0.97	0.00
19	0.68	0.69	0.01
20	0.63	0.63	0.00
21	0.57	0.58	0.01
22	0.75	0.75	0.00
23	0.88	0.89	0.01
24	0.97	0.98	0.01
25	0.91	0.92	0.01
26	0.73	0.74	0.01
27	0.71	0.71	0.00
28	0.51	0.53	0.02
29	0.41	0.42	0.01
30	0.46	0.49	0.03
Average	0.80	0.80	0.01

Table 13
Grade 4 Mathematics NRT Item p-values

Item	2005 (N=63330)	2006 (N=61785)	Difference
1	0.81	0.84	0.03
2	0.66	0.70	0.04
3	0.83	0.86	0.03
4	0.60	0.62	0.02
5	0.55	0.57	0.02
6	0.82	0.83	0.01
7	0.70	0.72	0.02
8	0.52	0.53	0.01
9	0.56	0.59	0.03
10	0.53	0.55	0.02
11	0.98	0.98	0.00
12	0.95	0.96	0.01
13	0.38	0.39	0.01
14	0.73	0.75	0.02
15	0.82	0.83	0.01
16	0.85	0.87	0.02
17	0.91	0.92	0.01
18	0.56	0.59	0.03
19	0.89	0.90	0.01
20	0.74	0.75	0.01
21	0.66	0.71	0.05
22	0.84	0.85	0.01
23	0.68	0.70	0.02
24	0.86	0.87	0.01
25	0.85	0.87	0.02
26	0.71	0.73	0.02
27	0.61	0.65	0.04
28	0.87	0.89	0.02
29	0.64	0.64	0.00
30	0.82	0.84	0.02
31	0.51	0.53	0.02
32	0.53	0.57	0.04
Average	0.72	0.74	0.02

Table 14
Grade 5 Mathematics NRT Item p-values

Item	2005 (N=65123)	2006 (N=63473)	Difference
1	0.85	0.87	0.02
2	0.74	0.78	0.04
3	0.78	0.81	0.03
4	0.69	0.72	0.03
5	0.73	0.74	0.01
6	0.81	0.84	0.03
7	0.77	0.79	0.02
8	0.63	0.65	0.02
9	0.92	0.93	0.01
10	0.92	0.93	0.01
11	0.94	0.94	0.00
12	0.75	0.78	0.03
13	0.81	0.81	0.00
14	0.71	0.73	0.02
15	0.82	0.83	0.01
16	0.91	0.92	0.01
17	0.67	0.66	-0.01
18	0.98	0.98	0.00
19	0.62	0.64	0.02
20	0.95	0.96	0.01
21	0.82	0.84	0.02
22	0.75	0.76	0.01
23	0.68	0.71	0.03
24	0.58	0.62	0.04
25	0.56	0.57	0.01
26	0.55	0.60	0.05
27	0.66	0.70	0.04
28	0.66	0.68	0.02
29	0.81	0.82	0.01
30	0.73	0.75	0.02
31	0.61	0.64	0.03
32	0.84	0.87	0.03
Average	0.76	0.78	0.02

Table 15
Grade 6 Mathematics NRT Item p-values

Item	2005 (N=65846)	2006 (N=64750)	Difference
1	0.74	0.77	0.03
2	0.78	0.79	0.01
3	0.70	0.72	0.02
4	0.75	0.77	0.02
5	0.81	0.82	0.01
6	0.71	0.73	0.02
7	0.56	0.60	0.04
8	0.75	0.76	0.01
9	0.92	0.93	0.01
10	0.83	0.83	0.00
11	0.56	0.59	0.03
12	0.65	0.68	0.03
13	0.79	0.81	0.02
14	0.78	0.78	0.00
15	0.76	0.77	0.01
16	0.71	0.73	0.02
17	0.74	0.76	0.02
18	0.75	0.77	0.02
19	0.69	0.70	0.01
20	0.70	0.72	0.02
21	0.53	0.55	0.02
22	0.58	0.60	0.02
23	0.58	0.59	0.01
24	0.69	0.70	0.01
25	0.63	0.65	0.02
26	0.61	0.65	0.04
27	0.64	0.64	0.00
28	0.56	0.56	0.00
29	0.52	0.56	0.04
30	0.58	0.63	0.05
31	0.56	0.56	0.00
Average	0.68	0.70	0.02

Table 16
Grade 7 Mathematics NRT Item p-values

Item	2005 (N=68130)	2006 (N=65829)	Difference
1	0.73	0.76	0.03
2	0.65	0.67	0.02
3	0.52	0.54	0.02
4	0.81	0.83	0.02
5	0.73	0.75	0.02
6	0.82	0.84	0.02
7	0.50	0.51	0.01
8	0.63	0.64	0.01
9	0.69	0.70	0.01
10	0.90	0.91	0.01
11	0.67	0.68	0.01
12	0.74	0.74	0.00
13	0.62	0.65	0.03
14	0.57	0.58	0.01
15	0.89	0.90	0.01
16	0.57	0.64	0.07
17	0.72	0.74	0.02
18	0.67	0.69	0.02
19	0.68	0.69	0.01
20	0.68	0.70	0.02
21	0.81	0.83	0.02
22	0.72	0.73	0.01
23	0.74	0.77	0.03
24	0.59	0.60	0.01
25	0.63	0.64	0.01
26	0.52	0.55	0.03
27	0.67	0.68	0.01
28	0.54	0.56	0.02
29	0.51	0.53	0.02
30	0.38	0.40	0.02
31	0.55	0.59	0.04
32	0.52	0.54	0.02
Average	0.66	0.67	0.02

Table 17
Grade 8 Mathematics NRT Item p-values

Item	2005 (N=68696)	2006 (N=67740)	Difference
1	0.81	0.81	0.00
2	0.57	0.60	0.03
3	0.32	0.34	0.02
4	0.84	0.83	-0.01
5	0.76	0.75	-0.01
6	0.65	0.66	0.01
7	0.64	0.64	0.00
8	0.83	0.84	0.01
9	0.66	0.65	-0.01
10	0.89	0.89	0.00
11	0.90	0.90	0.00
12	0.79	0.79	0.00
13	0.88	0.90	0.02
14	0.76	0.76	0.00
15	0.64	0.65	0.01
16	0.87	0.88	0.01
17	0.75	0.77	0.02
18	0.58	0.59	0.01
19	0.63	0.63	0.00
20	0.42	0.44	0.02
21	0.74	0.75	0.01
22	0.70	0.72	0.02
23	0.72	0.70	-0.02
24	0.52	0.51	-0.01
25	0.63	0.64	0.01
26	0.42	0.41	-0.01
27	0.59	0.61	0.02
28	0.53	0.53	0.00
29	0.55	0.56	0.01
30	0.43	0.43	0.00
31	0.49	0.51	0.02
Average	0.66	0.67	0.01

Criterion Referenced Test (CRT)

CRT Test Design

The MSA Criterion-Referenced Test is composed of *TerraNova* items that are closely aligned with the Maryland content standards, plus custom selected-response (SR) and constructed-response (CR) items written to measure performance on the Maryland content standards. The Mathematics tests in Grades 7 and 8 also contain student-produced-response (SPR) items, sometimes referred to as "gridded response" items. *TerraNova* Form D was administered in Grade 6; *TerraNova* Form C was administered in all other grades.

Table 18 shows the number of items, by item type, in each test form. The column "SR from NRT" in that table shows the number of NRT items that contribute to CRT scores. For the Mathematics tests, Forms A, C, and E contain the same operational items and are designated as Form 1; similarly, Forms B and D contain the same operational items and are designated as Form 2. [2] As can be seen in Table 18, the total number of operational items and score points was the same for all test forms within a grade.

Table 19 shows the number of items by item function (anchor items, common items, unique items, and field test items). Anchor items were used for placing the 2006 scale on the 2005 scale. Common items (which included many, but not necessarily all, of the anchor items) were used for linking alternate forms.

Tables 20 to 25 present the number of items and score points by Maryland content reporting standards. There are five reporting standards for Mathematics across grades. For all grades, the number of items and score points for each reporting standard were identical across forms within each grade. The actual values shown in Tables 20 to 25 align with the target values (shown in Table 1) for all grades and the sums in these tables are identical to the values shown in Table 18.

---

[2] The forms designated as operational Form 1 contain the same operational items in the same item positions, and are identical to one another except for the field test items included in Section 5 of each form. This is also true of the forms designated as operational Form 2. Although Forms 1 and 2 are distinct operational forms, they also share some common items.

Table 18
The Number of Items by Item Type

Grade Content	Form	CRT				Total CRT Items	Total CRT Score Points
		SR from NRT	SR	CR	SPR		
MA3	1	11	40	14	-	65	72
	2	11	40	14	-	65	72
MA4	1	10	40	14	-	64	71
	2	10	40	14	-	64	71
MA5	1	13	36	16	-	65	74
	2	13	36	16	-	65	74
MA6	1	5	43	14	-	62	70
	2	5	43	14	-	62	70
MA7	1	6	30	14	12	62	72
	2	6	30	14	12	62	72
MA8	1	11	25	16	12	64	75
	2	11	25	16	12	64	75

- For all grades, Form 1 consists of Forms A, C, & E and Form 2 consists of Forms B & D.
- For all grades, counts are without field test items.

Table 19
The Number of Items by Function

Content Grade	Form	Total Items*	Anchor Items	Common Items	Unique Items	Field-Test Items
MA3	A	83	26	39	26	18
	B	83	26	39	26	18
	C	83	26	39	26	18
	D	83	26	39	26	18
	E	83	26	39	26	18
MA4	A	82	26	32	32	18
	B	82	26	32	32	18
	C	82	26	32	32	18
	D	82	26	32	32	18
	E	82	26	32	32	18
MA5	A	85	27	40	25	20
	B	85	27	40	25	20
	C	85	27	40	25	20
	D	81	27	40	25	16
	E	81	27	40	25	16
MA6	A	77	27	31	31	15
	B	77	27	31	31	15
	C	78	27	31	31	16
	D	78	27	31	31	16
	E	78	27	31	31	16
MA7	A	78	23	34	28	16
	B	76	23	34	28	14
	C	79	23	34	28	17
	D	79	23	34	28	17
	E	79	23	34	28	17
MA8	A	81	22	38	26	17
	B	79	22	38	26	15
	C	80	22	38	26	16
	D	80	22	38	26	16
	E	78	22	38	26	14

- * Total = Common + Unique + Field Test
- For all grades, common items are items that appear both on Form 1 (Forms A, C, & E) and Form 2 (Forms B, & D).

Table 20
The Number of Items and Score Points by Maryland Content Standard for Grade 3

| | Forms A, C & E | | | | | | | Forms B & D | | | | | | |
| | NRT | Custom | | Total | | | | NRT | Custom | | Total | | | |
Standards	SR	SR	CR	Items	%	Points	%	SR	SR	CR	Items	%	Points	%
01	1	11	1	13	20	13	18	1	11	1	13	20	13	18
02/03	4	9	2	15	23	15	21	4	9	2	15	23	15	21
04/05	1	12	1	14	22	14	19	1	12	1	14	22	14	19
06	5	8	3	16	25	16	22	5	8	3	16	25	16	22
07	0	0	7	7	11	14	19	0	0	7	7	11	14	19
Sum	11	40	14	65	100	72	100	11	40	14	65	100	72	100

Table 21
The Number of Items and Score Points by Maryland Content Standard for Grade 4

| | Forms A, C & E | | | | | | | Forms B & D | | | | | | |
| | NRT | Custom | | Total | | | | NRT | Custom | | Total | | | |
Standards	SR	SR	CR	Items	%	Points	%	SR	SR	CR	Items	%	Points	%
01	0	13	1	14	22	14	20	0	13	1	14	22	14	20
02/03	2	10	2	14	22	14	20	2	10	2	14	22	14	20
04/05	0	13	2	15	23	15	21	0	13	2	15	23	15	21
06	8	4	2	14	22	14	20	8	4	2	14	22	14	20
07	0	0	7	7	11	14	20	0	0	7	7	11	14	20
Sum	10	40	14	64	100	71	100	10	40	14	64	100	71	100

Table 22
The Number of Items and Score Points by Maryland Content Standard for Grade 5

| | Forms A, C & E | | | | | | | Forms B & D | | | | | | |
| | NRT | Custom | | Total | | | | NRT | Custom | | Total | | | |
Standards	SR	SR	CR	Items	%	Points	%	SR	SR	CR	Items	%	Points	%
01	2	11	2	15	23	15	20	2	11	2	15	23	15	20
02/03	4	8	2	14	22	14	19	4	8	2	14	22	14	19
04/05	2	9	2	13	20	13	18	2	9	2	13	20	13	18
06	5	8	2	15	23	15	20	5	8	2	15	23	15	20
07	0	0	8	8	12	17	23	0	0	8	8	12	17	23
Sum	13	36	16	65	100	74	100	13	36	16	65	100	74	100

Table 23
The Number of Items and Score Points by Maryland Content Standard for Grade 6

| | Forms A, C & E | | | | | | | Forms B & D | | | | | | |
| | NRT | Custom | | Total | | | | NRT | Custom | | Total | | | |
Standards	SR	SR	CR	Items	%	Points	%	SR	SR	CR	Items	%	Points	%
01	1	11	2	14	23	14	20	1	11	2	14	23	14	20
02/03	1	11	2	14	23	14	20	1	11	2	14	23	14	20
04/05	0	12	1	13	21	13	19	0	12	1	13	21	13	19
06	3	9	2	14	23	14	20	3	9	2	14	23	14	20
07	0	0	7	7	11	15	21	0	0	7	7	11	15	21
Sum	5	43	14	62	100	70	100	5	43	14	62	100	70	100

Table 24
The Number of Items and Score Points by Maryland Content Standard for Grade 7

| | Forms A, C & E | | | | | | | | Forms B & D | | | | | | | |
| | NRT | Custom | | | Total | | | | NRT | Custom | | | Total | | | |
Standards	SR	SR	CR	GR	Items	%	Points	%	SR	SR	CR	GR	Items	%	Points	%
01	0	9	2	3	14	23	14	19	0	9	2	3	14	23	14	19
02/03	1	7	2	3	13	21	13	18	1	7	2	3	13	21	13	18
04/05	0	8	3	3	14	23	14	19	0	8	3	3	14	23	14	19
06	5	6	0	3	14	23	14	19	5	6	0	3	14	23	14	19
07	0	0	7	0	7	11	17	24	0	0	7	0	7	11	17	24
Sum	6	30	14	12	62	100	72	100	6	30	14	12	62	100	72	100

Table 25
The Number of Items and Score Points by Maryland Content Standard for Grade 8

| | Form A, C & E | | | | | | | | Form B & D | | | | | | | |
| | NRT | Custom | | | Total | | | | NRT | Custom | | | Total | | | |
Standards	SR	SR	CR	GR	Items	%	Points	%	SR	SR	CR	GR	Items	%	Points	%
01	2	6	3	4	15	23	15	20	2	6	3	4	15	23	15	20
02/03	2	6	2	3	13	20	13	17	2	6	2	3	13	20	13	17
04/05	1	7	3	3	14	22	14	19	1	7	3	3	14	22	14	19
06	6	6	0	2	14	22	14	19	6	6	0	2	14	22	14	19
07	0	0	8	0	8	13	19	25	0	0	8	0	8	13	19	25
Sum	11	25	16	12	64	100	75	100	11	25	16	12	64	100	75	100

Classical Item Analysis

Tables A1- A18 of Appendix A present item-level descriptive statistics for each of the test forms. These tables contain the following information: item function (common or unique), item type (SR, CR, or SPR), item *p*-value (P_VAL), item correlation with the total test score (R_ITT), and correlation between each item choice and the total test score (P_BIS1, etc.). The *p*-value for an SR item represents the proportion of students who answered the item correctly. The *p*-value for a CR item represents the mean raw score for the item divided by the number of points possible for the item. A point-biserial correlation between the item score and the total score on the test was also computed for the SR items. For the CR items, a Pearson product-moment correlation between the item score and the total score on the test was computed. For the item analysis, the studied item was excluded from the computation of the total score so as to not inflate the correlation artificially. This effect would be most noticeable for CR items worth several points. For the correct answer choice, the correlation between item choice and total score is the same as the point-biserial correlation of the item. A similar formula was applied to compute the correlation between each distracter and the total score. In general, negative correlations are expected for all distracters when an item is good.

Note that items were evaluated using the following criteria: a *p*-value below 0.30 for SR items and 0.20 for CR and SPR items, and a point-biserial below 0.15. Items flagged for any of these criteria were referred to CTB's content specialists for further review to ensure that each item was measuring the intended construct(s), that the scoring key or scoring rubric was correct, and (for multiple-choice items) that there was one and only one correct answer to the item.

Rater Agreement

All CR items were scored by at least two raters. If the scores assigned by the raters differed by one point, the student received the higher of the two scores. Discrepancies of more than one point were resolved by a third expert rater.

Rater agreement was assessed using only the scores assigned by the first two raters.   Indices of rater agreement and consistency were obtained using the scores from the first two raters. Appendix tables B1-B6 present rater agreement statistics for the CR items across all grades. These tables provide the percentages of pairs of raters' scores that did not differ (i.e., perfect agreement) and the percentages of pairs of raters' scores that differed by one point (i.e., adjacent agreement) for all CR items over all test forms.

When rater agreement was defined as the percentage of same scores plus adjacent scores, rater agreement across all grade levels ranged from 97.6% to 100% for the Mathematics items. The percentage of perfect agreement (i.e., identical scores assigned by rater 1 and rater 2) ranged from 77.4% to 99.8% in Grade 3, from 68.6% to 99.1% in Grade 4, from 75.5% to 99.6% in Grade 5, from 73.7% to 99.5% in Grade 6, from 77.1% to 99.7% in Grade 7, and from 69.8% to 99.4% in Grade 8.

Note that each CR item for Mathematics consists of two parts, A and B. Because Part A is dichotomously scored (1 point for a correct response), the percentage of perfect agreement for part A was usually higher than for part B, ranging from 96.1% to 99.8% in Grade 3, 94.9% to 99.1% in Grade 4, 93.9% to 99.6% in Grade 5, 96.2% to 99.5% in Grade 6, 94.3% to 99.7% in Grade 7 and 94.8% to 99.4% in Grade 8.

In addition to the percentage of agreement, the tables present the mean item score and item standard deviation of the item scores assigned by each rater group. The mean score points awarded by the two rater groups are very close. The product moment correlations between first and second ratings are also included in these tables.

Appendix Tables B7-B12 show the distributions of scores on the CR items. In these tables, ITEMNO represents item number in test book. "Omit" denotes the number of student cases that did not respond to the item. Code B is an answer that cannot be scored. Each number, 0, 1, 2, 3, represents a score of 0, 1, 2, and 3, respectively. "%_omit" represents the percent of omits. Note that parts A and B of the Mathematics items were treated as independent items and were separately scored.

Differential Item Functioning (DIF)

An item flagged for differential item functioning (DIF) is more difficult for a particular group of students than would be expected based on their total test scores, compared to the performance of the other group. The groups compared in the DIF analyses were female and male students, and African–American, Hispanic, and white students. Male and white were reference groups.

The statistical procedures used by CTB to identify items thought to exhibit substantial DIF are the same procedures used by the Educational Testing Service (ETS) and the National Assessment of Educational Progress (NAEP). For SR items, the Mantel-Haenszel (χ^2_{MH}) statistic was used to evaluate potential DIF items. In this procedure, the "C" - level DIF items are flagged, where a "C" item indicates a large amount of DIF and has an absolute value of the Mantel-Haenszel (Δ_{MH}) that is significantly greater than zero (at the .05 level) and $|\Delta_{MH}|$ exceeds 1.5. Also, the "B" - level DIF items are flagged, where a "B" item indicates DIF and has an absolute value of the Mantel-Haenszel (Δ_{MH}) that is significantly greater than zero (at the .05 level) and $-1.5 \leq \Delta_{MH} \leq -1$ or $1 \leq \Delta_{MH} \leq 1.5$ (Zwick, Donoghue, & Grima, 1993).

For the CR items, an effect size (ES) statistic based on Mantel χ^2 was used. ES is obtained by dividing the standardized mean difference (SMD) statistics by the standard deviation of the item. A detailed description of these procedures can be found in Zwick, et al., (1993).

Tentative flagging criteria followed the same rules as are used in NAEP:
BB: If the Mantel statistic is significant (p < .05) and the |ES| is between 0.17 and 0.25
CC: If the Mantel statistic is significant (p < .05) and the |ES| ≥ 0.25

Appendix tables C1-C6 show items flagged based on the above criteria. In the column "Focal", for those items flagged for ethnicity, AA represents African American and Hisp represents Hispanic. Positive values in the "DIF" column mean that the item favors the focal group, while negative values imply that the item disadvantages the focal group.

Item Fit Assessment

Item fit was assessed using the *Q1* statistic described by Yen (1984). *Q1* is a Pearson chi-square statistic,

$$Q1_j = \sum_{i=1}^{I} \frac{N_{ji}(O_{ji} - E_{ji})^2}{E_{ji}} + \sum_{i=1}^{I} \frac{N_{ji}[(1 - O_{ji}) - (1 - E_{ji})]^2}{1 - E_{ji}}$$

where N_{ji} is the number of examinees in cell i for item j, and O_{ji} and E_{ji} are the observed and expected proportion of examinees in in cell i obtaining the maximum possible score on item j.

Because *Q1* is influenced by sample size and by the number of possible score points for an item, this statistic was transformed to a Z-statistic,

$$Z_j = \frac{(Q_{1j} - DF_j)}{\sqrt{2DF_j}}$$

where Q_{1j} is the item chi-square statistic defined above,

j is an item, and
DF is the degrees of freedom for a given item j.

The Z-statistic is an index of the degree to which obtained proportions of students with each item score are close to the proportions that would be predicted by the estimated student ability and item parameters. These values, along with the associated chi-squares (*Q1*) are computed for ten intervals corresponding to deciles of the ability distribution. Because the expected value of Z increases as the sample size increases, critical values for Z were established using the following equation (Yen, 1991a):

$$Z_{crit,j} = \frac{4N_j}{1500}$$

where $Z_{crit,j}$ is critical value of Z for item j and
N_j is the number of students who responded to item j.

In the 2006 calibration of the Mathematics items, several items exhibited moderate misfit. Across all operational test forms, one misfitting item was identified at Grade 3, five items at Grade 4, two at Grade 5, four at Grade 6, two at Grade 7, and nine at Grade 8. The figures in Appendix D show the estimated and observed item characteristic curves (ICC's) of these items. No items were dropped from scoring because of model misfit. Appendix D contains the plots of the field test items flagged for misfit as well.

Calibration and Equating

IRT Model

Student item responses were calibrated using the combination of two IRT models. The three-parameter logistic model (3PL) was used to scale the SR items, and the two-parameter partial credit (2PPC) model was employed to scale the CR items. A brief explanation of the models is provided below.

Two types of IRT models have most commonly been used to scale large-scale education assessments containing mixed item types or formats. For SR items, the 3PL model has been employed. The 3PL model (Lord & Novick, 1968; Lord, 1980) defines a SR item in terms of three item parameters: item difficulty or location, item discrimination, and probability of a student with very low ability answering the item correctly (guessing parameter). In this model, the probability that a student with scale score θ responds correctly to item j is

$$p_j(\theta) = c_j + \frac{(1 - c_j)}{1 + \exp[-1.7a_j(\theta - b_j)]},$$

where a_j is the item discrimination, b_j is the item difficulty, and c_j is the probability of a correct response by a very low-scoring student.

The 2PPC model defines a CR item in terms of item discrimination as well as location parameter for each score point. The 2PPC model is a special case of Bock's (1972) nominal model. Bock's model states that the probability of an examinee with ability θ having a score at the kth level of the jth item is

$$P_{jk}(\theta) = P(x_j = k - 1|\theta) = \frac{\exp Z_{jk}}{\sum_{i=1}^{m_j} \exp Z_{ji}}, k = 1,...,m_j,$$

where m_j is the number of score levels, and
$$Z_{jk} = A_{jk}\theta + C_{jk},$$
$$A_{jk} = \alpha_j(k - 1), \ k = 1, 2,...m_j, \text{ and}$$
$$C_{jk} = -\sum_{i=0}^{k-1} \gamma_{ji}, \text{ where } \gamma_{j0} = 0,$$

where A_{jk} is the discrimination parameter of the kth category of item j, C_{jk} is the intercept parameter of the nonlinear response function associated with the kth category of item j, α_j and γ_{ji} are the parameters to be estimated from the data.

For each item there are $m_j - 1$ independent γ_{ji} parameters and one α_j parameter; a total of m_j independent item parameters are estimated.

Calibration and Equating Procedure

In this report, **common items** indicate items that appear across all alternate forms and are used for Form-to-Form equating. **Anchor items** indicate items used for Year-to-Year equating. Most anchor items are common items. No constructed response (CR) items or student-produced response (SPR) items were used as anchor items. As in previous years, each Mathematics CR item is composed of two parts, A and B. Each part is considered one item.

The following procedures were applied to calibrate and equate the 2006 MSA CRT items:

Calibration and Form-to-Form equating

Only items that contribute to the CRT score were calibrated. The following two steps were applied for Form-to-Form equating.

Step 1: Stability of equating items was checked using following the procedure.

(1) Each of the two operational forms for each grade was separately calibrated. Plots of the Form 1 vs. Form 2 item parameters (*a* parameters (using log of *a*) and *b* parameters) were produced. These plots were examined to identify items that were not behaving consistently across forms. For the 2006 assessments, there was only one item (Grade 3, item #33) with inconsistent parameters across the two forms. On 5/3/06, MSDE approved the suppression of this item for the 2006 administration.

Step 2: Thus, all of the shared items other than grade 3, item #33 were treated as common items for purposes of calibration and equating, and the two alternate Forms 1 and 2 at each grade level were calibrated together.

Year-to-Year Equating

The following two steps were applied for Year-to-Year equating.

Step 1: Stability of anchor items was checked using the following procedure.

(1) Item parameters for the 2006 test forms were transformed to the MSA CRT reporting scale using the test characteristic curve procedure suggested by Stocking and Lord (1983).
(2) The original *a* and *b* parameters of the anchor items were plotted against the recalibrated parameters from the 2006 calibration. Item p-values were also plotted.

Step 2: Results were evaluated to determine whether or not all of the anchor items were stable enough across years to use for year-to-year equating. For the 2006 tests, all of the anchor items were judged to be sufficiently stable, an all were used as equating anchors. Item parameters for the 2006 tests were transformed to the MSA CRT reporting scale using these anchor items and Stocking and Lord's transformation procedure.

Calibration and Equating Results

The untransformed (theta metric) item parameters for all items are contained in Appendix E. Stability of common items was checked using the method described above in Step 1 of the Form-to-Form equating procedures. Figures F1-F6 in Appendix F show the alignment of "a" parameters (using the log of a) and the alignment of "b" parameters. Note that only selected response (SR) items were used for common items.   Based on these plots, all items were judged to be sufficiently stable to serve as common items for calibration and equating purposes. Please note that grade 3, item #33 had already been removed.

Figures F7-F24 show the item parameters and p-values by grade and test form. Figures F25-F30 show test characteristic curves (TCC) and standard errors of measurement (SEM) curves based on the final item parameters. TCCs and SEMs for alternate forms were similar across all grades.

Distribution of the Maryland Score Scale

Table 26 presents the lowest obtainable scale scores (LOSS) and the highest obtainable scale scores (HOSS). For the 2006 assessments, MSDE requested that the LOSS and HOSS values remain at a LOSS of 240 and HOSS of 650 across all grades.

Table 26
LOSS and HOSS

Grade	LOSS	HOSS
MA3	240	650
MA4	240	650
MA5	240	650
MA6	240	650
MA7	240	650
MA8	240	650
RD10	240	650

The 2006 item parameters were placed on the MSA CRT reporting scale using previously calibrated items from the 2004 and 2005 tests as anchors in a Stocking and Lord test-characteristic curve equating procedure (Stocking & Lord, 1983). Student scores were computed using IRT pattern scoring with the transformed parameters. As shown in Table 27, and 28, distributions of raw scores and scale scores were similar across forms. Due to relatively long test lengths for every grade, reliability (Cronbach's alpha) was high for all grades. Reliability coefficients ranged from 0.92 to 0.96 across grades.

Tables 29 and 30 show the scale score statistics (means and standard deviations) for ethnic and gender subgroups on each test form. Across grades, white students generally performed better than African American and Hispanic students. The scale score differences ranged from about 30 to 40 scale score points. Female students performed slightly better than male students across all grades.

Figures G1-G18 in Appendix G show histograms for the distribution of scale scores for the total population and for subgroups defined by gender and ethnicity.

Table 27
CRT Raw Score Descriptive Statistics

Grade Content	Form	N Count	Mean	Mean P-Value	SD	Min	Max	Alpha	SEM
MA3	1	36268	52.54	0.73	11.23	0	72	0.92	3.11
	2	24120	52.89	0.73	11.51	0	72	0.93	3.05
	Total	60388	52.68	0.73	11.35	0	72	.	
MA4	1	37011	45.35	0.65	13.68	0	70	0.94	3.41
	2	24774	44.53	0.63	13.93	0	71	0.94	3.48
	Total	61785	45.02	0.64	13.79	0	71	.	
MA5	1	38101	45.82	0.62	14.25	0	74	0.94	3.49
	2	25372	45.20	0.61	14.31	0	74	0.94	3.51
	Total	63473	45.58	0.62	14.28	0	74	.	
MA6	1	38922	39.18	0.56	15.28	0	70	0.95	3.53
	2	25828	39.50	0.56	14.67	0	69	0.94	3.53
	Total	64750	39.31	0.56	15.04	0	70	.	
MA7	1	39533	36.54	0.51	16.88	0	72	0.96	3.54
	2	26296	36.67	0.51	17.35	0	72	0.96	3.59
	Total	65829	36.59	0.51	17.07	0	72	.	
MA8	1	40707	35.07	0.47	16.89	0	75	0.95	3.73
	2	27033	34.02	0.45	17.24	0	75	0.95	3.71
	Total	67740	34.65	0.46	17.04	0	75	.	

Table 28
CRT Scale Score Descriptive Statistics

Grade Content	Form	N Count	Mean	SD	MIN	MAX
	1	36268	410.21	43.99	240	650
MA3	2	24120	412.33	43.07	240	650
	Total	60388	411.06	43.64	240	650
	1	37011	410.04	43.68	240	650
MA4	2	24774	411.10	43.33	240	650
	Total	61785	410.47	43.54	240	650
	1	38101	414.38	44.82	240	650
MA5	2	25372	415.71	45.61	240	650
	Total	63473	414.91	45.14	240	650
	1	38922	405.65	49.64	240	650
MA6	2	25828	407.19	46.43	240	553
	Total	64750	406.27	48.39	240	650
	1	39533	401.35	50.85	240	650
MA7	2	26296	403.02	51.00	240	650
	Total	65829	402.02	50.92	240	650
	1	40707	408.50	46.94	240	650
MA8	2	27033	407.51	48.92	240	650
	Total	67740	408.10	47.74	240	650

Table 29
CRT Scale Score Descriptive Statistics by Ethnicity

Grade Content	Test Form	White					African American					Hispanic				
		N	Mean	SD	Min	Max	N	Mean	SD	Min	Max	N	Mean	SD	Min	Max
MA3	1	17339	424.30	40.33	240	650	13613	391.94	41.84	240	568	3050	396.98	40.27	240	568
	2	11526	426.41	39.35	240	650	9088	394.34	40.63	240	650	2071	399.47	39.91	240	551
	Total	28865	425.14	39.96	240	650	22701	392.90	41.38	240	650	5121	397.99	40.14	240	568
MA4	1	18044	423.57	39.01	240	650	13770	391.63	42.05	240	554	3073	396.73	42.64	240	546
	2	11979	425.17	38.05	240	650	9279	392.44	41.81	240	541	2068	397.45	43.11	240	525
	Total	30023	424.21	38.64	240	650	23049	391.95	41.95	240	554	5141	397.02	42.83	240	546
MA5	1	18485	427.56	39.82	240	650	14391	396.10	43.46	240	540	3047	401.47	45.17	240	546
	2	12304	429.24	40.30	240	650	9755	396.51	44.41	240	553	1891	404.96	44.03	240	564
	Total	30789	428.23	40.02	240	650	24146	396.27	43.85	240	553	4938	402.81	44.77	240	564
MA6	1	18442	421.64	41.70	240	650	15379	384.68	50.55	240	528	2897	393.07	49.35	240	502
	2	12346	422.58	39.86	240	553	10212	387.16	45.83	240	519	1909	395.44	45.25	240	553
	Total	30788	422.02	40.97	240	650	25591	385.67	48.73	240	528	4806	394.01	47.77	240	553
MA7	1	19064	419.51	42.75	240	650	15597	377.83	49.78	240	530	2817	384.97	50.63	240	515
	2	12610	421.71	43.15	240	650	10421	378.62	49.16	240	650	1816	388.61	48.66	240	516
	Total	31674	420.39	42.92	240	650	26018	378.14	49.53	240	650	4633	386.39	49.89	240	516
MA8	1	19836	425.18	40.21	240	650	15996	386.31	44.70	240	556	2734	394.31	45.90	240	528
	2	13323	425.34	41.04	240	650	10501	382.32	47.04	240	519	1766	396.29	45.44	240	549
	Total	33159	425.25	40.54	240	650	26497	384.73	45.68	240	556	4500	395.09	45.73	240	549

Table 30
CRT Scale Score Descriptive Statistics by Gender

Grade Content	Test Form	Male					Female				
		N	Mean	SD	MIN	MAX	N	Mean	SD	MIN	MAX
MA3	1	18665	408.89	44.44	240	650	17600	411.62	43.47	240	650
	2	12353	412.15	42.83	240	650	11764	412.53	43.33	240	650
	Total	31018	410.19	43.83	240	650	29364	411.99	43.42	240	650
MA4	1	18953	409.18	45.27	240	650	18055	410.96	41.93	240	650
	2	12524	410.10	44.75	240	650	12247	412.14	41.77	240	650
	Total	31477	409.55	45.07	240	650	30302	411.44	41.87	240	650
MA5	1	19554	412.78	46.69	240	650	18543	416.09	42.66	240	577
	2	12922	414.80	47.75	240	650	12447	416.67	43.26	240	650
	Total	32476	413.59	47.12	240	650	30990	416.32	42.90	240	650
MA6	1	20249	403.32	52.25	240	650	18663	408.23	46.45	240	569
	2	13257	405.54	48.91	240	553	12565	408.98	43.51	240	553
	Total	33506	404.20	50.96	240	650	31228	408.53	45.29	240	569
MA7	1	20293	398.74	53.75	240	650	19233	404.11	47.45	240	555
	2	13473	399.81	53.85	240	650	12820	406.41	47.56	240	650
	Total	33766	399.17	53.79	240	650	32053	405.03	47.51	240	650
MA8	1	20939	406.14	50.06	240	650	19761	411.03	43.21	240	650
	2	13948	404.51	52.28	240	650	13080	410.70	44.85	240	650
	Total	34887	405.49	50.97	240	650	32841	410.90	43.87	240	650

The Relationship between NRT and CRT

Each of the 2006 MSA tests included both NRT and CRT items. Even though the specific content standards for the NRT and CRT assessments are somewhat different, the two tests are designed to measure similar knowledge, skills, and abilities. To examine how much these two tests measure the same performance, the correlation between scale scores on the NRT and scale scores on the CRT were produced and are presented in Table 31. The correlation was relatively high and similar across alternate forms within grade. The correlations ranged from 0.80 to 0.85 in Mathematics.

Table 31
Correlation between NRT and CRT

CRT Form	Content/Grade					
	MA3	MA4	MA5	MA6	MA7	MA8
Total	0.81	0.82	0.85	0.82	0.82	0.83
1	0.81	0.82	0.85	0.81	0.82	0.83
2	0.80	0.82	0.85	0.82	0.82	0.82

The Score Distributions and Correlations of Content Standards

Scale scores based on total test performance were reported to students, schools, and LEAs. Scale scores based on content standards were reported only to MSDE. These content-standard scale scores were estimated using a maximum-likelihood IRT pattern scoring procedure with item parameters estimated from performance on the total test form. Tables 32 and 33 show the raw score and scale score results for each content standard.

Tables 34 and 35 show the raw score Pearson product-moment and Spearman Rho correlations among the content standards at each grade level. Tables 36 and 37 show the scale score Pearson product-moment and Spearman Rho correlations among the content standards at each grade level. At every grade level, the Pearson raw score correlations are higher than the scale score correlations. This result is to be expected, given the differences between the raw score and scale score distributions.[3] Because of the properties of the scale score distributions, a nonparametric correlation procedure such as the Spearman Rho is more appropriate than the Pearson product-moment correlation. Indeed, when the Spearman Rho scale score correlations are compared with either the Pearson or Spearman Rho raw score correlations, the differences are negligible.

---

[3] Because a perfect raw score on any of the content standards is assigned the highest obtainable scale score on the total test, regardless of the difficulty or number of items included in the content standard, there tend to be very large gaps between the HOSS and the penultimate scale score. In addition, the scale score distributions differ substantially from one content standard to another. Given these distributions, a nonparametric correlation procedure such as the Spearman Rho seems more appropriate than the Pearson product-moment correlation.

Table 32
Distribution of Raw Scores on Content Standards

Grade	Form	Content Standard	N	Maximum Possible	Mean	SD	Minimum	Maximum
3	1	1	36268	13	10.09	2.39	0	13
		2&3	36268	14	11.76	2.18	0	14
		4&5	36268	14	10.76	2.72	0	14
		6	36268	16	12.68	2.81	0	16
		7	36268	14	6.61	2.82	0	14
	2	1	24120	13	10.15	2.31	0	13
		2&3	24120	14	11.44	2.38	0	14
		4&5	24120	14	11.26	2.68	0	14
		6	24120	16	12.67	2.79	0	16
		7	24120	14	6.71	2.98	0	14
4	1	1	37011	14	9.27	2.92	0	14
		2&3	37011	13	8.74	2.88	0	13
		4&5	37011	15	10.24	3.55	0	15
		6	37011	14	10.35	2.78	0	14
		7	37011	14	6.76	3.31	0	14
	2	1	24774	14	9.56	3.11	0	14
		2&3	24774	14	8.94	2.97	0	14
		4&5	24774	15	10.00	3.64	0	15
		6	24774	14	10.44	2.70	0	14
		7	24774	14	5.60	3.34	0	14
5	1	1	38101	15	10.80	3.23	0	15
		2&3	38101	14	8.75	2.88	0	14
		4&5	38101	13	9.16	2.74	0	13
		6	38101	15	9.95	3.59	0	15
		7	38101	17	7.16	3.60	0	17
	2	1	25372	15	10.72	3.21	0	15
		2&3	25372	14	8.23	3.10	0	14
		4&5	25372	13	8.97	2.84	0	13
		6	25372	15	9.88	3.45	0	15
		7	25372	17	7.40	3.47	0	17

Table 32 (cont.)
Distribution of Raw Scores on Content Standards

Grade	Form	Content Standard	N	Maximum Possible	Mean	SD	Minimum	Maximum
6	1	1	38922	14	9.11	3.44	0	14
		2&3	38922	14	7.52	3.41	0	14
		4&5	38922	13	7.71	3.04	0	13
		6	38922	14	8.58	3.55	0	14
		7	38922	15	6.26	3.53	0	15
	2	1	25828	14	8.92	3.20	0	14
		2&3	25828	14	7.84	2.91	0	14
		4&5	25828	13	7.94	3.02	0	13
		6	25828	14	8.11	3.58	0	14
		7	25828	15	6.69	3.67	0	15
7	1	1	39533	14	7.51	3.81	0	14
		2&3	39533	13	5.57	3.69	0	13
		4&5	39533	14	7.97	3.63	0	14
		6	39533	14	7.51	3.54	0	14
		7	39533	17	7.98	3.83	0	17
	2	1	26296	14	7.51	4.02	0	14
		2&3	26296	13	6.43	3.57	0	13
		4&5	26296	14	7.68	3.63	0	14
		6	26296	14	7.82	3.54	0	14
		7	26296	17	7.24	4.14	0	17
8	1	1	40707	15	7.69	3.75	0	15
		2&3	40707	13	6.13	3.11	0	13
		4&5	40707	14	7.08	3.36	0	14
		6	40707	14	6.39	3.47	0	14
		7	40707	19	7.79	4.89	0	19
	2	1	27033	15	7.27	3.88	0	15
		2&3	27033	13	6.41	3.24	0	13
		4&5	27033	14	7.20	3.55	0	14
		6	27033	14	6.58	3.55	0	14
		7	27033	19	6.57	4.74	0	19

Table 33
Distribution of Scale Scores on Content Standards

Grade	Form	Content Standard	N	Maximum Possible	Mean	SD	Minimum	Maximum
3	1	1	36268	650	436.09	91.80	240	650
		2&3	36268	650	454.12	112.77	240	650
		4&5	36268	650	436.04	92.97	240	650
		6	36268	650	434.71	91.98	240	650
		7	36268	650	396.27	55.90	240	650
	2	1	24120	650	437.16	91.26	240	650
		2&3	24120	650	452.40	108.08	240	650
		4&5	24120	650	450.77	104.54	240	650
		6	24120	650	436.09	91.71	240	650
		7	24120	650	403.38	51.21	240	650
4	1	1	37011	650	417.74	66.61	240	650
		2&3	37011	650	423.03	81.26	240	650
		4&5	37011	650	424.47	79.53	240	650
		6	37011	650	432.52	89.49	240	650
		7	37011	650	402.66	53.68	240	650
	2	1	24774	650	425.12	80.20	240	650
		2&3	24774	650	418.21	71.66	240	650
		4&5	24774	650	426.38	81.20	240	650
		6	24774	650	433.23	89.86	240	650
		7	24774	650	399.27	59.22	240	650
5	1	1	38101	650	434.11	85.77	240	650
		2&3	38101	650	420.18	68.16	240	650
		4&5	38101	650	431.80	82.64	240	650
		6	38101	650	428.75	81.61	240	650
		7	38101	650	402.95	53.64	240	650
	2	1	25372	650	432.46	85.32	240	650
		2&3	25372	650	420.58	66.88	240	650
		4&5	25372	650	430.85	82.38	240	650
		6	25372	650	426.40	76.24	240	650
		7	25372	650	408.28	51.27	240	650

Table 33 (cont.)
Distribution of Scale Scores on Content Standards

Grade	Form	Content Standard	N	Maximum Possible	Mean	SD	Minimum	Maximum
6	1	1	38922	650	419.28	80.99	240	650
		2&3	38922	650	406.77	74.68	240	650
		4&5	38922	650	411.92	74.69	240	650
		6	38922	650	414.40	82.20	240	650
		7	38922	650	398.17	57.22	240	650
	2	1	25828	650	413.73	69.67	240	650
		2&3	25828	650	410.18	63.42	240	650
		4&5	25828	650	414.77	74.69	240	650
		6	25828	650	411.08	83.29	240	650
		7	25828	650	400.87	54.43	240	650
7	1	1	39533	650	402.40	84.61	240	650
		2&3	39533	650	392.27	87.41	240	650
		4&5	39533	650	405.98	75.78	240	650
		6	39533	650	407.71	73.08	240	650
		7	39533	650	394.70	52.44	240	650
	2	1	26296	650	403.31	90.88	240	650
		2&3	26296	650	404.94	78.95	240	650
		4&5	26296	650	406.25	74.03	240	650
		6	26296	650	413.64	78.40	240	650
		7	26296	650	396.45	56.78	240	650
8	1	1	40707	650	411.93	68.48	240	650
		2&3	40707	650	408.58	62.61	240	650
		4&5	40707	650	408.28	62.34	240	650
		6	40707	650	398.92	83.58	240	650
		7	40707	650	402.96	54.60	240	650
	2	1	27033	650	410.01	71.11	240	650
		2&3	27033	650	412.24	65.57	240	650
		4&5	27033	650	410.77	69.93	240	650
		6	27033	650	400.68	85.96	240	650
		7	27033	650	396.66	60.05	240	650

Table 34
Raw Score Correlations (Pearson Product-Moment) between Content Standards

Grade	Content Standard	Mean	SD	1	2&3	4&5	6	7
				Mathematics				
3	1	10.12	2.36	1.00	0.69	0.73	0.74	0.63
	2&3	11.63	2.27		1.00	0.70	0.72	0.63
	4&5	10.96	2.72			1.00	0.76	0.66
	6	12.68	2.80				1.00	0.68
	7	6.65	2.88					1.00
4	1	9.38	3.00	1.00	0.72	0.74	0.72	0.70
	2&3	8.82	2.92		1.00	0.73	0.70	0.71
	4&5	10.14	3.59			1.00	0.71	0.76
	6	10.38	2.75				1.00	0.69
	7	6.29	3.37					1.00
5	1	10.77	3.22	1.00	0.69	0.74	0.76	0.74
	2&3	8.54	2.98		1.00	0.72	0.72	0.71
	4&5	9.09	2.78			1.00	0.75	0.74
	6	9.92	3.54				1.00	0.77
	7	7.26	3.55					1.00
6	1	9.04	3.35	1.00	0.73	0.75	0.78	0.80
	2&3	7.65	3.23		1.00	0.70	0.72	0.77
	4&5	7.80	3.03			1.00	0.74	0.75
	6	8.39	3.57				1.00	0.79
	7	6.43	3.59					1.00
7	1	7.51	3.89	1.00	0.78	0.80	0.82	0.81
	2&3	5.91	3.67		1.00	0.76	0.78	0.77
	4&5	7.86	3.63			1.00	0.78	0.83
	6	7.63	3.55				1.00	0.76
	7	7.68	3.97					1.00
8	1	7.52	3.81	1.00	0.77	0.77	0.77	0.85
	2&3	6.24	3.17		1.00	0.74	0.73	0.79
	4&5	7.12	3.44			1.00	0.74	0.80
	6	6.47	3.51				1.00	0.76
	7	7.30	4.87					1.00

Table 35
Raw Score Correlations (Spearman Rho) between Content Standards

Grade	Content Standard	Mean	SD	1	2&3	4&5	6	7
					Mathematics			
3	1	10.12	2.36	1.00	0.64	0.68	0.70	0.62
	2&3	11.63	2.27		1.00	0.64	0.67	0.62
	4&5	10.96	2.72			1.00	0.70	0.64
	6	12.68	2.80				1.00	0.67
	7	6.65	2.88					1.00
4	1	9.38	3.00	1.00	0.72	0.73	0.71	0.70
	2&3	8.82	2.92		1.00	0.73	0.70	0.72
	4&5	10.14	3.59			1.00	0.70	0.76
	6	10.38	2.75				1.00	0.70
	7	6.29	3.37					1.00
5	1	10.77	3.22	1.00	0.70	0.73	0.76	0.75
	2&3	8.54	2.98		1.00	0.72	0.72	0.71
	4&5	9.09	2.78			1.00	0.75	0.75
	6	9.92	3.54				1.00	0.78
	7	7.26	3.55					1.00
6	1	9.04	3.35	1.00	0.73	0.75	0.78	0.80
	2&3	7.65	3.23		1.00	0.69	0.73	0.77
	4&5	7.80	3.03			1.00	0.74	0.75
	6	8.39	3.57				1.00	0.79
	7	6.43	3.59					1.00
7	1	7.51	3.89	1.00	0.77	0.81	0.82	0.82
	2&3	5.91	3.67		1.00	0.77	0.78	0.78
	4&5	7.86	3.63			1.00	0.79	0.83
	6	7.63	3.55				1.00	0.77
	7	7.68	3.97					1.00
8	1	7.52	3.81	1.00	0.75	0.77	0.75	0.84
	2&3	6.24	3.17		1.00	0.74	0.71	0.78
	4&5	7.12	3.44			1.00	0.73	0.80
	6	6.47	3.51				1.00	0.73
	7	7.30	4.87					1.00

Table 36
Scale Score Correlations (Pearson Product-Moment) between Content Standards

Grade	Content Standard	Mean	SD	1	2&3	4&5	6	7
					Mathematics			
3	1	436.51	91.58	1.00	0.51	0.51	0.54	0.54
	2&3	453.43	110.92		1.00	0.49	0.52	0.54
	4&5	441.93	98.02			1.00	0.51	0.55
	6	435.26	91.87				1.00	0.57
	7	399.11	54.19					1.00
4	1	420.70	72.46	1.00	0.57	0.57	0.56	0.62
	2&3	421.10	77.59		1.00	0.57	0.56	0.63
	4&5	425.24	80.21			1.00	0.54	0.63
	6	432.81	89.64				1.00	0.59
	7	401.30	55.99					1.00
5	1	433.45	85.59	1.00	0.60	0.59	0.61	0.66
	2&3	420.34	67.65		1.00	0.60	0.61	0.67
	4&5	431.42	82.53			1.00	0.60	0.66
	6	427.81	79.51				1.00	0.68
	7	405.08	52.77					1.00
6	1	417.07	76.72	1.00	0.61	0.62	0.62	0.69
	2&3	408.13	70.42		1.00	0.60	0.60	0.69
	4&5	413.06	74.70			1.00	0.61	0.67
	6	413.08	82.65				1.00	0.66
	7	399.25	56.14					1.00
7	1	402.76	87.17	1.00	0.64	0.68	0.68	0.73
	2&3	397.33	84.36		1.00	0.65	0.66	0.70
	4&5	406.09	75.09			1.00	0.68	0.75
	6	410.08	75.31				1.00	0.69
	7	395.40	54.22					1.00
8	1	411.17	69.55	1.00	0.68	0.69	0.61	0.74
	2&3	410.04	63.83		1.00	0.67	0.59	0.71
	4&5	409.27	65.48			1.00	0.60	0.74
	6	399.62	84.54				1.00	0.60
	7	400.44	56.92					1.00

Table 37
Scale Score Correlations (Spearman Rho) between Content Standards

Grade	Content Standard	Mean	SD	1	2&3	4&5	6	7
				Mathematics				
3	1	436.51	91.58	1.00	0.66	0.70	0.71	0.64
	2&3	453.43	110.92		1.00	0.66	0.68	0.63
	4&5	441.93	98.02			1.00	0.72	0.66
	6	435.26	91.87				1.00	0.68
	7	399.11	54.19					1.00
4	1	420.70	72.46	1.00	0.73	0.74	0.73	0.75
	2&3	421.10	77.59		1.00	0.74	0.71	0.75
	4&5	425.24	80.21			1.00	0.71	0.78
	6	432.81	89.64				1.00	0.73
	7	401.30	55.99					1.00
5	1	433.45	85.59	1.00	0.72	0.74	0.77	0.76
	2&3	420.34	67.65		1.00	0.73	0.75	0.74
	4&5	431.42	82.53			1.00	0.76	0.76
	6	427.81	79.51				1.00	0.80
	7	405.08	52.77					1.00
6	1	417.07	76.72	1.00	0.75	0.77	0.80	0.82
	2&3	408.13	70.42		1.00	0.72	0.75	0.79
	4&5	413.06	74.70			1.00	0.77	0.78
	6	413.08	82.65				1.00	0.82
	7	399.25	56.14					1.00
7	1	402.76	87.17	1.00	0.80	0.83	0.84	0.84
	2&3	397.33	84.36		1.00	0.79	0.80	0.81
	4&5	406.09	75.09			1.00	0.81	0.84
	6	410.08	75.31				1.00	0.79
	7	395.40	54.22					1.00
8	1	411.17	69.55	1.00	0.79	0.80	0.77	0.86
	2&3	410.04	63.83		1.00	0.77	0.73	0.82
	4&5	409.27	65.48			1.00	0.75	0.83
	6	399.62	84.54				1.00	0.75
	7	400.44	56.92					1.00

Factor analysis of the MSA Assessments

Exploratory factor analysis was used to examine the structure of the 2006 MSA assessments. At each grade, principal axis factor analysis was applied to extract factor(s) from each of the two operational forms (Form 1 and Form 2), with varimax rotation of the extracted factors. For each test, the number of factors extracted was equal to the number of reported content standards (i.e., 5 factors for each of the Mathematics assessments). Squared multiple correlations (SMC) were used as prior communality estimates (Harman, 1976). The results of these analyses are shown in Appendix H, Tables H1 to H24.

Each test form had between 9 and 16 initial eigenvalues greater than 1.0, with one dominant factor accounting for approximately 17 to 27 percent of the variance, with each additional factor accounting for less than 4 percent of the total variance. After extraction and rotation of 5 factors for each of the Mathematics tests, the variance explained by the factors ranged from 7.6 to 12.1 percent for the first factor, 4.9 to 10.4 percent for the second factor, 1.9 to 6.2 percent for the third factor, 1.3 to 5.1 percent for the fourth factor, and 1.1 to 3.0 for the fifth factor.

While these analyses did yield multifactorial solutions for all of the tests, there was generally no clear relationship between the content standards and the loadings on the extracted factors.

Percent At or Above Cut (PAC)

At the Bookmark standard-setting workshops in 2003 and 2004, performance level cut scores were established for three proficiency levels: Basic, Proficient, and Advanced. Table 38 shows the resulting scale score ranges for each performance level. Note that the Maryland scale was not constructed as a vertical scale, so meaningful comparisons can not be made between performance cut scores at different grades.

Table 39 shows the percentages of students at each performance level on the 2006 MSA assessments. The last column "Proficient + Advanced" represents the percent at or above the cut (PAC) that will be reported for the NCLB act. The 2006 PAC for Mathematics showed a steady decline from grade 4 to grade, 8 dropping from approximately 82 percent in Grade 4 to approximately 55 percent in Grade 8. Tables 40 and 41 show the PAC classified by ethnicity and gender group. Tables 42 to 47 present the PAC by local education agencies (LEA) for each grade. Figures 2 to 7 show changes in the PAC between 2004 and 2005 for each LEA.

Table 38
Scale Score Ranges for Each Performance Level
Based on 2003 and 2004 Standard Setting

Grade	Basic	Proficient	Advanced
3	240-378	379-440	441-650
4	240-373	374-432	433-650
5	240-391	392-452	453-650
6	240-395	396-446	447-650
7	240-395	396-450	451-650
8	240-406	407-443	444-650

Table 39
Percentages of Students at Each Performance Level

Grade Content	Form	N	Basic	Proficient	Advanced	Proficient +Advanced
MA3	1	36268	21.48	54.33	24.19	78.52
	2	24120	20.53	53.94	25.52	79.47
	Total	60388	21.10	54.17	24.72	78.90
MA4	1	37011	18.37	49.87	31.76	81.63
	2	24774	17.87	49.68	32.45	82.13
	Total	61785	18.17	49.79	32.04	81.83
MA5	1	38101	27.14	54.12	18.74	72.86
	2	25372	26.40	53.83	19.77	73.60
	Total	63473	26.84	54.00	19.15	73.16
MA6	1	38922	34.28	47.28	18.44	65.72
	2	25828	35.02	46.00	18.98	64.98
	Total	64750	34.57	46.77	18.66	65.43
MA7	1	39533	40.29	44.35	15.36	59.71
	2	26296	39.97	43.50	16.53	60.03
	Total	65829	40.16	44.01	15.83	59.84
MA8	1	40707	45.20	32.44	22.36	54.80
	2	27033	44.90	32.48	22.62	55.10
	Total	67740	45.08	32.46	22.46	54.92

Table 40

Percentages of Students at Each Performance Level by Ethnicity

Grade Content	Ethnicity	N	Basic	Proficient	Advanced	Proficient +Advanced
MA3	White	28865	11.05	53.87	35.07	88.95
	African American	22701	34.02	54.73	11.25	65.98
	Hispanic	5121	28.88	57.88	13.24	71.12
	Others	3701	9.46	47.99	42.56	90.54
MA4	White	30023	9.04	46.88	44.07	90.96
	African American	23049	30.14	54.22	15.64	69.86
	Hispanic	5141	25.60	55.24	19.16	74.40
	Others	3572	6.94	37.82	55.24	93.06
MA5	White	30789	16.28	56.61	27.11	83.72
	African American	24146	41.07	51.56	7.37	58.93
	Hispanic	4938	35.30	54.46	10.25	64.70
	Others	3600	10.22	47.44	42.33	89.78
MA6	White	30788	20.69	51.56	27.76	79.31
	African American	25591	52.47	41.23	6.30	47.53
	Hispanic	4806	43.32	48.00	8.68	56.68
	Others	3565	14.22	43.56	42.22	85.78
MA7	White	31674	24.04	51.73	24.22	75.96
	African American	26018	60.80	34.81	4.39	39.20
	Hispanic	4633	52.15	41.87	5.98	47.85
	Others	3504	16.81	45.35	37.84	83.19
MA8	White	33159	27.76	38.76	33.48	72.24
	African American	26497	68.14	25.05	6.82	31.86
	Hispanic	4500	56.82	31.44	11.73	43.18
	Others	3584	20.12	30.19	49.69	79.88

Table 41
Percentages of Students at Each Performance Level by Gender

Grade Content	Gender	N	Basic	Proficient	Advanced	Proficient +Advanced
MA3	Male	31018	21.45	54.38	24.17	78.55
	Female	29364	20.74	53.95	25.31	79.26
MA4	Male	31477	19.37	48.28	32.35	80.63
	Female	30302	16.92	51.36	31.72	83.08
MA5	Male	32476	28.02	52.48	19.51	71.98
	Female	30990	25.60	55.61	18.79	74.40
MA6	Male	33506	36.70	44.04	19.25	63.30
	Female	31228	32.26	49.72	18.03	67.74
MA7	Male	33766	42.36	41.83	15.81	57.64
	Female	32053	37.84	46.31	15.85	62.16
MA8	Male	34887	46.73	30.61	22.66	53.27
	Female	32841	43.31	34.43	22.26	56.69

Table 42
Percentages of Students at Grade 3 Performance Levels by LEA

LEA #	N	Basic	Proficient	Advanced	Proficient +Advanced
1	682	22.43	51.91	25.66	77.57
2	5241	11.85	53.60	34.55	88.15
3	7417	22.56	53.89	23.55	77.44
4	1208	9.02	47.68	43.29	90.98
5	421	19.00	60.81	20.19	81.00
6	1969	12.04	60.18	27.78	87.96
7	1153	19.51	63.92	16.57	80.49
8	1871	22.02	56.01	21.97	77.98
9	332	37.95	51.20	10.84	62.05
10	2879	18.27	60.40	21.33	81.73
11	297	14.14	68.35	17.51	85.86
12	2931	14.71	60.35	24.94	85.30
13	3577	12.30	51.19	36.51	87.70
14	168	8.33	54.17	37.50	91.67
15	9644	16.05	48.51	35.44	83.95
16	9171	30.96	56.18	12.87	69.04
17	515	13.01	61.75	25.24	86.99
18	1147	14.91	54.49	30.60	85.09
19	181	25.41	61.88	12.71	74.59
20	303	16.50	54.79	28.71	83.50
21	1573	14.62	57.41	27.97	85.38
22	1125	19.02	56.09	24.89	80.98
23	449	8.91	44.54	46.55	91.09
30	5818	39.81	51.55	8.65	60.19
31	270	45.56	49.63	4.81	54.44
55	46	23.91	69.57	6.52	76.09

Table 43
Percentages of Students at Grade 4 Performance Levels by LEA

LEA #	N	Basic	Proficient	Advanced	Proficient +Advanced
1	666	17.57	46.70	35.74	82.43
2	5358	9.26	45.61	45.13	90.74
3	7636	15.89	51.91	32.20	84.11
4	1270	8.11	42.36	49.53	91.89
5	354	14.69	54.52	30.79	85.31
6	2086	10.16	55.94	33.89	89.84
7	1171	22.80	55.76	21.43	77.20
8	1840	19.35	52.07	28.59	80.65
9	306	30.07	52.29	17.65	69.93
10	2974	14.53	51.61	33.86	85.47
11	339	12.68	58.11	29.20	87.32
12	2965	13.32	54.74	31.94	86.68
13	3679	10.52	43.08	46.40	89.48
14	148	10.14	48.65	41.22	89.86
15	10008	13.58	43.90	42.52	86.42
16	9521	28.35	54.13	17.52	71.65
17	577	15.25	51.13	33.62	84.75
18	1173	13.30	49.87	36.83	86.70
19	213	13.62	68.08	18.31	86.39
20	300	19.00	42.67	38.33	81.00
21	1574	10.42	51.65	37.93	89.58
22	1057	13.91	49.20	36.90	86.09
23	443	14.00	43.57	42.44	86.00
30	5809	37.51	51.20	11.29	62.49
31	282	34.40	51.42	14.18	65.60
55	35	22.86	51.43	25.71	77.14

Table 44
Percentages of Students at Grade 5 Performance Levels by LEA

LEA #	N	Basic	Proficient	Advanced	Proficient +Advanced
1	659	29.59	52.96	17.45	70.41
2	5496	16.94	56.60	26.46	83.06
3	7917	27.80	54.96	17.24	72.20
4	1301	12.99	58.19	28.82	87.01
5	390	25.90	61.28	12.82	74.10
6	2114	15.42	62.25	22.33	84.58
7	1222	23.00	63.34	13.67	77.00
8	1923	25.53	55.23	19.24	74.47
9	308	39.61	52.60	7.79	60.39
10	3047	22.68	57.24	20.09	77.32
11	367	29.97	57.49	12.53	70.03
12	3053	22.21	61.42	16.38	77.79
13	3901	13.00	51.50	35.50	87.00
14	158	30.38	56.33	13.29	69.62
15	10182	19.36	51.11	29.53	80.64
16	9786	40.82	50.50	8.68	59.18
17	538	15.99	63.01	21.00	84.01
18	1202	23.79	54.83	21.38	76.21
19	179	27.93	62.57	9.50	72.07
20	314	19.43	59.24	21.34	80.57
21	1514	25.30	57.27	17.44	74.70
22	1075	26.98	54.70	18.33	73.02
23	450	20.44	59.33	20.22	79.56
30	6032	46.52	48.13	5.35	53.48
31	304	50.33	45.39	4.28	49.67
55	40	35.00	57.50	7.50	65.00

Table 45
Percentages of Students at Grade 6 Performance Levels by LEA

LEA #	N	Basic	Proficient	Advanced	Proficient +Advanced
1	669	32.59	46.94	20.48	67.41
2	5468	27.82	48.96	23.23	72.18
3	7832	36.43	47.70	15.87	63.57
4	1345	25.80	51.60	22.60	74.20
5	399	29.57	53.13	17.29	70.43
6	2238	20.69	54.65	24.66	79.31
7	1289	32.74	50.58	16.68	67.26
8	2011	31.68	52.11	16.21	68.32
9	350	55.14	39.14	5.71	44.86
10	2988	21.75	54.45	23.80	78.25
11	365	29.59	55.07	15.34	70.41
12	3081	30.61	50.73	18.66	69.39
13	3774	16.72	49.63	33.65	83.28
14	178	45.51	49.44	5.06	54.49
15	10015	23.96	46.99	29.05	76.04
16	10480	45.13	46.82	8.04	54.87
17	578	23.70	51.73	24.57	76.30
18	1293	26.99	47.33	25.68	73.01
19	255	42.35	44.31	13.33	57.65
20	318	34.28	51.89	13.84	65.72
21	1597	19.66	53.48	26.86	80.34
22	1022	37.48	43.25	19.28	62.52
23	497	18.91	49.50	31.59	81.09
30	6393	68.79	27.73	3.47	31.21
31	274	58.39	37.96	3.65	41.61
55	41	46.34	43.90	9.76	53.66

Table 46
Percentages of Students at Grade 7 Performance Levels by LEA

LEA #	N	Basic	Proficient	Advanced	Proficient +Advanced
1	744	31.05	54.70	14.25	68.95
2	5565	30.58	44.65	24.76	69.42
3	8149	42.25	43.40	14.35	57.75
4	1384	29.12	54.55	16.33	70.88
5	403	36.97	53.10	9.93	63.03
6	2299	28.93	51.98	19.10	71.07
7	1348	37.54	51.41	11.05	62.46
8	2111	39.32	49.64	11.04	60.68
9	337	58.75	37.98	3.26	41.25
10	3048	26.71	51.44	21.85	73.29
11	419	27.45	62.53	10.02	72.55
12	3014	35.63	49.87	14.50	64.37
13	3959	19.30	49.84	30.87	80.70
14	176	50.00	41.48	8.52	50.00
15	10286	29.36	46.44	24.20	70.64
16	10376	54.68	39.02	6.29	45.32
17	598	23.58	59.36	17.06	76.42
18	1208	34.93	47.27	17.80	65.07
19	236	51.27	39.41	9.32	48.73
20	361	38.78	45.98	15.24	61.22
21	1586	23.14	55.42	21.44	76.86
22	1089	42.42	43.99	13.59	57.58
23	492	21.14	54.47	24.39	78.86
30	6596	75.46	22.42	2.12	24.55
55	44	61.36	36.36	2.27	38.64

Table 47
Percentages of Students at Grade 8 Performance Levels by LEA

LEA #	N	Basic	Proficient	Advanced	Proficient +Advanced
1	751	37.82	42.21	19.97	62.18
2	5790	31.00	36.86	32.14	69.00
3	8481	43.26	35.34	21.40	56.74
4	1398	37.27	38.98	23.75	62.73
5	440	41.14	37.27	21.59	58.86
6	2341	37.98	38.19	23.84	62.02
7	1332	38.89	41.59	19.52	61.11
8	2105	45.08	36.34	18.57	54.92
9	360	65.56	26.94	7.50	34.44
10	3154	29.14	37.86	33.01	70.86
11	357	28.57	44.82	26.61	71.43
12	3181	38.60	36.50	24.90	61.40
13	3935	23.63	38.55	37.81	76.37
14	187	56.15	30.48	13.37	43.85
15	10618	33.58	32.04	34.39	66.42
16	10791	66.29	24.83	8.89	33.71
17	611	30.93	42.88	26.19	69.07
18	1261	46.79	35.61	17.61	53.21
19	258	56.59	32.95	10.47	43.41
20	349	49.00	35.53	15.47	51.00
21	1641	25.47	38.57	35.95	74.53
22	1087	50.78	36.43	12.79	49.22
23	547	21.94	35.28	42.78	78.06
30	6717	78.53	17.79	3.68	21.47
55	43	60.47	30.23	9.30	39.53

Figure 2

Percent at or Above Proficiency Cut Score (PAC) by LEA for Mathematics Grade 3
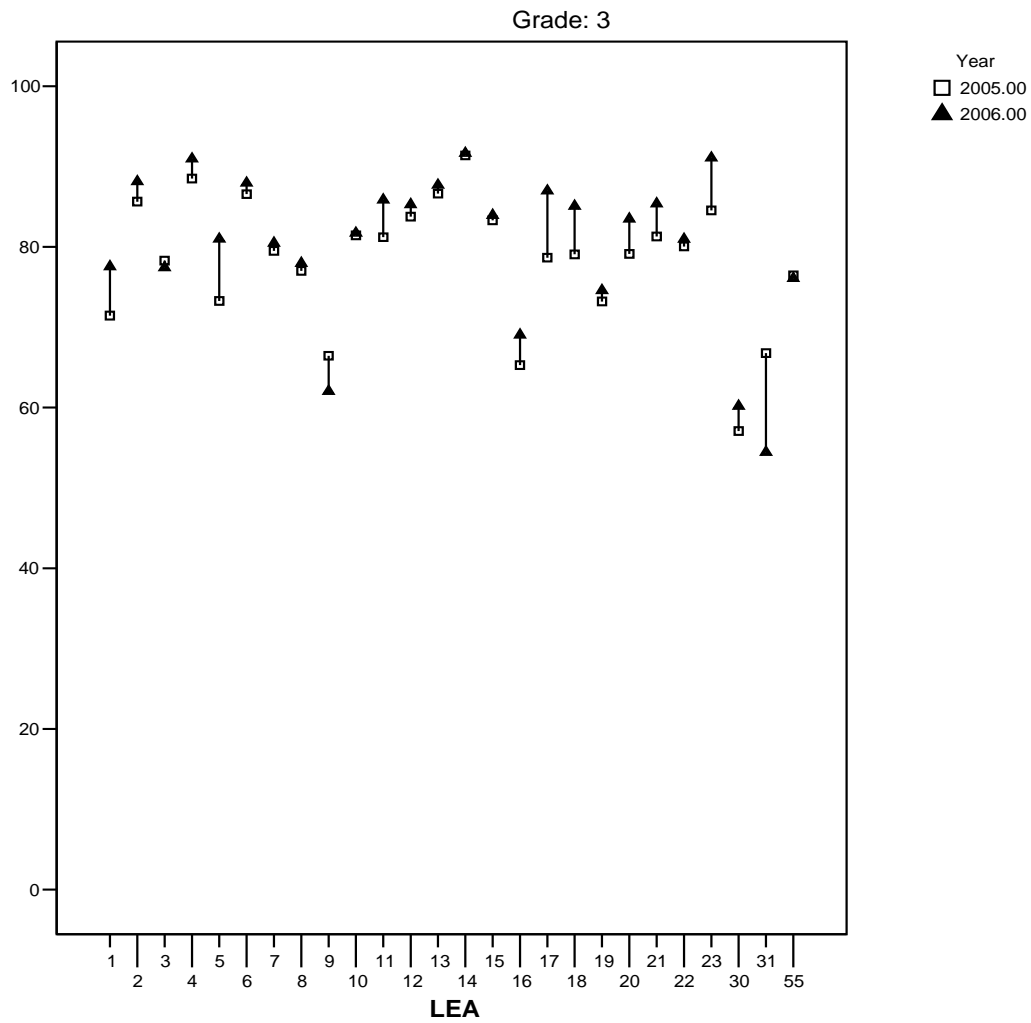
Figure 3

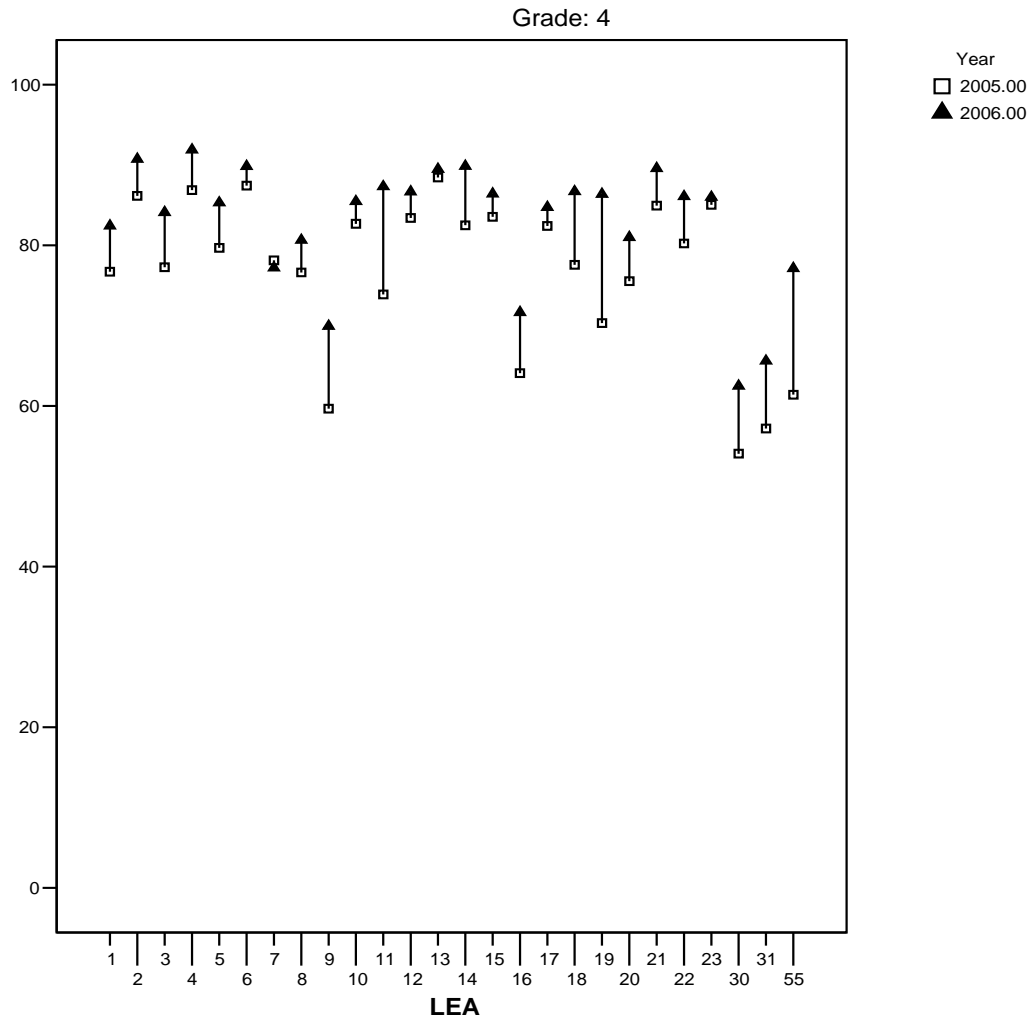Percent at or Above Proficiency Cut Score (PAC) by LEA for Mathematics Grade 4

Figure 4

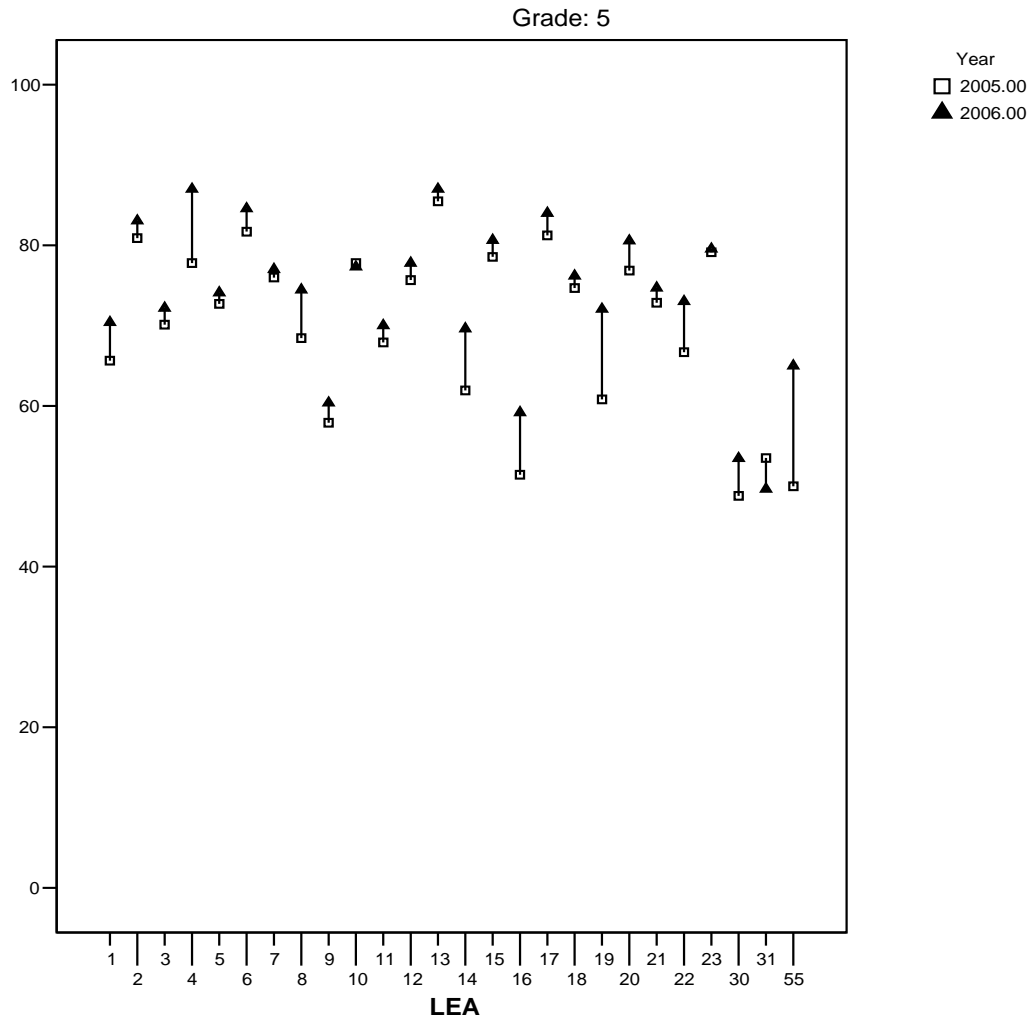Percent at or Above Proficiency Cut Score (PAC) by LEA for Mathematics Grade 5

Figure 5

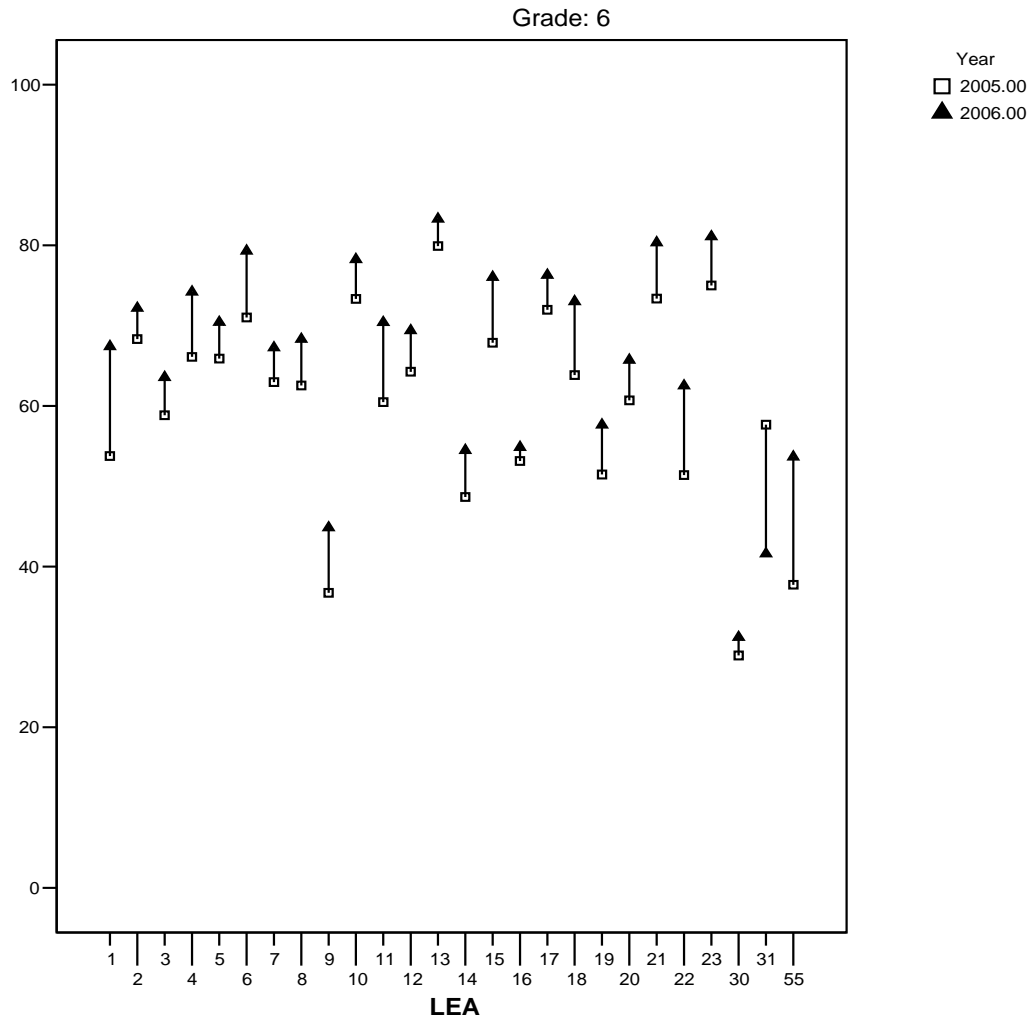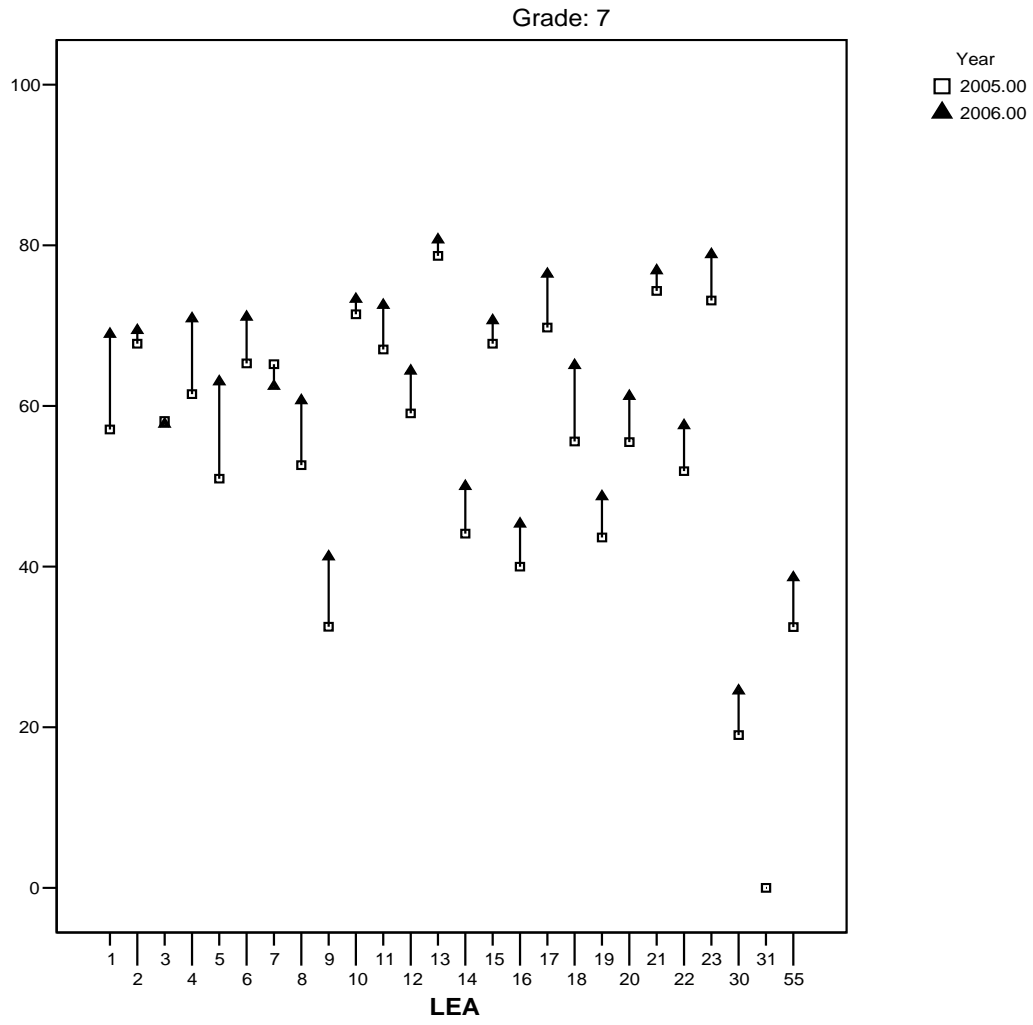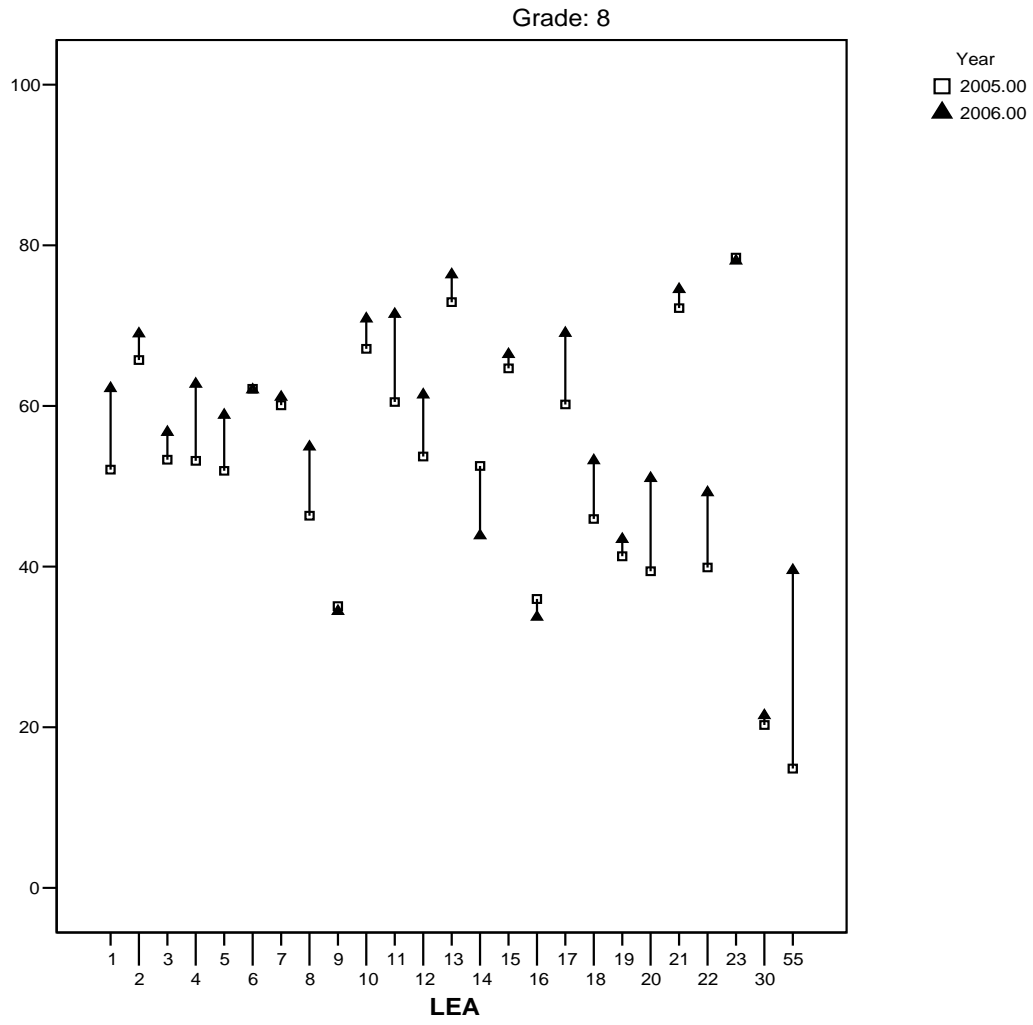Percent at or Above Proficiency Cut Score (PAC) by LEA for Mathematics Grade 6

Figure 6
Percent at or Above Proficiency Cut Score (PAC) by LEA for Mathematics Grade 7

Figure 7
Percent at or Above Proficiency Cut Score (PAC) by LEA for Mathematics Grade 8

References

Bock, R. D., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: An application of an EM algorithm. *Psychometrika, 46,* 443–459.

Burket, G. R. (1991). PARDUX [Computer program]. Unpublished.

Harman, H. H. (1976). Modern Factor analysis (3rd edition). The University of Chicago Press: Chicago.

Kolen, M.J., & Brennan, R.L. (1995). *Test equating: Methods and practices*. New York: Springer-Verlag.

Muraki, E. (1990). Fitting a polytomous item response model to Likert-type data. *Applied Psychological Measurement, 14,* 59–71.

Muraki, E. (1992). A generalized partial credit model: Application of an EM algorithm. *Applied Psychological Measurement, 16,* 159–176.

Thissen, D. (1982). Marginal maximum-likelihood estimation for the one-parameter logistic model. *Psychometrika, 47,* 175–186.

Yen, W. M. (1984). Effects of local item dependence on the fit and equating performance of the three-parameter logistic model. *Applied Psychological Measurement, 2*, 125–145.

Yen, W. M. (1991a). *How to evaluate FIT for Maryland*. Unpublished manuscript.

Yen, W. M. (1991b). *Specifications for CAT/5 Benchmark Scaling: Setting LOSS/HOSS values*. Unpublished manuscript.

Zwick, R., Donoghue, J.R., & Grima, A. (1993). Assessment of differential item functioning for performance tasks. *Journal of Educational Measurement, 30*, 233–251.