Maryland School Assessment
2005

Technical Report 2005

for

Maryland School Assessment

Submitted to Maryland State Department of Education

September 05

A subsidiary of The McGraw-Hill Companies

CTB/McGraw-Hill

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. This technical report addresses only those assessments for which CTB/McGraw-Hill was responsible.

The MSA Mathematics and Grade 10 Reading 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, 8 and 10; *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.

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 and Reading 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 and Reading Grade 10. 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 Reading, each reporting category corresponds to a single content standard. For Mathematics, however, 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	0-1	7-8	10%-11%
3	Measurement	24	3	1		0-1	5-6	7%-9%
4	Statistics		5	1	1	1	8	12%
5	Probability		3	2			5	7%
6	Number Relationships and Computation	3, 9, 13, 15, 32	6	3			14	20%
7	Process of Mathematics				5	3	17	24%
	Total Score Points	6	30	12	13	12	71	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	0-1	0-1	8	11%
3	Measurement	16	2	1	0-1	0-1	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 1 (cont.)
Test Designs by Grade / Content

Grade 10 Reading

	Content Standard	TerraNova Items that Contribute to CRT Score	Number of CRT SR Items	Number of CRT CR Items	Score Points	Percentage of Score Points
G	General Reading Processes	3, 4, 23, 25, 32, 33, 34, 37, 43, 44, 45, 47, 48	3		16	26%
I	Informational Reading Processes	13, 14, 15, 16, 17, 22, 24, 26, 27, 46	6	2	22	36%
L	Literary Reading Processes	1, 2, 5, 6, 35, 36, 38, 51, 52, 53, 54	6	2	23	38%
	Total Score Points	**34**	**15**	**12**	**61**	**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	17.6%
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	19.7%	15	20.0%
2&3	14	20.0%	13	18.3%	13	17.3%
4&5	13	18.5%	13	18.3%	14	18.7%
6	14	20.0%	14	19.7%	14	18.7%
7	15	21.5%	17	23.9%	19	25.3%
Total	70	100%	71	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.

Handscoring Process

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 2005 MSA. Because percentages are rounded up to whole numbers, the percentages in this table do not always sum to 100. Among the Mathematics examinees, 48 to 51 percent were White, 37 to 40 percent were African American, and 6 to 8 percent were Hispanic. Among the Reading examinees, 76 percent were White, 20 percent were African American, and 2 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 2005 distributions of ethnicity and gender for the Mathematics tests are essentially the same as the 2003 and 2004 distributions. The ethnic composition of the 2005 Reading examinee population was substantially different from previous years' distributions. However, it should be noted that the 2005 Reading assessment was administered to only a small group of students (6,934 students in 2005 vs. 62,732 in 2004), who were not expected to be comparable to previous years' examinees.

Table 3
2005 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	12657	49	37	8	5
	B	12442	49	37	8	6
	C	12258	50	37	8	6
	D	12132	49	38	8	6
	E	12020	49	38	7	6
	Total	61509	49	37	8	6
4	A	12965	50	38	7	6
	B	12796	49	38	8	5
	C	12638	49	38	7	6
	D	12530	49	38	7	5
	E	12401	49	38	7	6
	Total	63330	49	38	7	6
5	A	13330	49	39	7	5
	B	13153	49	38	7	5
	C	13043	49	39	7	5
	D	12861	49	39	7	5
	E	12736	48	39	7	6
	Total	65123	49	39	7	5
6	A	13388	48	40	7	5
	B	13247	48	40	7	5
	C	13090	49	40	7	5
	D	13090	49	39	7	5
	E	13031	49	39	7	5
	Total	65846	48	40	7	5
7	A	13812	49	39	6	5
	B	13667	49	40	6	5
	C	13645	49	39	6	5
	D	13533	49	39	6	5
	E	13473	50	39	6	5
	Total	68130	49	39	6	5
8	A	11647	50	39	7	5
	B	11508	50	39	6	5
	C	11468	50	39	6	5
	D	11434	50	38	6	5
	E	11363	50	39	6	5
	F	11276	51	38	6	5
	Total	68696	50	39	6	5
10	A	6934	76	20	2	2

* 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
2005 MSA Student Gender by Grade Level and Test Form*

Grade	Test Form	Number of Students**	Percent Male	Percent Female
3	A	12657	52	48
	B	12442	51	49
	C	12258	51	49
	D	12132	50	50
	E	12020	51	49
	Total	61509	51	49
4	A	12965	53	47
	B	12796	51	49
	C	12638	51	49
	D	12530	51	49
	E	12401	50	50
	Total	63330	51	49
5	A	13330	52	47
	B	13153	51	49
	C	13043	51	49
	D	12861	51	49
	E	12736	52	48
	Total	65123	52	48
6	A	13388	52	48
	B	13247	52	48
	C	13090	51	49
	D	13090	51	49
	E	13031	51	49
	Total	65846	51	49
7	A	13812	53	47
	B	13667	52	48
	C	13645	51	49
	D	13533	51	49
	E	13473	51	49
	Total	68130	52	48
8	A	11647	52	48
	B	11508	51	49
	C	11468	51	49
	D	11434	51	49
	E	11363	51	49
	F	11276	51	49
	Total	68696	51	49
10	A	6934	51	49

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

Norm Referenced Test (NRT)

NRT Test Design

In 2005, 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. The MSA Grade 10 Reading test included the *TerraNova* English Language Arts Survey Form C.   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. The *TerraNova* English Language Arts Survey for Grade 10 consists of 34 reading items and 26 language items. *TerraNova* Mathematics scale scores based on IRT pattern scoring were reported. Scores on the *TerraNova* English Language Arts Survey were not reported in 2005.[1]

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
RD10	34	Not Reported

MA: Mathematics   RD: Reading

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[2].   The NP shows that Maryland students' performance on the NRT was higher (53rd through 61st percentile) than the national average. As can be seen from NCS, SS, and NP in Tables 8 and 9, students' 2005 performance at most grade levels was somewhat higher than in 2004. Note that performance cuts in Table 9 were obtained from the *TerraNova* standard setting, not the Maryland standard setting.

---

[1] The 2005 Reading assessment included an intact *TerraNova* English Language Arts survey form, and selected *TerraNova* items contributed to the reported CRT Reading scores. However, no separate *TerraNova* scores were reported for Reading examinees in 2005.

[2] 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.

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, 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 2005 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 Hispanic students showed smaller score increases between Grade 6 and Grade 7 than did White students.

Tables 12 to 18 show the proportion of students answering each NRT item correctly (i.e., item p-values) in 2004 and 2005. Tables 12 to 17 show the performance on the *TerraNova* Mathematics items in Grades 3 through 8; Table 18 shows the performance on *TerraNova* Reading items in Grade 10. For most items in Grades 3 through 6, p-values were higher in 2005 than in 2004, indicating an improvement in student performance. At Grades 7 and above, the 2004 and 2005 the average p-values were approximately the same in 2004 and 2005, 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	61366	23.97	4.431	-0.944	0.719	0.817	1.897
4	63183	23.02	5.934	-0.503	-0.489	0.863	2.195
5	64967	24.28	5.768	-0.729	-0.189	0.863	2.134
6	65255	21.34	6.475	-0.487	-0.721	0.875	2.288
7	67261	21.24	7.058	-0.371	-0.886	0.892	2.323
8	67839	20.75	5.937	-0.278	-0.678	0.851	2.291

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

Grade	N	Mean	SD	Skewness	Kurtosis	NP
3	61366	619	46.4	0.435	1.461	61
4	63183	636	45.2	0.341	2.756	57
5	64967	662	53.4	0.413	0.882	61
6	65255	673	50.7	-0.097	1.989	58
7	67261	677	53.6	0.312	2.489	53
8	67839	700	51.7	0.316	2.023	58

Table 8
NRT Summary Statistics: 2004 and 2005 State Means

Grade	2004			2005		
	NCS	SS	NP	NCS	SS	NP
3	23.522	615	59	23.966	619	61
4	22.400	632	54	23.017	636	57
5	23.646	657	58	24.281	662	61
6	20.665	668	55	21.340	673	58
7	21.152	676	53	21.239	677	53
8	20.719	700	58	20.745	700	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 2004 and 2005

Grade	2004 Performance Level					2005 Performance Level				
	1	2	3	4	5	1	2	3	4	5
3	35.4	35.3	20.7	4.2	4.5	31.9	35.3	22.4	4.9	5.5
4	19.4	34.8	29.6	10.4	5.8	16.6	33.6	30.9	11.9	7.0
5	11.2	22.2	28.4	20.7	17.5	9.7	20.3	27.1	22.1	20.9
6	31.9	24.9	26.2	12.2	5.0	28.3	24.2	27.3	14.0	6.1
7	27.5	23.5	26.6	15.7	6.8	27.8	22.7	26.1	16.0	7.4
8	12.8	17.3	28.9	25.1	15.9	13.7	17.2	27.7	24.6	16.7

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	6239	630.99	46.87	385	740	4699	598.57	40.96	385	740	995	604.58	40.72	385	740
	B	6064	632.49	46.24	385	740	4612	600.96	40.46	385	740	1032	604.27	39.81	385	740
	C	6059	633.23	45.87	385	740	4526	601.81	38.98	385	740	971	605.58	39.50	461	740
	D	5876	632.51	44.57	470	740	4554	602.37	40.53	385	740	956	606.88	38.02	496	740
	E	5936	632.33	46.66	385	740	4499	601.12	40.00	385	740	872	610.40	39.48	385	740
	Total	30174	632.30	46.07	385	740	22890	600.95	40.22	385	740	4826	606.22	39.57	385	740
MA4	A	6412	647.09	44.20	403	770	4855	616.10	40.62	403	770	947	621.22	41.53	403	770
	B	6277	649.64	44.17	403	770	4828	618.20	39.01	403	770	962	624.72	42.06	403	770
	C	6150	649.51	43.22	403	770	4831	619.50	38.71	403	770	913	627.29	40.79	403	770
	D	6125	649.86	43.17	403	770	4770	619.65	38.12	403	770	920	625.45	41.91	403	770
	E	6112	649.16	43.04	403	770	4670	617.88	39.28	403	770	906	627.97	41.87	403	770
	Total	31076	649.04	43.58	403	770	23954	618.26	39.18	403	770	4648	625.29	41.69	403	770
MA5	A	6469	676.74	53.46	430	797	5214	636.51	46.42	430	797	901	650.03	50.90	511	797
	B	6393	677.94	51.68	430	797	5053	640.28	45.52	430	797	967	642.60	50.34	430	797
	C	6338	680.46	50.90	430	797	5055	641.23	45.09	430	797	917	646.78	50.44	430	797
	D	6268	679.16	49.87	430	797	5016	640.55	43.57	430	797	911	649.89	50.20	430	797
	E	6114	680.87	51.82	430	797	4985	640.10	44.64	430	797	893	646.26	48.45	430	797
	Total	31582	679.01	51.59	430	797	25323	639.71	45.10	430	797	4589	647.05	50.14	430	797
MA6	A	6432	687.11	47.99	477	820	5273	649.07	46.55	477	820	897	659.20	47.99	477	820
	B	6301	687.45	48.24	477	820	5231	652.38	46.13	477	820	885	658.87	48.13	477	820
	C	6334	689.14	47.32	477	820	5122	654.05	45.12	477	820	847	662.08	46.48	477	820
	D	6339	688.35	46.26	477	820	5088	651.74	47.22	477	820	865	661.79	49.60	477	820
	E	6347	689.06	47.04	477	820	5062	653.30	47.12	477	820	843	657.27	48.12	477	820
	Total	31753	688.22	47.38	477	820	25776	652.09	46.46	477	820	4337	659.84	48.09	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	6769	691.83	52.39	487	850	5317	652.45	45.16	487	850	873	656.09	52.53	487	850
	B	6626	692.87	51.06	487	850	5299	653.24	45.48	487	850	869	659.39	52.63	487	850
	C	6647	693.66	51.50	487	850	5268	654.44	46.88	487	850	857	659.71	48.33	487	850
	D	6601	694.63	50.42	487	850	5206	653.94	45.70	487	850	859	662.16	51.66	487	850
	E	6684	695.19	51.14	487	850	5162	654.48	45.66	487	850	793	660.37	48.93	487	850
	Total	33327	693.63	51.32	487	850	26252	653.71	45.78	487	850	4251	659.52	50.91	487	850
MA8	A	5778	714.57	50.88	502	872	4404	672.63	44.49	502	872	778	682.20	49.47	502	872
	B	5671	715.70	49.91	502	872	4398	676.74	43.38	502	872	702	683.34	44.65	502	872
	C	5696	715.62	48.41	502	872	4369	677.50	42.75	502	872	700	687.48	39.81	502	872
	D	5720	716.58	48.56	502	872	4318	677.21	43.76	502	872	671	684.67	49.17	502	872
	E	5674	717.41	48.42	502	872	4282	675.69	43.99	502	872	661	682.48	43.77	502	872
	F	5711	717.40	48.66	502	872	4167	677.14	44.29	502	872	685	680.29	48.94	502	872
	Total	34250	716.21	49.16	502	872	25938	676.14	43.80	502	872	4197	683.40	46.18	502	872
RD10	A	Not Reported														

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	6516	617.99	50.06	385	740	6102	616.48	44.37	385	740
	B	6310	620.14	48.36	385	740	6105	617.35	44.49	385	740
	C	6243	621.75	47.94	385	740	5991	618.08	44.04	385	740
	D	6108	621.66	47.65	385	740	5990	617.58	42.96	385	740
	E	6136	621.37	49.30	385	740	5861	618.18	43.60	385	740
	Total	31313	620.55	48.70	385	740	30049	617.53	43.90	385	740
MA4	A	6794	634.57	47.42	403	770	6137	634.34	44.64	403	770
	B	6559	637.19	47.43	403	770	6210	635.81	43.42	403	770
	C	6424	638.02	46.31	403	770	6185	636.65	43.14	403	770
	D	6377	638.77	46.69	403	770	6114	635.74	41.92	403	770
	E	6221	638.68	46.39	403	770	6156	634.52	43.41	403	770
	Total	32375	637.40	46.88	403	770	30802	635.41	43.32	403	770
MA5	A	6971	660.28	56.48	430	797	6316	660.20	53.07	430	797
	B	6721	662.76	55.12	430	797	6399	660.91	51.76	430	797
	C	6694	664.81	54.23	430	797	6326	662.29	51.50	430	797
	D	6568	664.35	52.96	430	797	6264	661.32	50.47	430	797
	E	6570	664.11	54.48	430	797	6132	662.90	52.46	430	797
	Total	33524	663.23	54.71	430	797	31437	661.52	51.86	430	797
MA6	A	6929	669.88	52.99	477	820	6335	672.20	49.76	477	820
	B	6735	672.38	52.14	477	820	6363	672.03	49.11	477	820
	C	6577	675.06	50.99	477	820	6403	673.44	48.90	477	820
	D	6619	672.83	52.41	477	820	6355	673.06	48.97	477	820
	E	6622	674.16	51.61	477	820	6309	673.20	49.60	477	820
	Total	33482	672.83	52.07	477	820	31765	672.79	49.27	477	820
MA7	A	7159	674.24	55.13	487	850	6455	675.79	52.87	487	850
	B	6976	676.86	55.22	487	850	6500	674.76	51.05	487	850
	C	6850	677.49	55.09	487	850	6626	677.02	53.16	487	850
	D	6813	677.65	54.41	487	850	6559	677.40	52.06	487	850
	E	6816	678.52	55.21	487	850	6502	677.71	51.33	487	850
	Total	34614	676.93	55.03	487	850	32642	676.54	52.11	487	850
MA8	A	5926	697.73	56.44	502	872	5577	696.22	49.61	502	872
	B	5807	701.07	54.82	502	872	5545	697.55	48.27	502	872
	C	5743	703.05	53.60	502	872	5601	697.10	46.37	502	872
	D	5734	702.21	54.36	502	872	5555	698.58	47.74	502	872
	E	5675	702.82	55.42	502	872	5530	698.15	48.04	502	872
	F	5656	703.26	53.68	502	872	5484	697.92	49.01	502	872
	Total	34541	701.66	54.77	502	872	33292	697.58	48.18	502	872
RD10	A	Not Reported									

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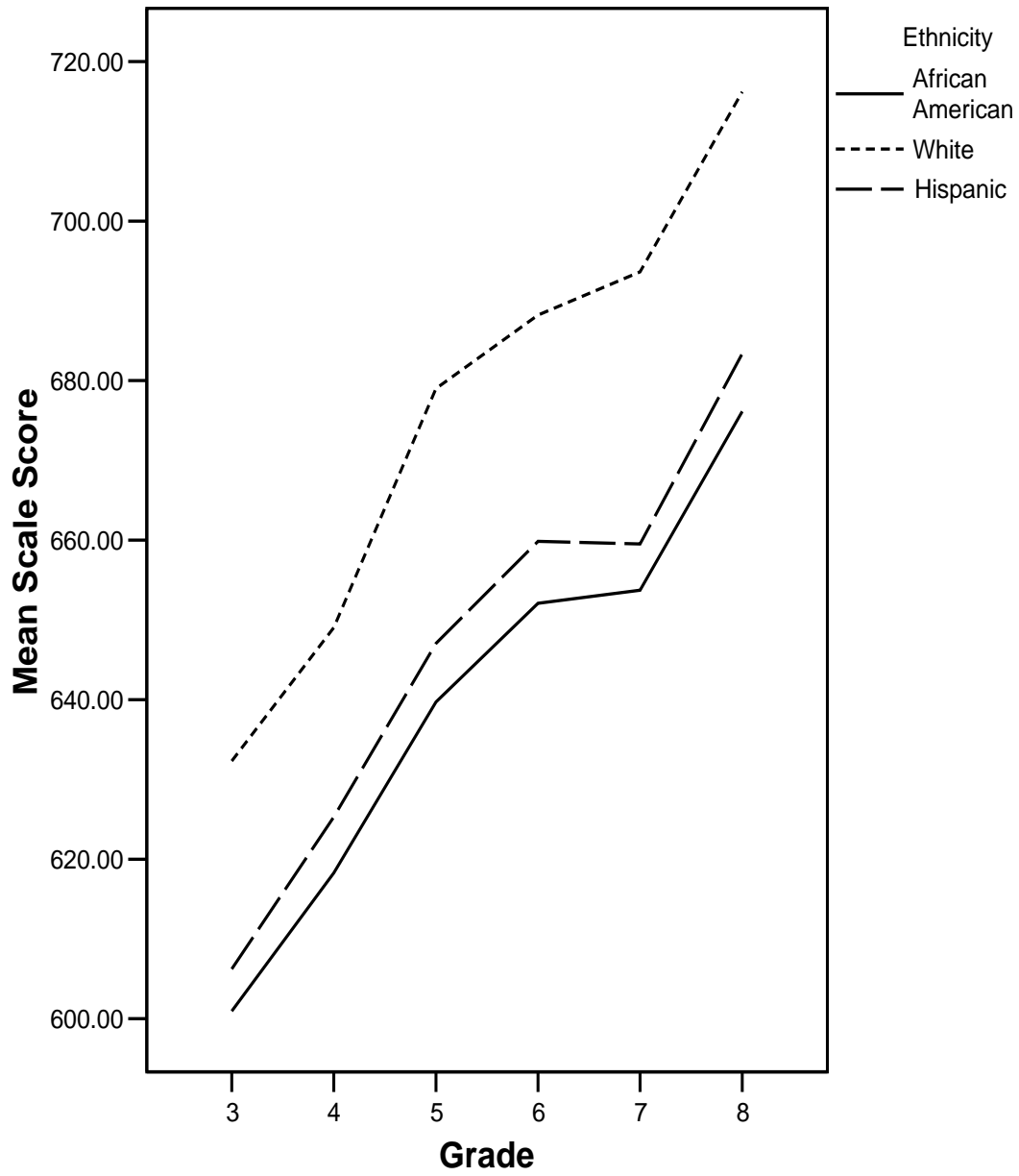
Table 12
Grade 3 Mathematics NRT Item p-values

Item	2004 (N=62961)	2005 (N=61509)	Difference
1	0.85	0.86	0.02
2	0.91	0.92	0.01
3	0.91	0.91	0.01
4	0.88	0.89	0.01
5	0.74	0.76	0.02
6	0.95	0.96	0.00
7	0.67	0.69	0.02
8	0.69	0.72	0.03
9	0.82	0.83	0.01
10	0.88	0.90	0.01
11	0.90	0.91	0.01
12	0.89	0.90	0.01
13	0.86	0.88	0.02
14	0.86	0.88	0.02
15	0.83	0.86	0.03
16	0.95	0.96	0.01
17	0.90	0.91	0.01
18	0.97	0.97	0.00
19	0.64	0.68	0.03
20	0.62	0.63	0.01
21	0.56	0.57	0.01
22	0.74	0.75	0.02
23	0.86	0.88	0.02
24	0.97	0.97	0.01
25	0.90	0.91	0.02
26	0.72	0.73	0.01
27	0.69	0.71	0.02
28	0.50	0.51	0.02
29	0.38	0.41	0.03
30	0.44	0.46	0.02
Average	0.78	0.80	0.01

Table 13
Grade 4 Mathematics NRT Item p-values

Item	2004 (N=65013)	2005 (N=63330)	Difference
1	0.79	0.81	0.02
2	0.63	0.66	0.04
3	0.78	0.83	0.05
4	0.58	0.60	0.02
5	0.55	0.55	0.00
6	0.81	0.82	0.01
7	0.69	0.70	0.02
8	0.51	0.52	0.02
9	0.54	0.56	0.02
10	0.52	0.53	0.01
11	0.97	0.98	0.00
12	0.95	0.95	0.01
13	0.35	0.38	0.03
14	0.72	0.73	0.01
15	0.80	0.82	0.02
16	0.83	0.85	0.02
17	0.89	0.91	0.01
18	0.53	0.56	0.03
19	0.87	0.89	0.02
20	0.74	0.74	0.01
21	0.65	0.66	0.01
22	0.82	0.84	0.02
23	0.64	0.68	0.04
24	0.84	0.86	0.02
25	0.82	0.85	0.03
26	0.68	0.71	0.03
27	0.57	0.61	0.03
28	0.85	0.87	0.01
29	0.62	0.64	0.02
30	0.80	0.82	0.02
31	0.48	0.51	0.03
32	0.50	0.53	0.03
Average	0.70	0.72	0.02

Table 14
Grade 5 Mathematics NRT Item p-values

Item	2004 (N=66176)	2005 (N=65123)	Difference
1	0.84	0.85	0.01
2	0.71	0.74	0.03
3	0.73	0.78	0.06
4	0.65	0.69	0.03
5	0.71	0.73	0.01
6	0.80	0.81	0.02
7	0.74	0.77	0.03
8	0.60	0.63	0.03
9	0.91	0.92	0.01
10	0.91	0.92	0.01
11	0.93	0.94	0.01
12	0.72	0.75	0.03
13	0.80	0.81	0.00
14	0.70	0.71	0.01
15	0.81	0.82	0.01
16	0.90	0.91	0.01
17	0.66	0.67	0.00
18	0.97	0.98	0.00
19	0.60	0.62	0.02
20	0.95	0.95	0.00
21	0.81	0.82	0.01
22	0.73	0.75	0.01
23	0.66	0.68	0.02
24	0.53	0.58	0.05
25	0.54	0.56	0.02
26	0.51	0.55	0.05
27	0.62	0.66	0.04
28	0.64	0.66	0.02
29	0.81	0.81	0.00
30	0.70	0.73	0.03
31	0.59	0.61	0.02
32	0.78	0.84	0.05
Average	0.74	0.76	0.02

Table 15
Grade 6 Mathematics NRT Item p-values

Item	2004 (N=68229)	2005 (N=65846)	Difference
1	0.75	0.74	0.00
2	0.76	0.78	0.03
3	0.70	0.70	0.00
4	0.73	0.75	0.02
5	0.79	0.81	0.01
6	0.69	0.71	0.02
7	0.54	0.56	0.03
8	0.74	0.75	0.01
9	0.92	0.92	0.00
10	0.81	0.83	0.01
11	0.54	0.56	0.02
12	0.61	0.65	0.04
13	0.73	0.79	0.06
14	0.76	0.78	0.01
15	0.73	0.76	0.03
16	0.69	0.71	0.02
17	0.69	0.74	0.05
18	0.74	0.75	0.01
19	0.67	0.69	0.02
20	0.66	0.70	0.03
21	0.50	0.53	0.03
22	0.56	0.58	0.02
23	0.58	0.58	0.01
24	0.67	0.69	0.02
25	0.61	0.63	0.02
26	0.57	0.61	0.03
27	0.62	0.64	0.01
28	0.55	0.56	0.01
29	0.47	0.52	0.05
30	0.52	0.58	0.05
31	0.54	0.56	0.02
Average	0.66	0.68	0.02

Table 16
Grade 7 Mathematics NRT Item p-values

Item	2004 (N=69237)	2005 (N=68130)	Difference
1	0.71	0.73	0.01
2	0.69	0.65	-0.04
3	0.53	0.52	-0.01
4	0.82	0.81	-0.01
5	0.74	0.73	-0.01
6	0.82	0.82	0.00
7	0.54	0.50	-0.04
8	0.63	0.63	0.00
9	0.69	0.69	0.00
10	0.90	0.90	0.00
11	0.68	0.67	-0.01
12	0.73	0.74	0.00
13	0.61	0.62	0.01
14	0.57	0.57	0.00
15	0.88	0.89	0.01
16	0.53	0.57	0.04
17	0.74	0.72	-0.02
18	0.68	0.67	-0.01
19	0.68	0.68	0.00
20	0.66	0.68	0.02
21	0.80	0.81	0.01
22	0.73	0.72	-0.01
23	0.72	0.74	0.02
24	0.59	0.59	0.01
25	0.63	0.63	0.00
26	0.51	0.52	0.02
27	0.65	0.67	0.02
28	0.55	0.54	0.00
29	0.50	0.51	0.01
30	0.35	0.38	0.02
31	0.52	0.55	0.04
32	0.52	0.52	0.00
Average	0.65	0.66	0.00

Table 17
Grade 8 Mathematics NRT Item p-values

Item	2004 (N=68659)	2005 (N=68696)	Difference
1	0.82	0.81	-0.01
2	0.55	0.57	0.02
3	0.34	0.32	-0.02
4	0.86	0.84	-0.03
5	0.77	0.76	-0.02
6	0.60	0.65	0.05
7	0.65	0.64	-0.02
8	0.84	0.83	-0.01
9	0.67	0.66	-0.01
10	0.89	0.89	-0.01
11	0.89	0.90	0.01
12	0.80	0.79	-0.01
13	0.89	0.88	-0.01
14	0.78	0.76	-0.02
15	0.64	0.64	0.00
16	0.87	0.87	0.00
17	0.74	0.75	0.02
18	0.58	0.58	0.00
19	0.66	0.63	-0.03
20	0.38	0.42	0.04
21	0.72	0.74	0.02
22	0.68	0.70	0.02
23	0.75	0.72	-0.03
24	0.52	0.52	0.00
25	0.63	0.63	0.00
26	0.42	0.42	0.00
27	0.53	0.59	0.06
28	0.55	0.53	-0.01
29	0.54	0.55	0.01
30	0.42	0.43	0.00
31	0.47	0.49	0.02
Average	0.66	0.66	0.00

Table 18
Grade 10 Reading NRT Item p-values

Item	2004 (N=62958)	2005 (N=6934)	Difference	Item	2004 (N=62958)	2005 (N=6934)	Difference
1	0.71	0.71	0.00	31	0.71	0.74	0.03
2	0.56	0.57	0.01	32	0.53	0.49	-0.04
3	0.77	0.80	0.03	33	0.53	0.54	0.01
4	0.86	0.87	0.01	34	0.54	0.51	-0.03
5	0.42	0.40	-0.02	35	0.73	0.73	0.01
6	0.75	0.76	0.01	36	0.69	0.68	-0.01
7	0.82	0.82	0.00	37	0.61	0.59	-0.02
8	0.60	0.58	-0.03	38	0.64	0.64	0.00
9	0.39	0.39	0.01	39	0.66	0.66	0.00
10	0.66	0.65	-0.01	40	0.68	0.68	0.00
11	0.64	0.63	-0.01	41	0.54	0.51	-0.02
12	0.84	0.85	0.01	42	0.60	0.59	-0.02
13	0.78	0.78	0.00	43	0.78	0.79	0.01
14	0.66	0.64	-0.02	44	0.74	0.74	-0.01
15	0.52	0.50	-0.01	45	0.79	0.80	0.01
16	0.42	0.44	0.02	46	0.42	0.38	-0.03
17	0.56	0.54	-0.02	47	0.72	0.71	-0.01
18	0.69	0.70	0.02	48	0.61	0.59	-0.02
19	0.67	0.64	-0.02	49	0.57	0.58	0.01
20	0.57	0.57	0.00	50	0.68	0.66	-0.03
21	0.69	0.70	0.02	51	0.70	0.70	0.00
22	0.54	0.55	0.01	52	0.58	0.58	0.00
23	0.71	0.72	0.01	53	0.42	0.40	-0.02
24	0.56	0.57	0.00	54	0.57	0.58	0.01
25	0.54	0.55	0.00	55	0.47	0.46	-0.02
26	0.72	0.73	0.01	56	0.59	0.58	-0.01
27	0.53	0.52	-0.01	57	0.72	0.72	-0.01
28	0.65	0.66	0.01	58	0.45	0.45	0.00
29	0.56	0.57	0.01	59	0.52	0.53	0.01
30	0.46	0.46	0.00	60	0.71	0.72	0.01
				Average	0.62	0.62	0.00

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 19 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, D (and F in grade 8) contain the same operational items and are designated as Form 2.[3] For Grade 10 Reading, only one form (Form A) was administered in 2005. As can be seen in Table 19, the total number of operational items and score points was the same for all test forms within a grade.

Table 20 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 2005 scale on the 2004 scale. Common items (which included many, but not necessarily all, of the anchor items) were used for linking alternate forms.

Tables 21 to 27 present the number of items and score points by Maryland content reporting standards. There are five reporting standards for Mathematics across grades, and three standards for Reading. For Grades 3 through 7, the number of items and score points for each reporting standard were identical across forms within each grade. For Grade 8, the two operational forms differed by one point on standard 01 (Algebra, Patterns, and Functions) and standard 06 (Number Relationships and Computation). The actual values shown in Tables 21 to 27 are identical to the target values (shown in Table 1) for Reading and for Grades 3 through 6 Mathematics, and are within one point of all target values for Grades 7 and 8 Mathematics.

---

[3] 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 19
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	76
	2	11	25	16	12	64	76
RD10	A	34	15	4	-	53	61

- For grades 3 through 7, Form 1 consists of Forms A, C, & E and Form 2 consists of Forms B & D.
- For grade 8, Form 1 consists of Forms A, C, & E and Form 2 consists of Forms B, D, & F.
- For all grades, counts are without field test items.

Table 20
The Number of Items by Function

Content Grade	Form	Total Items*	Anchor Items	Common Items	Unique Items	Field-Test Items
MA3	A	77	32	57	8	12
	B	77	32	57	8	12
	C	83	32	57	8	18
	D	83	32	57	8	18
	E	83	32	57	8	18
MA4	A	76	42	36	28	12
	B	76	45	36	28	12
	C	82	42	36	28	18
	D	82	45	36	28	18
	E	82	42	36	28	18
MA5	A	81	28	55	10	16
	B	81	28	55	10	16
	C	80	28	55	10	15
	D	80	28	55	10	15
	E	80	28	55	10	15
MA6	A	77	40	32	30	15
	B	73	41	32	30	11
	C	73	40	32	30	11
	D	72	41	32	30	10
	E	72	40	32	30	10
MA7	A	83	33	33	29	21
	B	83	33	33	29	21
	C	75	33	33	29	13
	D	73	33	33	29	11
	E	75	33	33	29	13
MA8	A	85	30	31	33	21
	B	80	24	31	33	16
	C	80	30	31	33	16
	D	78	24	31	33	14
	E	82	30	31	33	18
	F	78	24	31	33	14
RD10	A	53	-	-	-	0

- * Total = Common + Unique
- For grades 3 through 7, common items are items that appear both on Form 1 (Forms A, C, & E) and Form 2 (Forms B & D).
- For grade 8, common items are items that appear both on Form 1 (Forms A, C, & E) and Form 2 (Forms B, D, & F).

Table 21

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 22

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 23

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 24

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

Standards	Forms A, C & E							Forms B & D						
	NRT	Custom		Total				NRT	Custom		Total			
	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 25

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

Standards	Forms A, C & E								Forms B & D							
	NRT	Custom			Total				NRT	Custom			Total			
	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 26

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

Standards	Form A, C & E								Form B & D							
	NRT	Custom			Total				NRT	Custom			Total			
	SR	SR	CR	GR	Items	%	Points	%	SR	SR	CR	GR	Items	%	Points	%
01	2	5	3	4	14	22	14	18	2	6	3	4	15	23	15	20
02/03	2	6	2	2	12	19	12	16	2	5	2	3	12	19	12	16
04/05	1	7	3	4	15	23	15	20	1	8	3	3	15	23	15	20
06	6	7	0	2	15	23	15	20	6	6	0	2	14	22	14	18
07	0	0	8	0	8	13	20	26	0	0	8	0	8	13	20	26
Sum	11	25	16	12	64	100	76	100	11	25	16	12	64	100	76	100

Table 27

The Number of Items and Score Points by Maryland Content Standard for Grade 10

| | Form A | | | | | | |
| | NRT | Custom | | Total | | | |
Standards	SR	SR	CR	Items	%	Points	%
01	13	3	0	16	30	16	26
02	10	6	2	18	34	22	36
03	11	6	2	19	36	23	38
Sum	34	15	4	53	100	61	100

Classical Item Analysis

Tables A1- A30 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-B7 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 97.7% to 100% for Mathematics items and from 98.9% to 99.3% for Reading items. The percentage of perfect agreement (i.e., identical scores assigned by rater 1 and rater 2) ranged from 74.2% to 99.7% in Grade 3, from 74.5% to 99.2% in Grade 4, from 77.6% to 99.7% in Grade 5, from 76.5% to 99.6% in Grade 6, from 74.8% to 99.7% in Grade 7, from 82.5% to 99.5% in Grade 8, and from 63.5% to 70.9% in Grade 10.

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 94.7% to 99.7% in Grade 3, 94.0% to 99.2% in Grade 4, 93.4% to 99.7% in Grade 5, 97.2% to 99.6% in Grade 6, 93.9% to 99.7% in Grade 7 and 86.8 to 99.5% 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 B8-B16 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 $-1.5 \le \Delta_{MH} \le -1$ or $1 \le \Delta_{MH} \le 1.5$ (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| $\ge$ 0.25

Appendix tables C1-C7 show items flagged based on the above criteria. In the column "Focal", for those items flagged for ethnicity, the number 2 represents African American and the number 4 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 2005 calibration of the Mathematics items, several items exhibited moderate misfit. Across all operational test forms, one misfitting item was identified at Grade 3, three items at Grade 4, two at Grade 5, two at Grade 6, six at Grade 7, and seven 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.

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

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

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 2005 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 2005 assessments, there were no items with inconsistent parameters across the two forms.

Step 2: Thus, all of the shared items 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 2005 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 2005 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 2005 tests, all of the anchor items were judged to be sufficiently stable, an all were used as equating anchors. Item parameters for the 2005 tests were transformed to the MSA CRT reporting scale using these anchor items and Stocking and Lord's transformation procedure.

Calibration and Equating Results

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.

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 28 presents the lowest obtainable scale scores (LOSS) and the highest obtainable scale scores (HOSS). For the 2005 assessments, MSDE requested that the previous grade-specific LOSS and HOSS values be reset to a common LOSS of 240 and HOSS of 650 across all grades.

Table 28
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 2005 item parameters were placed on the MSA CRT reporting scale using previously calibrated items from the 2003 and 2004 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 29, and 30, distributions of raw scores and scale scores were similar across forms, except at Grade 7, where raw scores were more than 4 points higher on Form 2 than on Form 1.   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 31 and 32 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. The largest difference between male and female students was on the Grade 10 Reading test, with female students scoring more than 14 points higher than male students.

Figures G1-G21 in Appendix G show histograms for the distribution of scale scores for the total population and for subgroups defined by ethnicity and gender.

Table 29
CRT Raw Score Descriptive Statistics

Grade Content	Form	N Count	Mean	Mean P-Value	SD	Min	Max	Alpha	SEM
MA3	1	36935	54.03	0.75	11.96	0	72	0.93	3.20
	2	24574	52.93	0.74	11.32	0	72	0.92	3.20
	Total	61509	53.59	0.74	11.72	0	72	.	.
MA4	1	38004	43.63	0.61	14.22	0	71	0.94	3.51
	2	25326	43.88	0.62	14.06	0	71	0.94	3.45
	Total	63330	43.73	0.62	14.16	0	71	.	.
MA5	1	39109	43.83	0.59	15.49	0	74	0.94	3.68
	2	26014	45.23	0.61	15.03	0	74	0.94	3.76
	Total	65123	44.39	0.60	15.32	0	74	.	.
MA6	1	39509	37.12	0.53	15.27	0	70	0.94	3.63
	2	26337	37.07	0.53	14.73	0	70	0.94	3.62
	Total	65846	37.10	0.53	15.05	0	70	.	.
MA7	1	40930	32.95	0.46	16.44	0	71	0.96	3.43
	2	27200	37.09	0.52	17.89	0	72	0.96	3.59
	Total	68130	34.60	0.48	17.16	0	72	.	.
MA8	1	34478	35.42	0.47	17.20	0	76	0.95	3.78
	2	34218	33.91	0.45	16.66	0	76	0.95	3.75
	Total	68696	34.67	0.46	16.95	0	76	.	.
RD10	A	6934	36.05	0.59	12.50	0	60	0.93	3.41

Table 30
CRT Scale Score Descriptive Statistics

Grade Content	Form	N Count	Mean	SD	MIN	MAX
MA3	1	36935	411.05	48.39	240	650
	2	24574	410.89	46.59	240	650
	Total	61509	410.99	47.68	240	650
MA4	1	38004	403.75	46.33	240	650
	2	25326	404.19	44.07	240	650
	Total	63330	403.93	45.44	240	650
MA5	1	39109	411.25	45.68	240	650
	2	26014	411.16	45.38	240	650
	Total	65123	411.22	45.56	240	650
MA6	1	39509	402.62	46.97	240	650
	2	26337	401.20	46.44	240	650
	Total	65846	402.05	46.76	240	650
MA7	1	40930	397.54	51.67	240	564
	2	27200	398.10	50.10	240	650
	Total	68130	397.76	51.05	240	650
MA8	1	34478	404.99	46.65	240	650
	2	34218	405.23	45.94	240	650
	Total	68696	405.11	46.30	240	650
RD10	A	6934	387.49	49.08	240	561

Table 31
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	18262	425.12	45.88	240	650	13765	391.30	43.61	240	650	2847	397.47	42.77	240	650
	2	11959	424.78	43.80	240	650	9198	392.42	43.47	240	650	1995	396.71	41.47	240	650
	Total	30221	424.99	45.06	240	650	22963	391.75	43.56	240	650	4842	397.16	42.23	240	650
MA4	1	18709	418.35	42.04	240	650	14397	383.16	43.41	240	650	2770	391.84	43.47	240	540
	2	12421	417.90	38.56	240	650	9633	385.70	42.84	240	531	1890	391.27	45.06	240	534
	Total	31130	418.17	40.69	240	650	24030	384.18	43.20	240	650	4660	391.61	44.11	240	540
MA5	1	18956	426.37	41.40	240	650	15306	390.88	41.83	240	650	2717	397.84	44.30	240	650
	2	12680	425.29	41.66	240	650	10099	392.79	41.43	240	535	1884	394.47	45.92	240	541
	Total	31636	425.94	41.51	240	650	25405	391.64	41.68	240	650	4601	396.46	45.00	240	650
MA6	1	19201	417.26	41.42	240	650	15665	383.36	45.74	240	650	2602	389.47	44.51	240	509
	2	12709	415.31	40.77	240	650	10486	382.53	45.97	240	650	1763	389.56	43.94	240	497
	Total	31910	416.48	41.17	240	650	26151	383.03	45.84	240	650	4365	389.51	44.27	240	509
MA7	1	20220	415.60	44.53	240	564	16096	372.60	49.66	240	523	2552	385.14	46.86	240	510
	2	13308	417.02	42.23	240	650	10735	372.63	47.48	240	518	1752	386.08	48.65	240	543
	Total	33528	416.16	43.63	240	650	26831	372.61	48.80	240	523	4304	385.52	47.60	240	543
MA8	1	17261	420.00	40.34	240	650	13337	383.01	45.60	240	510	2162	395.56	41.10	240	536
	2	17228	420.36	39.21	240	650	13144	383.91	45.03	240	526	2091	391.25	45.19	240	542
	Total	34489	420.18	39.78	240	650	26481	383.46	45.32	240	526	4253	393.44	43.21	240	542
RD10	A	5248	396.28	45.85	240	561	1420	355.15	47.14	240	480	158	377.30	47.02	240	469

Table 32
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	18938	409.82	49.46	240	650	17992	412.38	47.15	240	650
	2	12449	409.88	47.85	240	650	12123	411.92	45.25	240	650
	Total	31387	409.84	48.83	240	650	30115	412.20	46.40	240	650
MA4	1	19480	402.51	47.68	240	650	18520	405.06	44.81	240	650
	2	12969	403.07	45.62	240	650	12353	405.40	42.33	240	650
	Total	32449	402.73	46.87	240	650	30873	405.20	43.84	240	650
MA5	1	20293	410.35	47.87	240	650	18813	412.23	43.16	240	650
	2	13323	410.91	47.13	240	650	12688	411.42	43.47	240	650
	Total	33616	410.57	47.58	240	650	31501	411.90	43.29	240	650
MA6	1	20329	400.67	49.48	240	650	19164	404.80	43.87	240	650
	2	13526	398.72	49.38	240	650	12805	403.88	42.85	240	650
	Total	33855	399.89	49.45	240	650	31969	404.43	43.47	240	650
MA7	1	21118	394.22	54.90	240	562	19798	401.18	47.61	240	564
	2	14008	395.93	53.54	240	650	13185	400.46	45.97	240	650
	Total	35126	394.90	54.37	240	650	32983	400.89	46.96	240	650
MA8	1	17605	401.72	50.04	240	650	16872	408.41	42.56	240	650
	2	17458	402.17	49.11	240	650	16754	408.44	42.12	240	549
	Total	35063	401.95	49.58	240	650	33626	408.42	42.34	240	650
RD10	A	3526	380.56	51.11	240	537	3408	394.66	45.80	240	561

The Relationship between NRT and CRT

Each of the 2005 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 33. The correlation was relatively high and similar across alternate forms within grade. The correlations ranged from 0.80 to 0.86 in Mathematics. Reading NRT scores were not computed in 2005.

Table 33
Correlation between NRT and CRT

CRT Form	Content/Grade					
	MA3	MA4	MA5	MA6	MA7	MA8
Total	0.81	0.84	0.86	0.82	0.82	0.83
1	0.80	0.84	0.86	0.82	0.81	0.83
2	0.81	0.83	0.86	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 34 and 35 show the raw score and scale score results for each content standard.

Tables 36 and 37 show the raw score Pearson product-moment and Spearman Rho correlations among the content standards at each grade level. Tables 38 and 39 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.[4] 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.

---

[4] 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 34
Distribution of Raw Scores on Content Standards

Grade	Form	Content Standard	N	Maximum Possible	Mean	SD	Minimum	Maximum
3	1	1	36935	13	9.73	2.38	0	13
		2&3	36935	15	12.00	2.59	0	15
		4&5	36935	14	10.72	2.59	0	14
		6	36935	16	12.95	2.91	0	16
		7	36935	14	8.63	3.27	0	14
	2	1	24574	13	9.81	2.29	0	13
		2&3	24574	15	11.80	2.50	0	15
		4&5	24574	14	10.75	2.55	0	14
		6	24574	16	12.76	2.83	0	16
		7	24574	14	7.81	3.03	0	14
4	1	1	38004	14	10.11	2.94	0	14
		2&3	38004	14	8.54	3.14	0	14
		4&5	38004	15	8.95	3.79	0	15
		6	38004	14	10.14	2.82	0	14
		7	38004	14	5.89	3.51	0	14
	2	1	25326	14	9.95	2.91	0	14
		2&3	25326	14	8.18	3.12	0	14
		4&5	25326	15	9.66	3.82	0	15
		6	25326	14	10.57	2.69	0	14
		7	25326	14	5.53	3.39	0	14
5	1	1	39109	15	9.31	3.29	0	15
		2&3	39109	14	7.50	3.15	0	14
		4&5	39109	13	8.93	3.08	0	13
		6	39109	15	9.49	3.50	0	15
		7	39109	17	8.61	4.22	0	17
	2	1	26014	15	9.29	3.34	0	15
		2&3	26014	14	7.98	2.97	0	14
		4&5	26014	13	8.86	2.99	0	13
		6	26014	15	9.91	3.25	0	15
		7	26014	17	9.19	4.39	0	17

Table 34 (cont.)
Distribution of Raw Scores on Content Standards

Grade	Form	Content Standard	N	Maximum Possible	Mean	SD	Minimum	Maximum
6	1	1	39509	14	7.84	3.41	0	14
		2&3	39509	14	7.37	3.35	0	14
		4&5	39509	13	7.84	3.07	0	13
		6	39509	14	7.27	3.36	0	14
		7	39509	15	6.78	3.85	0	15
	2	1	26337	14	7.66	3.21	0	14
		2&3	26337	14	7.48	3.14	0	14
		4&5	26337	13	7.25	3.01	0	13
		6	26337	14	7.67	3.48	0	14
		7	26337	15	7.01	3.78	0	15
7	1	1	40930	14	6.48	3.66	0	14
		2&3	40930	13	5.55	3.50	0	13
		4&5	40930	14	7.32	3.75	0	14
		6	40930	14	7.13	3.67	0	14
		7	40930	17	6.46	3.46	0	17
	2	1	27200	14	6.90	3.78	0	14
		2&3	27200	13	6.56	3.78	0	13
		4&5	27200	14	7.80	3.79	0	14
		6	27200	14	8.30	3.74	0	14
		7	27200	17	7.52	4.37	0	17
8	1	1	34478	14	7.02	3.67	0	14
		2&3	34478	12	5.97	3.08	0	12
		4&5	34478	15	7.39	3.41	0	15
		6	34478	15	6.91	3.62	0	15
		7	34478	20	8.14	5.16	0	20
	2	1	34218	15	7.08	3.85	0	15
		2&3	34218	12	5.00	2.92	0	12
		4&5	34218	15	7.86	3.32	0	15
		6	34218	14	6.46	3.49	0	14
		7	34218	20	7.51	4.85	0	20
10	1	1	6934	16	10.51	3.56	0	16
		2	6934	22	12.64	4.94	0	22
		3	6934	23	12.90	4.93	0	23

Table 35
Distribution of Scale Scores on Content Standards

Grade	Form	Content Standard	N	Maximum Possible	Mean	SD	Minimum	Maximum
3	1	1	36935	650	427.80	86.58	240	650
		2&3	36935	650	438.93	102.67	240	650
		4&5	36935	650	430.75	90.23	240	650
		6	36935	650	444.01	106.32	240	650
		7	36935	650	414.75	68.49	240	650
	2	1	24574	650	427.18	86.31	240	650
		2&3	24574	650	434.71	94.77	240	650
		4&5	24574	650	430.57	89.03	240	650
		6	24574	650	435.84	95.44	240	650
		7	24574	650	411.29	60.39	240	650
4	1	1	38004	650	421.35	87.48	240	650
		2&3	38004	650	408.69	72.80	240	650
		4&5	38004	650	409.66	75.62	240	650
		6	38004	650	421.49	89.67	240	650
		7	38004	650	400.46	62.21	240	650
	2	1	25326	650	422.29	85.58	240	650
		2&3	25326	650	407.44	64.50	240	650
		4&5	25326	650	416.43	82.30	240	650
		6	25326	650	426.62	93.30	240	650
		7	25326	650	397.59	58.04	240	650
5	1	1	39109	650	413.95	65.81	240	650
		2&3	39109	650	408.29	70.99	240	650
		4&5	39109	650	428.10	90.83	240	650
		6	39109	650	416.65	70.74	240	650
		7	39109	650	411.22	52.32	240	650
	2	1	26014	650	414.01	65.44	240	650
		2&3	26014	650	411.91	65.93	240	650
		4&5	26014	650	425.00	85.29	240	650
		6	26014	650	418.09	70.84	240	650
		7	26014	650	412.33	59.39	240	650

Table 35 (cont.)
Distribution of Scale Scores on Content Standards

Grade	Form	Content Standard	N	Maximum Possible	Mean	SD	Minimum	Maximum
6	1	1	39509	650	405.75	70.92	240	650
		2&3	39509	650	399.10	76.60	240	650
		4&5	39509	650	408.60	73.34	240	650
		6	39509	650	396.27	77.74	240	650
		7	39509	650	401.49	58.33	240	650
	2	1	26337	650	401.41	63.80	240	650
		2&3	26337	650	402.80	68.95	240	650
		4&5	26337	650	400.41	69.95	240	650
		6	26337	650	400.88	76.50	240	650
		7	26337	650	399.02	57.12	240	650
7	1	1	40930	650	390.51	81.31	240	650
		2&3	40930	650	389.46	82.66	240	650
		4&5	40930	650	398.29	72.94	240	650
		6	40930	650	402.81	77.97	240	650
		7	40930	650	395.01	54.70	240	650
	2	1	27200	650	395.47	83.12	240	650
		2&3	27200	650	398.49	87.99	240	650
		4&5	27200	650	402.42	76.97	240	650
		6	27200	650	411.44	87.72	240	650
		7	27200	650	393.70	57.18	240	650
8	1	1	34478	650	408.79	73.08	240	650
		2&3	34478	650	408.89	75.23	240	650
		4&5	34478	650	403.61	59.61	240	650
		6	34478	650	391.12	82.44	240	650
		7	34478	650	401.52	53.71	240	650
	2	1	34218	650	405.32	66.52	240	650
		2&3	34218	650	402.22	68.22	240	650
		4&5	34218	650	404.95	57.56	240	650
		6	34218	650	393.64	84.93	240	650
		7	34218	650	400.18	54.94	240	650
10	1	1	6934	650	398.31	70.85	240	650
		2	6934	650	385.88	58.81	240	650
		3	6934	650	387.22	57.35	240	650

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

Mathematics

Grade	Content Standard	Mean	SD	1	2&3	4&5	6	7
3	1	9.76	2.34	1.00	0.68	0.71	0.72	0.62
	2&3	11.92	2.56		1.00	0.70	0.71	0.64
	4&5	10.73	2.57			1.00	0.74	0.66
	6	12.88	2.88				1.00	0.67
	7	8.30	3.20					1.00
4	1	10.04	2.93	1.00	0.70	0.73	0.71	0.71
	2&3	8.40	3.14		1.00	0.72	0.67	0.72
	4&5	9.23	3.82			1.00	0.71	0.74
	6	10.31	2.77				1.00	0.68
	7	5.75	3.47					1.00
5	1	9.30	3.31	1.00	0.71	0.74	0.76	0.79
	2&3	7.69	3.09		1.00	0.70	0.70	0.73
	4&5	8.90	3.05			1.00	0.75	0.76
	6	9.66	3.41				1.00	0.78
	7	8.84	4.30					1.00
6	1	7.77	3.33	1.00	0.73	0.73	0.75	0.80
	2&3	7.41	3.27		1.00	0.68	0.73	0.77
	4&5	7.61	3.06			1.00	0.71	0.72
	6	7.43	3.41				1.00	0.77
	7	6.87	3.83					1.00
7	1	6.65	3.71	1.00	0.78	0.78	0.81	0.80
	2&3	5.96	3.65		1.00	0.76	0.79	0.81
	4&5	7.51	3.77			1.00	0.79	0.83
	6	7.60	3.74				1.00	0.79
	7	6.88	3.89					1.00
8	1	7.05	3.76	1.00	0.73	0.76	0.78	0.85
	2&3	5.49	3.04		1.00	0.69	0.72	0.79
	4&5	7.62	3.38			1.00	0.72	0.79
	6	6.69	3.56				1.00	0.77
	7	7.82	5.02					1.00

Reading

Grade	Content Standard	Mean	SD	1	2	3
10	1	10.51	3.56	1.00	0.78	0.79
	2	12.64	4.94	0.78	1.00	0.82
	3	12.90	4.93	0.79	0.82	1.00

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

Mathematics								
Grade	Content Standard	Mean	SD	1	2&3	4&5	6	7
3	1	9.76	2.34	1.00	0.64	0.67	0.67	0.60
	2&3	11.92	2.56		1.00	0.66	0.66	0.62
	4&5	10.73	2.57			1.00	0.69	0.64
	6	12.88	2.88				1.00	0.65
	7	8.30	3.20					1.00
4	1	10.04	2.93	1.00	0.70	0.73	0.70	0.73
	2&3	8.40	3.14		1.00	0.72	0.67	0.73
	4&5	9.23	3.82			1.00	0.71	0.75
	6	10.31	2.77				1.00	0.70
	7	5.75	3.47					1.00
5	1	9.30	3.31	1.00	0.72	0.75	0.76	0.79
	2&3	7.69	3.09		1.00	0.71	0.71	0.74
	4&5	8.90	3.05			1.00	0.74	0.76
	6	9.66	3.41				1.00	0.79
	7	8.84	4.30					1.00
6	1	7.77	3.33	1.00	0.73	0.73	0.75	0.80
	2&3	7.41	3.27		1.00	0.68	0.72	0.77
	4&5	7.61	3.06			1.00	0.71	0.72
	6	7.43	3.41				1.00	0.77
	7	6.87	3.83					1.00
7	1	6.65	3.71	1.00	0.77	0.79	0.80	0.80
	2&3	5.96	3.65		1.00	0.77	0.79	0.81
	4&5	7.51	3.77			1.00	0.79	0.84
	6	7.60	3.74				1.00	0.79
	7	6.88	3.89					1.00
8	1	7.05	3.76	1.00	0.71	0.76	0.76	0.85
	2&3	5.49	3.04		1.00	0.68	0.70	0.77
	4&5	7.62	3.38			1.00	0.71	0.79
	6	6.69	3.56				1.00	0.76
	7	7.82	5.02					1.00
Reading								
Grade	Content Standard	Mean	SD	1	2	3		
10	1	10.51	3.56	1.000	0.768	0.778		
	2	12.64	4.94		1.000	0.815		
	3	12.90	4.93			1.000		

Table 38
Scale Score Correlations (Pearson Product-Moment) between Content Standards

Mathematics

Grade	Content Standard	Mean	SD	1	2&3	4&5	6	7
3	1	427.55	86.48	1.00	0.54	0.55	0.52	0.53
	2&3	437.25	99.61		1.00	0.52	0.51	0.52
	4&5	430.68	89.75			1.00	0.52	0.53
	6	440.75	102.19				1.00	0.53
	7	413.36	65.40					1.00
4	1	421.73	86.73	1.00	0.57	0.56	0.55	0.60
	2&3	408.19	69.60		1.00	0.59	0.57	0.65
	4&5	412.37	78.43			1.00	0.55	0.64
	6	423.54	91.18				1.00	0.60
	7	399.32	60.59					1.00
5	1	413.98	65.66	1.00	0.64	0.62	0.65	0.71
	2&3	409.74	69.04		1.00	0.59	0.61	0.67
	4&5	426.86	88.67			1.00	0.60	0.64
	6	417.22	70.79				1.00	0.68
	7	411.66	55.26					1.00
6	1	404.02	68.19	1.00	0.61	0.64	0.62	0.70
	2&3	400.58	73.66		1.00	0.60	0.59	0.67
	4&5	405.32	72.12			1.00	0.60	0.66
	6	398.11	77.28				1.00	0.65
	7	400.50	57.86					1.00
7	1	392.49	82.07	1.00	0.66	0.67	0.66	0.71
	2&3	393.07	84.95		1.00	0.66	0.64	0.72
	4&5	399.94	74.60			1.00	0.66	0.76
	6	406.26	82.11				1.00	0.67
	7	394.49	55.71					1.00
8	1	407.06	69.91	1.00	0.65	0.69	0.61	0.73
	2&3	405.57	71.90		1.00	0.65	0.58	0.67
	4&5	404.28	58.61			1.00	0.61	0.76
	6	392.38	83.70				1.00	0.61
	7	400.85	54.33					1.00

Reading

Grade	Content Standard	Mean	SD	1	2	3
10	1	398.31	70.85	1.00	0.65	0.66
	2	385.88	58.81		1.00	0.78
	3	387.22	57.35			1.00

Table 39
Scale Score Correlations (Spearman Rho) between Content Standards

Mathematics

Grade	Content Standard	Mean	SD	1	2&3	4&5	6	7
3	1	427.55	86.48	1.00	0.65	0.69	0.68	0.62
	2&3	437.25	99.61		1.00	0.68	0.67	0.62
	4&5	430.68	89.75			1.00	0.71	0.66
	6	440.75	102.19				1.00	0.66
	7	413.36	65.40					1.00
4	1	421.73	86.73	1.00	0.72	0.75	0.72	0.74
	2&3	408.19	69.60		1.00	0.74	0.70	0.74
	4&5	412.37	78.43			1.00	0.72	0.77
	6	423.54	91.18				1.00	0.72
	7	399.32	60.59					1.00
5	1	413.98	65.66	1.00	0.75	0.76	0.78	0.82
	2&3	409.74	69.04		1.00	0.74	0.74	0.76
	4&5	426.86	88.67			1.00	0.76	0.78
	6	417.22	70.79				1.00	0.80
	7	411.66	55.26					1.00
6	1	404.02	68.19	1.00	0.75	0.75	0.78	0.83
	2&3	400.58	73.66		1.00	0.70	0.74	0.79
	4&5	405.32	72.12			1.00	0.74	0.75
	6	398.11	77.28				1.00	0.79
	7	400.50	57.86					1.00
7	1	392.49	82.07	1.00	0.80	0.81	0.82	0.83
	2&3	393.07	84.95		1.00	0.79	0.80	0.82
	4&5	399.94	74.60			1.00	0.81	0.85
	6	406.26	82.11				1.00	0.80
	7	394.49	55.71					1.00
8	1	407.06	69.91	1.00	0.76	0.79	0.78	0.88
	2&3	405.57	71.90		1.00	0.74	0.72	0.81
	4&5	404.28	58.61			1.00	0.73	0.82
	6	392.38	83.70				1.00	0.77
	7	400.85	54.33					1.00

Reading

Grade	Content Standard	Mean	SD	1	2	3
10	1	398.31	70.85	1.00	0.79	0.79
	2	385.88	58.81	0.79	1.00	0.83
	3	387.22	57.35	0.79	0.83	1.00

Factor analysis of the MSA Assessments

Exploratory factor analysis was used to examine the structure of the 2005 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 and 3 factors for the Grade 10 Reading test). 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 H26.

Each test form had between 8 and 10 initial eigenvalues greater than 1.0, with one dominant factor accounting for approximately 18 to 30 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 6.9 to 12.3 percent for the first factor, 5.9 to 12.2 percent for the second factor, 2.5 to 8.5 percent for the third factor, and 1.3 to 4.7 percent for the fourth and fifth factors. After extraction and rotation of 3 factors for the Grade 10 Reading test, 8.9 percent of the variance was explained by the first factor, 7.8 percent by the second factor, and 6.3 percent by the third 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. The one notable exception was in Grade 3, where a weak but fairly distinct factor did emerge on both test forms for Content Standard 7 (Process of Mathematics), accounting for approximately 2.5 percent of the variance on Form 1 (Factor 3) and 2.2 percent of the variance on Form 2 (Factor 4).

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 40 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 41 shows the percentages of students at each performance level on the 2005 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 2005 PAC for Mathematics showed a steady decline from grade to grade, dropping from approximately 77 percent in Grade 3 to approximately 52 percent in Grade 8. Tables 42 and 43 show the PAC classified by ethnicity and gender group. Tables 44 to 50 present the PAC by local education agencies (LEA) for each grade. Figures 2 to 8 show changes in the PAC between 2004 and 2005 for each LEA.

Table 40
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
10	240-373	374-414	415-650

Table 41
Percentages of Students at Each Performance Level

Grade Content	Form	N	Basic	Proficient	Advanced	Proficient +Advanced
MA3	1	36935	23.17	51.04	25.79	76.83
	2	24574	22.78	51.80	25.42	77.22
	Total	61509	23.01	51.35	25.64	76.99
MA4	1	38004	24.11	48.85	27.04	75.89
	2	25326	22.54	50.66	26.80	77.46
	Total	63330	23.48	49.57	26.95	76.52
MA5	1	39109	31.04	51.36	17.60	68.96
	2	26014	30.37	52.93	16.70	69.63
	Total	65123	30.77	51.99	17.24	69.23
MA6	1	39509	39.45	45.12	15.43	60.55
	2	26337	40.05	45.59	14.36	59.95
	Total	65846	39.69	45.31	15.00	60.31
MA7	1	40930	44.19	41.86	13.95	55.81
	2	27200	44.58	41.64	13.78	55.42
	Total	68130	44.35	41.77	13.88	55.65
MA8	1	34478	48.29	32.27	19.44	51.71
	2	34218	47.34	34.11	18.55	52.66
	Total	68696	47.82	33.19	18.99	52.18
RD10	A	6934	32.68	38.04	29.28	67.32

Table 42

Percentages of Students at Each Performance Level by Ethnicity

Grade Content	Ethnicity	N	Basic	Proficient	Advanced	Proficient +Advanced
MA3	White	30221	13.15	51.24	35.61	86.85
	African American	22963	36.17	52.06	11.77	63.83
	Hispanic	4842	30.94	54.96	14.11	69.06
	Others	3483	10.85	42.55	46.60	89.15
MA4	White	31130	12.68	50.13	37.20	87.32
	African American	24030	37.90	50.07	12.03	62.10
	Hispanic	4660	31.70	51.37	16.93	68.30
	Others	3510	9.66	38.89	51.45	90.34
MA5	White	31636	18.07	56.55	25.38	81.93
	African American	25405	47.05	47.46	5.49	52.95
	Hispanic	4601	42.21	49.16	8.63	57.79
	Others	3481	12.24	47.37	40.39	87.76
MA6	White	31910	25.99	51.72	22.29	74.01
	African American	26151	57.29	38.03	4.68	42.71
	Hispanic	4365	51.09	42.20	6.71	48.91
	Others	3420	18.30	45.12	36.58	81.70
MA7	White	33528	27.61	51.12	21.28	72.39
	African American	26831	66.69	30.25	3.06	33.31
	Hispanic	4304	55.65	38.52	5.83	44.35
	Others	3467	19.30	44.62	36.08	80.70
MA8	White	34489	32.56	39.46	27.98	67.44
	African American	26481	69.13	25.35	5.51	30.87
	Hispanic	4253	59.77	31.72	8.51	40.23
	Others	3473	22.23	32.39	45.38	77.77
RD10	White	5248	24.71	39.71	35.58	75.29
	African American	1420	61.48	32.32	6.20	38.52
	Hispanic	158	43.04	34.18	22.78	56.96
	Others	108	25.93	37.96	36.11	74.07

Table 43
Percentages of Students at Each Performance Level by Gender

Grade Content	Gender	N	Basic	Proficient	Advanced	Proficient +Advanced
MA3	Male	31387	24.10	50.29	25.62	75.90
	Female	30115	21.88	52.46	25.67	78.12
MA4	Male	32449	24.95	47.71	27.34	75.05
	Female	30873	21.94	51.52	26.54	78.07
MA5	Male	33616	31.73	50.18	18.08	68.27
	Female	31501	29.74	53.92	16.34	70.26
MA6	Male	33855	41.85	42.70	15.45	58.15
	Female	31969	37.36	48.09	14.55	62.64
MA7	Male	35126	46.14	39.58	14.28	53.86
	Female	32983	42.40	44.13	13.46	57.60
MA8	Male	35063	50.31	30.84	18.85	49.69
	Female	33626	45.22	35.64	19.15	54.79
RD10	Male	3526	37.78	36.67	25.55	62.22
	Female	3408	27.41	39.47	33.13	72.59

Table 44
Percentages of Students at Grade 3 Performance Levels by LEA

LEA #	N	Basic	Proficient	Advanced	Proficient +Advanced
1	662	28.55	52.72	18.73	71.45
2	5270	14.36	52.79	32.85	85.64
3	7570	21.70	54.45	23.84	78.30
4	1236	11.49	49.27	39.24	88.51
5	348	26.72	53.16	20.11	73.28
6	2062	13.43	53.69	32.88	86.57
7	1147	20.49	61.73	17.79	79.51
8	1801	22.99	50.53	26.49	77.01
9	307	33.55	56.68	9.77	66.45
10	2946	18.57	55.60	25.83	81.43
11	330	18.79	60.91	20.30	81.21
12	2946	16.23	55.50	28.28	83.77
13	3560	13.37	45.59	41.04	86.63
14	151	8.61	44.37	47.02	91.39
15	9986	16.69	45.77	37.53	83.31
16	9608	34.72	51.24	14.04	65.28
17	567	21.34	55.03	23.63	78.66
18	1152	20.92	52.60	26.48	79.08
19	209	26.79	60.77	12.44	73.21
20	297	20.88	56.90	22.22	79.12
21	1535	18.70	54.14	27.17	81.30
22	1058	19.94	57.37	22.68	80.06
23	427	15.46	46.84	37.70	84.54
30	5971	42.92	48.99	8.09	57.08
31	310	33.23	54.84	11.94	66.77
55	51	23.53	60.78	15.69	76.47

Table 45
Percentages of Students at Grade 4 Performance Levels by LEA

LEA #	N	Basic	Proficient	Advanced	Proficient +Advanced
1	657	23.29	54.19	22.53	76.71
2	5463	13.84	49.51	36.65	86.16
3	7887	22.72	52.99	24.29	77.28
4	1287	13.13	48.72	38.15	86.87
5	374	20.32	52.67	27.01	79.68
6	2101	12.57	54.88	32.56	87.43
7	1184	21.88	58.28	19.85	78.13
8	1831	23.38	52.43	24.19	76.62
9	300	40.33	44.00	15.67	59.67
10	3008	17.32	53.89	28.79	82.68
11	360	26.11	53.89	20.00	73.89
12	3000	16.60	55.37	28.03	83.40
13	3808	11.56	43.70	44.75	88.45
14	160	17.50	45.00	37.50	82.50
15	10174	16.45	45.38	38.17	83.55
16	9860	35.92	49.03	15.05	64.08
17	523	17.59	56.02	26.39	82.41
18	1168	22.43	51.63	25.94	77.57
19	192	29.69	54.17	16.15	70.31
20	323	24.46	49.23	26.32	75.54
21	1487	15.06	51.38	33.56	84.94
22	1042	19.77	50.77	29.46	80.23
23	428	14.95	48.13	36.92	85.05
30	6323	45.94	45.67	8.38	54.06
31	327	42.81	48.01	9.17	57.19
55	57	38.60	49.12	12.28	61.40

Table 46
Percentages of Students at Grade 5 Performance Levels by LEA

LEA #	N	Basic	Proficient	Advanced	Proficient +Advanced
1	681	34.36	52.28	13.36	65.64
2	5528	19.10	58.16	22.74	80.90
3	7822	29.88	53.75	16.38	70.12
4	1297	22.21	56.28	21.51	77.79
5	392	27.30	62.76	9.95	72.70
6	2186	18.30	62.26	19.44	81.70
7	1237	24.01	62.41	13.58	75.99
8	1946	31.55	52.98	15.47	68.45
9	335	42.09	52.54	5.37	57.91
10	2985	22.21	56.88	20.90	77.79
11	349	32.09	57.02	10.89	67.91
12	3140	24.33	58.25	17.42	75.67
13	3706	14.52	54.43	31.06	85.48
14	176	38.07	53.41	8.52	61.93
15	10122	21.44	47.85	30.72	78.56
16	10611	48.57	45.36	6.07	51.43
17	565	18.76	59.12	22.12	81.24
18	1272	25.31	56.76	17.92	74.69
19	245	39.18	50.61	10.20	60.82
20	311	23.15	61.41	15.43	76.85
21	1525	27.15	55.74	17.11	72.85
22	1141	33.30	52.23	14.46	66.70
23	484	20.87	59.50	19.63	79.13
30	6666	51.20	44.64	4.16	48.80
31	327	46.48	49.85	3.67	53.52
55	66	50.00	43.94	6.06	50.00

Table 47
Percentages of Students at Grade 6 Performance Levels by LEA

LEA #	N	Basic	Proficient	Advanced	Proficient +Advanced
1	744	46.24	45.70	8.06	53.76
2	5500	31.67	45.67	22.65	68.33
3	8123	41.16	45.13	13.71	58.85
4	1372	33.89	53.43	12.68	66.11
5	390	34.10	52.56	13.33	65.90
6	2253	28.98	51.71	19.31	71.02
7	1342	37.03	49.25	13.71	62.97
8	2005	37.46	50.32	12.22	62.54
9	343	63.27	33.53	3.21	36.73
10	3030	26.67	53.10	20.23	73.33
11	420	39.52	50.95	9.52	60.48
12	2974	35.74	51.35	12.91	64.26
13	3882	20.09	51.31	28.59	79.91
14	187	51.34	35.29	13.37	48.66
15	10145	32.12	45.13	22.75	67.88
16	10610	46.85	45.25	7.90	53.15
17	578	28.03	57.61	14.36	71.97
18	1181	36.16	46.91	16.93	63.84
19	239	48.54	44.77	6.69	51.46
20	369	39.30	46.07	14.63	60.70
21	1547	26.63	53.72	19.65	73.37
22	1070	48.60	40.75	10.65	51.40
23	480	25.00	49.17	25.83	75.00
30	6753	71.07	27.08	1.85	28.94
31	248	42.34	50.40	7.26	57.66
55	53	62.26	32.08	5.66	37.74

Table 48
Percentages of Students at Grade 7 Performance Levels by LEA

LEA #	N	Basic	Proficient	Advanced	Proficient +Advanced
1	743	42.93	46.16	10.90	57.07
2	5871	32.24	47.06	20.69	67.76
3	8440	41.87	44.00	14.12	58.13
4	1381	38.52	49.89	11.59	61.48
5	422	49.05	43.13	7.82	50.95
6	2323	34.70	50.71	14.59	65.30
7	1316	34.80	50.84	14.36	65.20
8	2016	47.37	42.51	10.12	52.63
9	366	67.49	28.14	4.37	32.51
10	3138	28.59	51.47	19.95	71.41
11	361	32.96	57.34	9.70	67.04
12	3140	40.92	46.69	12.39	59.08
13	3852	21.31	50.42	28.27	78.69
14	195	55.90	36.41	7.69	44.10
15	10515	32.25	44.53	23.22	67.75
16	11064	60.01	34.99	5.00	39.99
17	602	30.23	55.32	14.45	69.77
18	1247	44.43	44.59	10.99	55.57
19	259	56.37	39.38	4.25	43.63
20	346	44.51	47.40	8.09	55.49
21	1628	25.68	52.95	21.38	74.32
22	1116	48.12	44.27	7.62	51.88
23	536	26.87	55.04	18.10	73.13
30	7160	80.98	17.88	1.15	19.02
31	1	100.00	.	.	.
55	77	67.53	27.27	5.19	32.47

Table 49
Percentages of Students at Grade 8 Performance Levels by LEA

LEA #	N	Basic	Proficient	Advanced	Proficient +Advanced
1	772	47.93	39.38	12.69	52.07
2	5790	34.28	38.89	26.82	65.72
3	8739	46.70	33.14	20.16	53.30
4	1439	46.84	33.01	20.15	53.16
5	468	48.08	34.19	17.74	51.92
6	2417	37.86	41.95	20.19	62.14
7	1323	39.91	43.16	16.93	60.09
8	2115	53.66	31.73	14.61	46.34
9	362	64.92	30.66	4.42	35.08
10	3163	32.88	40.91	26.21	67.12
11	410	39.51	35.12	25.37	60.49
12	3236	46.29	37.14	16.56	53.71
13	3937	27.08	41.12	31.80	72.92
14	198	47.48	36.36	16.16	52.53
15	10621	35.32	35.08	29.60	64.68
16	11035	64.03	27.74	8.23	35.97
17	653	39.82	38.59	21.59	60.18
18	1241	54.07	32.07	13.86	45.93
19	218	58.72	31.19	10.09	41.28
20	378	60.58	26.72	12.70	39.42
21	1668	27.82	41.43	30.76	72.18
22	1073	60.11	30.57	9.32	39.89
23	534	21.54	36.33	42.13	78.46
30	6789	79.72	17.40	2.89	20.28
55	101	85.15	12.87	1.98	14.85

Table 50
Percentages of Students at Grade 10 Performance Levels by LEA

LEA #	N	Basic	Proficient	Advanced	Proficient +Advanced
2	45	73.33	13.33	13.33	26.67
3	570	30.18	36.49	33.33	69.82
5	274	46.35	36.50	17.15	53.65
6	1051	22.84	38.06	39.11	77.16
7	481	35.14	39.92	24.95	64.86
9	226	40.71	34.96	24.34	59.29
10	1606	28.71	34.99	36.30	71.30
11	93	19.36	52.69	27.96	80.65
12	121	32.23	48.76	19.01	67.77
14	85	43.53	41.18	15.29	56.47
15	10	70.00	20.00	10.00	30.00
17	287	29.62	38.68	31.71	70.38
19	111	35.14	51.35	13.51	64.86
20	193	25.91	46.63	27.46	74.09
21	836	21.89	45.45	32.66	78.11
22	4	100.00	.	.	.
23	266	25.56	36.47	37.97	74.44
30	646	65.02	31.58	3.41	34.98
55	2	50.00	50.00	.	50.00

Figure 2

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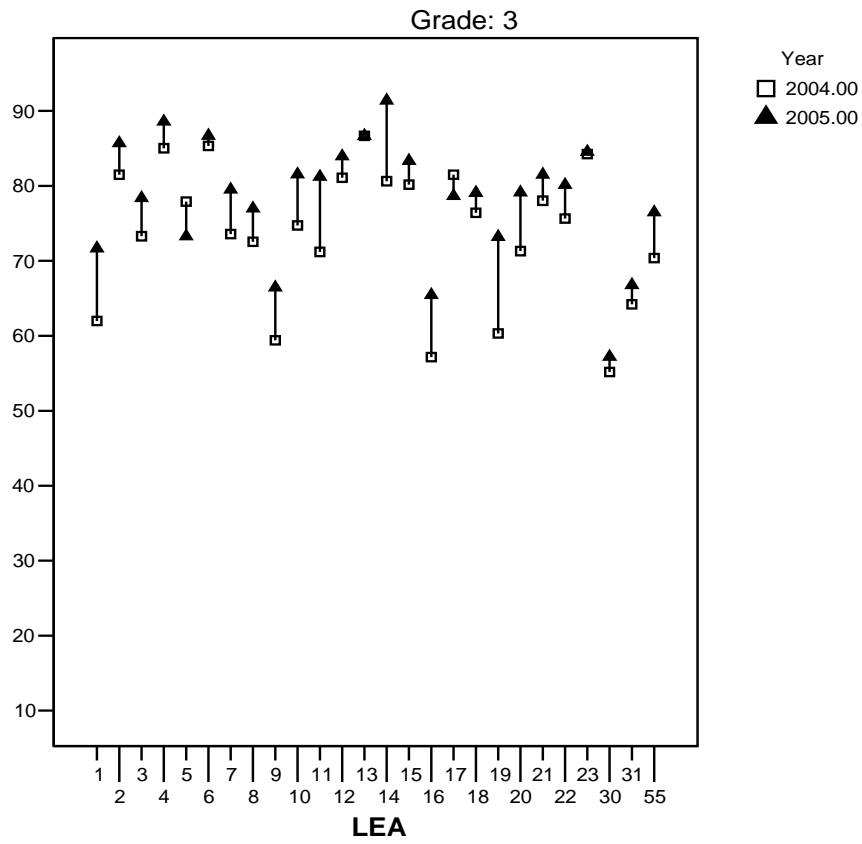
Figure 3

Percent at or Above Proficiency Cut Score (PAC) by LEA for Mathematics Grade 4
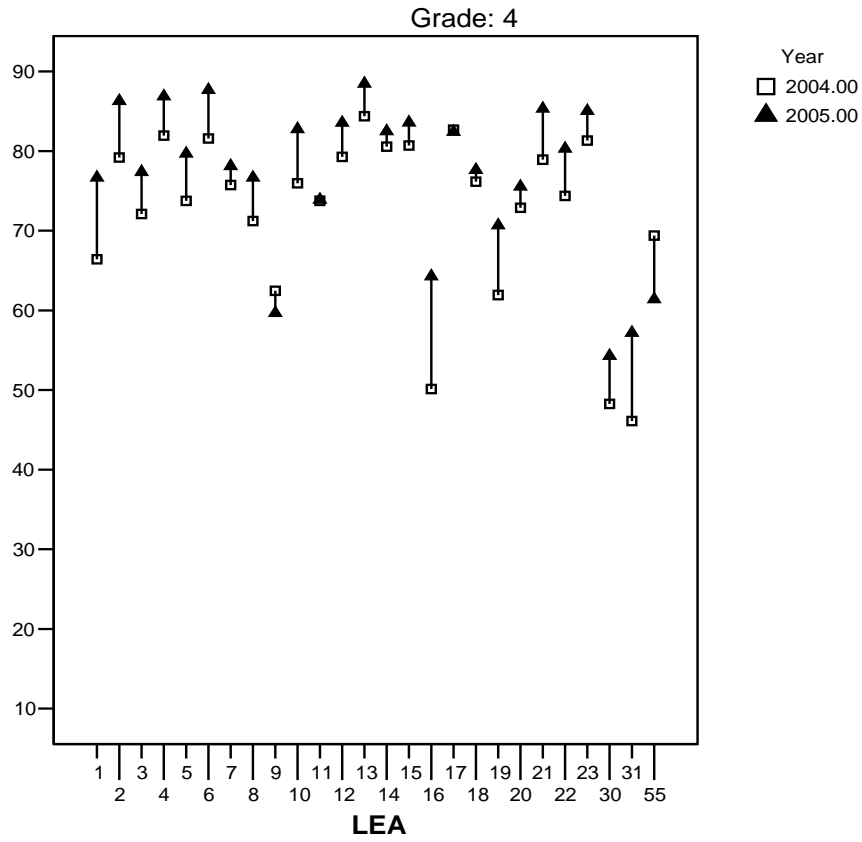
Figure 4
Percent at or Above Proficiency Cut Score (PAC) by LEA for Mathematics Grade 5
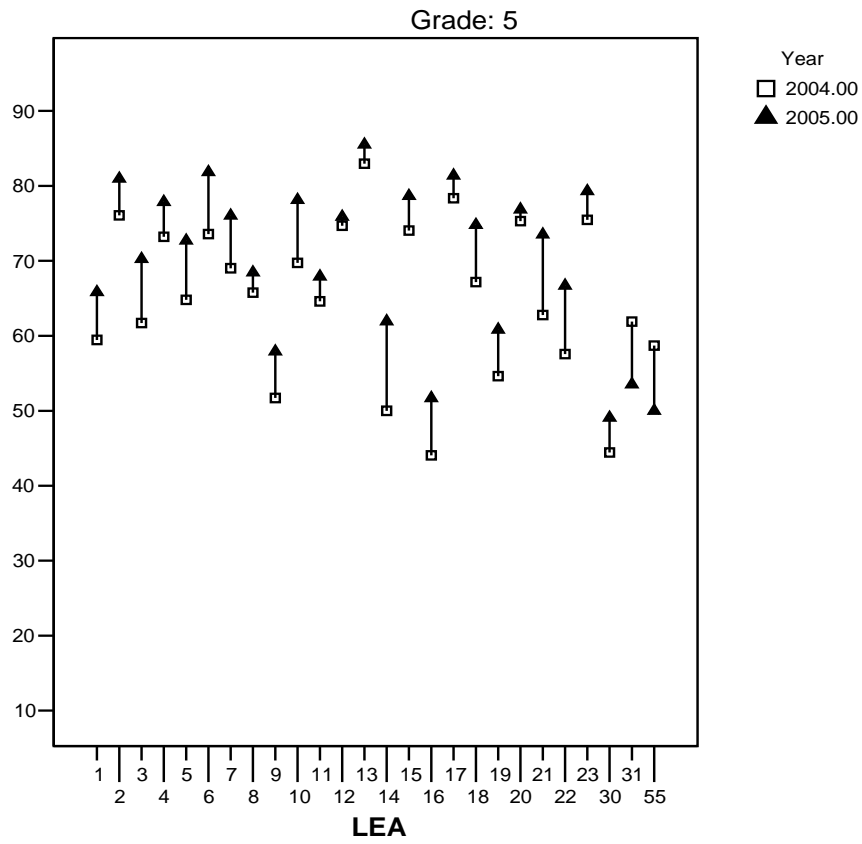
Figure 5

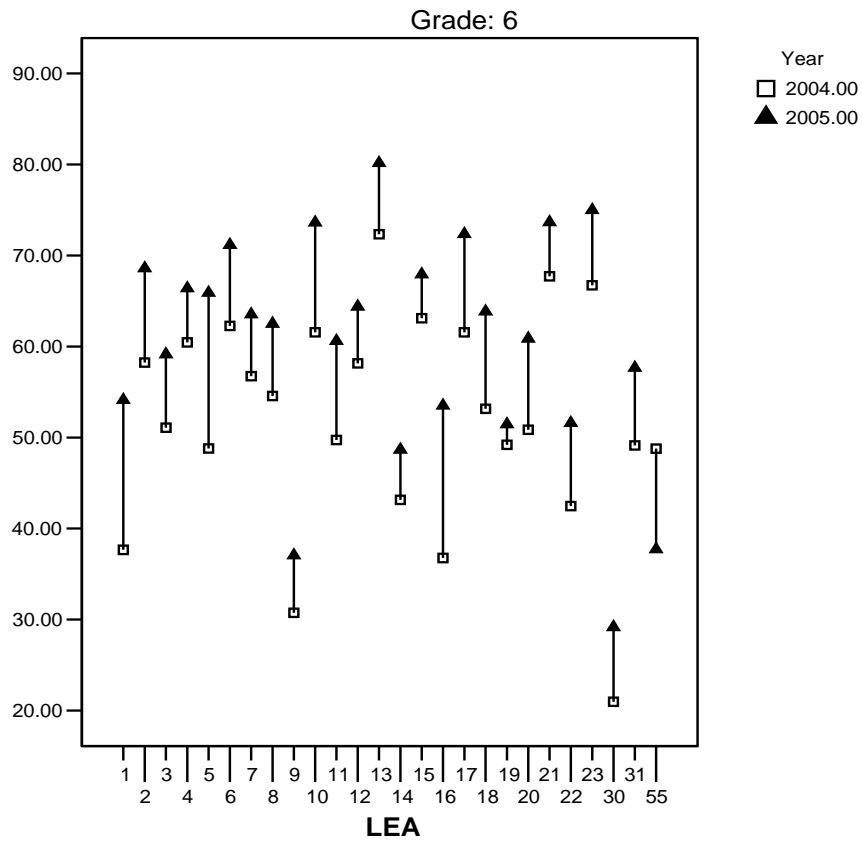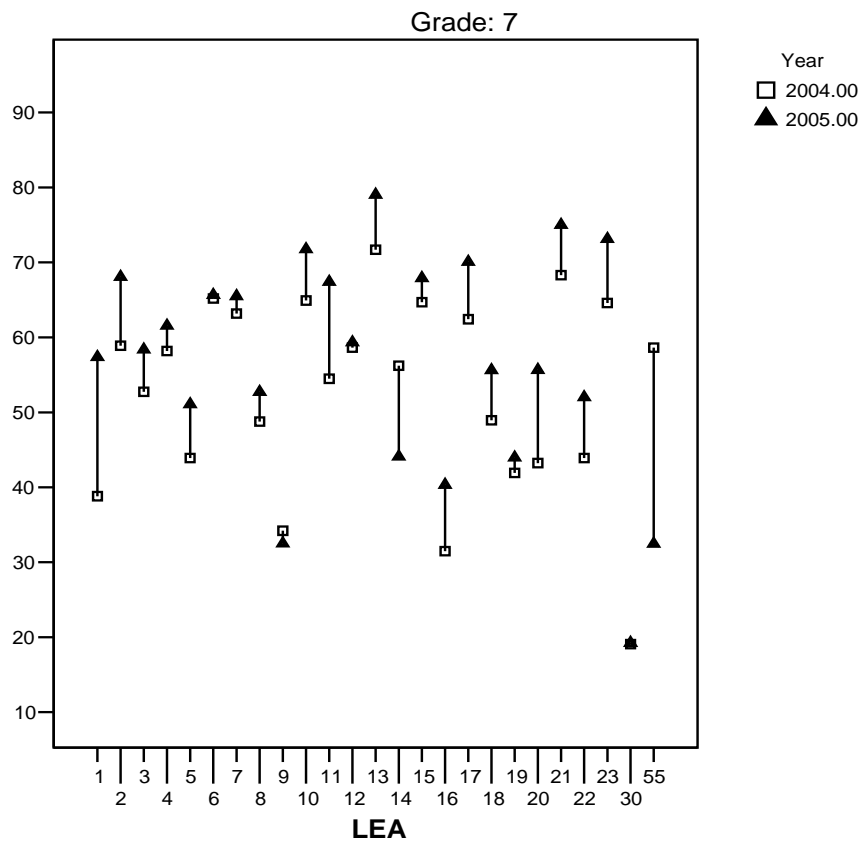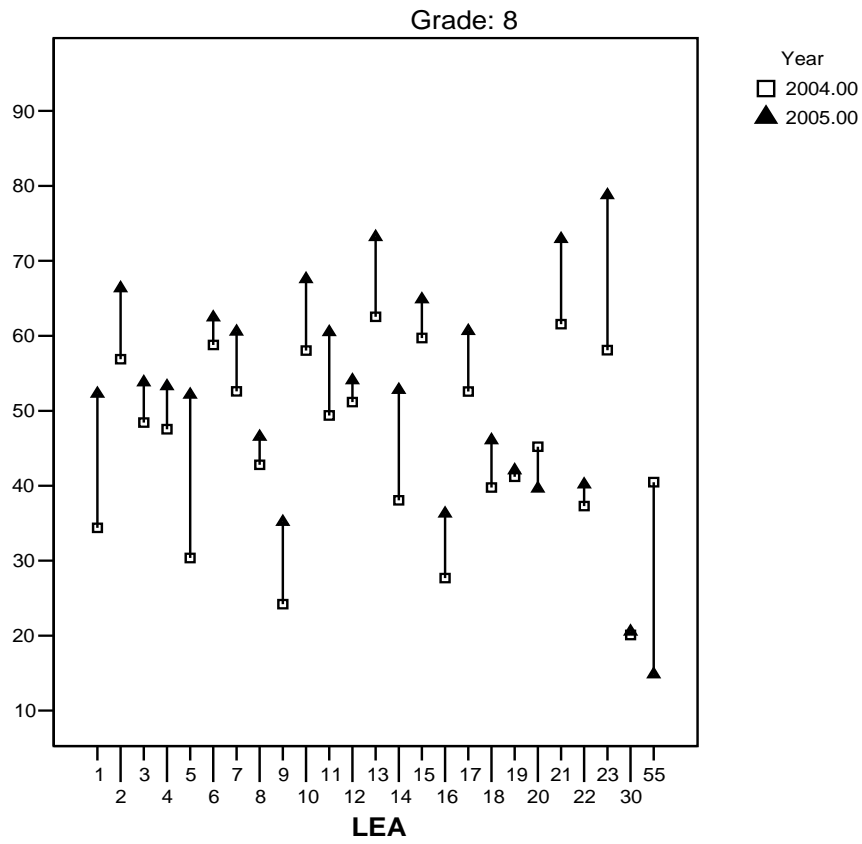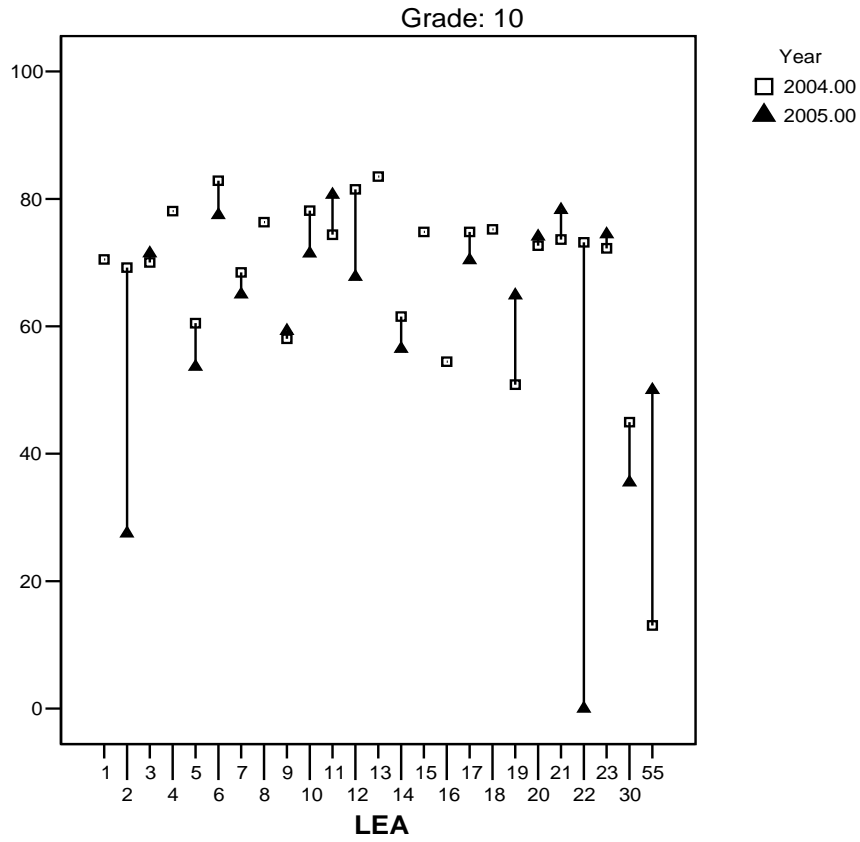Percent at or Above Proficiency Cut Score (PAC) by LEA for Mathematics Grade 6

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

Figure 8
Percent at or Above Proficiency Cut Score (PAC) by LEA for Reading Grade 10

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.

Harry, H. Harman, (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.