Maryland High School Assessment

2005

Scoring Contractor's Report

Measurement Incorporated
423 Morris St
Durham NC 27701

Introduction

As Scoring Contractor for the Maryland High School Assessment (HSA), Measurement Incorporated (MI), in collaboration with the Maryland State Department of Education (MSDE), conducts and coordinates all scoring activities related to the HSA. This includes:

- Retrieving all HSA materials from Maryland schools and LEA central offices
- Checking in materials and producing security reports
- Scanning of used answer books to capture student demographic information, bubbled selected responses (SR), and student-produced responses (SPR), and to prepare the answer books for handscoring of constructed response (CR) items
- Conducting rangefinding sessions for handscoring of CR items
- Preparing training materials for handscoring of CR items
- Handscoring of CR items, both operational and field test
- Providing student demographic information, SR and SPR responses, and raw CR score data in data files for MSDE and the Development Contractor
- Storing all test materials securely
- Preparing annotated item guides and training sets for public release on the MSDE web site
- Providing written reviews of CR items prior to field testing and written evaluations of field test CR items after test administration and scoring

Five HSA content areas were tested and scored in 2005: Algebra/Data Analysis, Biology, English, Geometry, and Government. (Geometry is a Maryland School Assessment content area test that was administered during the same test administrations as the HSA and scored by MI. For the purposes of this report, the term HSA will be used to refer collectively to all five content areas' assessments.)

The 2005 Maryland High School Assessment consisted of three test administrations: January, May, and summer. Each administration consisted of multiple forms, each form with multiple constructed and selected response items for each of the content areas assessed. Algebra/Data Analysis and Geometry

also included student produced response, or “gridded,” items for which the student provided numeric answers to the SPR questions.

Algebra/Data Analysis, Biology, Geometry and Government were each administered in January, May, and summer. English was administered in May and summer only.

In the January and May administrations of the Algebra/Data Analysis, Biology, Geometry and Government Maryland High School Assessments, there were both operational and field test items. The operational items, on which students’ scores were based, were identical in each form within a content area. These items had been field tested in prior administrations and were selected for use as operational items by MSDE. Field test items, which were items being administered for the first time in 2005, were unique in each form. Scores assigned to these field test items did not affect the students’ scores.

For the January test administrations in 2005, handscoring of field test constructed response items for each of these four content areas followed immediately after operational scoring was completed for that content area. For the May test administration, field test scoring took place in the fall per MSDE request, after operational scoring of the May and summer administrations had been completed.

Student responses to operational SR and SPR items were reported concurrently with student responses to field test SR and SPR items in score files posted for MSDE and the Development Contractor.

The new English HSA administered in May and summer consisted of multiple forms each containing unique test items. At MSDE direction, all English items were treated as operational items in MI’s scoring and reporting activities.

The scoring of the summer test administration for all content areas did not include a field test component.

A summary of the procedures followed for each administration follows. Please see attachments A - I for more information about each 2005 administration.

The following procedures were followed for each of the three test administrations in 2005:

Pick-up, Processing, and Scanning of Test Materials

Pick-up of Materials

MI arranged for a Maryland-based courier service to handle the retrieval of the HSA test booklets and at MSDE direction established a schedule for the four separate pick-ups that were needed for the January and May test administrations. A schedule of pick-up dates was implemented for each test administration: once during primary week, once during make-up week #1, once during make-up week #2, and once after the test administration was completed. Pick-ups for LEA 24 schools (private institutions, state-operated programs, or non-public educational centers serving students who have been placed by the Maryland public schools) and LEA central offices were limited to the fourth pick-up only.

For the summer administration, with a small number of participating schools and students, MI, at MSDE direction, provided schools with Fed Ex airbills and instructions for returning the materials to MI. In the future, larger summer administrations will require pick-ups at the schools.

Prior to each test administration in January and May, MSDE provided a list of schools (including LEA 24 schools and LAC central offices) and addresses to MI. The Development Contractor provided MI with shipping and tracking data for orders shipped to schools and LEA central offices. MI used this data to create a list of schools participating in the HSA. The Development Contractor supplied updates to the shipping and tracking data as new orders were filled.

To improve the pick-up process and to help avoid late deliveries of materials, MI provided, prior to pickups, a list to MSDE of all schools in the tracking data from the Development Contractor. MSDE supplied this information to the LACs, who confirmed participating schools and their addresses. LACs also provided notification of any schools that required pick-ups and were not on the current list. Also after this initial notification, since MI depends on shipping and tracking data supplied by the Development Contractor to identify schools requiring pick-ups, it was important for LACs to provide notification to MI of any schools to which they had distributed materials but that were not in the tracking data.

The Development Contractor supplied return shipment labels and included them with the shipments of the HSA materials sent to the schools. The MI Durham address printed on these labels helped to ensure that a definite destination was associated with each box of materials returned.

MI printed pick-up tickets on three-part form paper for each school and provided them to the courier. At each pick-up, the courier left one copy of the completed and signed ticket for the school to retain. The courier kept one copy for his

records, and one copy accompanied the materials shipped to MI. The pick-up tickets showed the number of boxes picked up from each school.

At the end of each day's pick-up of materials, the materials were transferred directly to the courier's central facility in Maryland. Upon receipt at the courier's central facility, the materials once again were verified against the database as to quantity of boxes and schools. After all of the pick-ups were completed for that day, the materials were packed for shipping directly to MI's main offices. MI arranged for a national shipping company to provide this service. In order to maintain tight security of the materials involved, this shipper was required to take the test materials from the courier's central facility directly to the MI receiving center in Durham. No stops were made for any other pick-ups or at any warehouses operated by the contracted carrier. As an extra security measure, no other materials were on the delivery vehicle(s), except for the HSA materials being delivered directly to Durham. The materials were sent directly to MI Durham overnight and arrived there the next day with pallet manifests and copies of all pick-up tickets from that pick-up date.

Schools and LACs contacted MI's toll-free HSA call center with any questions or problems involving pick-up of materials. MI maintains a toll-free number for schools and LEAs to contact regarding all pick-up issues. Operators at the MI Call Center have access to a database containing pick-up information for each school and communicate problems and inquiries to the MI Shipping Department and to Project Management. The operators are also able to initiate the printing of additional pick-up tickets for the courier as needed. MI is currently developing additional software applications to automate as many of these processes as possible.

After each scheduled pick-up day, the courier provided to MI, and MI to MSDE, a spreadsheet showing the total number of boxes retrieved from each school, including schools that did not turn over any materials to the courier and schools that did not receive a pick-up for any reason.

For the summer administration, schools sent materials directly to MI via Fed Ex. MI monitored Fed Ex tracking in order to ensure that all boxes sent from the schools were received by MI.

Check-in and Processing of Materials

Checking-in and processing test materials involved three major tasks:

- Initial checking-in of boxes to ensure that all boxes picked up by the courier on a particular date were present and accounted for in the delivery to MI

- Checking-in secure materials: test books, Algebra/Data Analysis answer books, and Geometry answer books

- Processing all used answer books for scoring

Initial Check-In of Boxes

Upon arrival in Durham, the truck containing pallets of boxes from the previous day's pickups was unloaded. First, the boxes on each pallet were counted to ensure that the quantity received matched the quantity recorded on the pallet manifest. Second, the boxes were sorted by school and by content area. Boxes for each school were counted to ensure that the quantity matched the quantity recorded on the pick-up ticket for that school. These two steps ensured that all boxes picked up the previous day had been delivered to Durham.

When these steps were accomplished, unpacking of the boxes began. The next tasks were security check-in and processing of used answer books for scoring, which took place concurrently.

Security Check-In of Test Books and Secure Answer Books

MI is responsible for the security and tracking of all test materials shipped from the schools and LEA central offices to MI. MI checked these returned test materials against the shipping and tracking data supplied by the Development Contractor. This data provided the LEA/school associated with the unique security barcode printed on each secure book shipped to the LEAs and schools. MI also used any TACM "Appendix A: Redistribution of Materials Record" documents included in the schools' returned test materials. These documents indicated the transfer of materials among LEAs and schools, helping to provide a final association with a LEA and/or school for each secure book.

Upon receipt of the test materials in Durham, a high priority was to match the security codes of returned material with the material originally shipped. This was accomplished in two ways. The security bar codes on the test books for all subject areas and the security bar codes on sealed, shrink-wrapped packages of unused mathematics answer documents (Algebra/Data Analysis and Geometry answer documents are the only content areas with security barcodes on the answer documents) were scanned by warehouse personnel using a handheld barcode scanner. In this report, this process will be called security scanning. All "loose" (i.e. not sealed in shrink-wrapped packages) Algebra/Data Analysis and Geometry answer documents had their individual security barcodes read by a barcode reader on NCS OPSCAN 21 Optical Mark Readers (OMRs). In this report this process will be called OMR scanning. (See "OMR Scanning to Capture Demographic Information and SR and SPR Responses" below.) The security bar code data from both scanning processes was downloaded into specific programs to be compared with the data from the original shipment of materials to schools and LEA central offices by the Development Contractor.

MI followed a number of quality control, back up, and identification procedures during the security check-in process. Each box opened in the MI warehouse was assigned and tagged with a unique school and data bar code identifier, so that if any questions about bar codes arose, the original documents could be located quickly. Also, all scanner files were backed up daily, or more frequently, depending on the volume being processed.

Processing of answer books for scoring

Header Sheets

At the school, test coordinators were to place a header, provided by the Development Contractor, in each box containing used answer books for a particular content area. This header provided not only a school and LEA association for the used answer books in each box, but also the quantity of used answer books in the box for that content area. Since processing of the answer books for scoring required that the forms within each content area were separated, multiple headers for each content in each box were required. When a box was opened at MI, additional headers were generated for each form to match the LEA and school number on the school header for each box.

The school header sheets were then set aside for later OMR scanning. When the school header sheets were scanned, they were compared programmatically to the MI header sheets that had been generated. This ensured that MI had the correct LEA and school association for each MI header and that the correct numbers of answer books were scanned under MI headers for each school header.

After MI header sheets were generated for each form of the content in the box, the answer books were separated, and each form was placed in a scan bin designated for that content and form, with the appropriate MI header on top of each group of books from that box. Each scan bin was labeled with the content, form, and a unique internal tracking number. The number of answer books in each scan bin varied, as a scan bin contained all of the same content and form. Headers with LEA and school information separated the schools in each bin.

Daily, or more often as required, the scan bins of answer documents were transferred to MI's scanning operations area for further processing.

Student ID Labels

At the school level, pre-printed student ID labels were affixed to the student answer books of those students included in the school's pretest file. Each affixed label identified the student associated with that answer book. For students not included in the pretest file, school personnel applied generic ID labels (with

school and LEA information and a unique barcode number) to those students' answer books and bubbled in any required information on each student's demographic page. If the school did not have a pre-printed ID label or school generic label to affix, school personnel bubbled in any required information for the student, as well as the LEA code number and school code number, on the demographic page. MI attached a MI generic ID label to answer books that did not have an affixed pre-print or school generic ID label. The MI generic label contained a unique barcode number that was then linked to the answer book in the data.

During scanning, any misread or missing ID label or security barcode information was corrected at the point of scanning. If the scanner created an error message for either the student barcode or the security barcode, the book was rescanned. If needed, a MI generic label was applied.

OMR Scanning to Capture Demographic Information and SR and SPR Responses

In 2005, all "loose" (i.e. not sealed in shrink-wrapped packages) answer books for all contents were prepared for OMR scanning, regardless of whether the schools had designated them as scorable, invalidated, or unused. MI headers and scan bin labels noted these school-level designations for each group of answer books. This procedure helped to ensure that all "loose" answer books were scanned to capture any demographic information and SR and SPR responses. This allowed review of scan files to attempt to identify scorable or invalidated answer books that had been mistakenly designated by the schools as "unused."

The stack of loose answer books was removed from each bin and machine-cut to remove the spines of the books in order to separate the pages for scanning. Books were then put back in the bin with the MI headers and brought to the scanners.

The pages of answer books and the accompanying MI headers in each bin were scanned on an NCS OPSCAN 21 Optical Mark Reader (OMR). Information was captured from the pre-printed or school-generic student ID label, security barcode on math answer books, bubbled information on the cover page, lithocode on each page of the answer book, and SR and SPR bubbles on each page of the answer book. Additionally, the scanner read sheet marks and timing marks to identify each page and to locate scannable areas on each page.

Header information included LEA, school, quantity of answer books, content, and the designations "scorable," "invalidated," or "unused." Comparison of the header information with the data in scan files helped to ensure that this information was correct for each scanned book under each header.

A computer program that instructed the scanner where the bar codes, lithocodes, and bubbled information areas should be located on each answer sheet controlled collection of the information from answer books. Each of these areas was mapped to a definition that specified what data was valid for that area of the answer sheet. The information recorded from a single answer book, including student identification information, lithocodes, and SR and SPR responses, was represented as a line of data, or record, in a text file. Additional information, such as the identity of the scanning program, the batch number, a unique sequence number for each document scanned, and the scanning date assigned during the scanning process also was present in each record. Each line of data in the file represented a different answer book, and each file contained information only from the books of a single batch, which represented a physical scan bin.

A unique batch number was assigned programmatically to each bin as its contents were scanned. Also during scanning, an endorsement string was printed along the margin of each page. This string included the batch number, the sequence of the answer book within the batch, and the date scanned. This then served as an additional unique ID for each scanned document.

For purposes of security check-in, the security barcodes on loose Algebra/Data Analysis and Geometry answer books were captured during scanning. These barcodes went into a table, were validated programmatically (i.e. confirmation made that the security barcode numbers were valid and in the correct range provided to MI by the Development Contractor), and then added to the security check-in database as “received.”

Cover Pages and Demographic Data

Cover pages of the answer books (with student demographic information) were scanned at the same time as the other answer book pages. During the OMR scanning process, these cover pages were then automatically shunted from the other answer book pages. The student demographic data associated with the cover page was thus included in the scanned data and linked with data captured from the pages of the answer book and from the MI header sheet associated with that answer book.

Immediately after scanning, MI Data Validation Rules (see Attachment J) were applied programmatically to the scanned data in the file. Demographic data that failed any of the edit rules were manually reviewed. A list of books failing edits was generated programmatically for MI’s Data Entry Department. For each batch, the cover sheets were reviewed in an attempt to correct the information. At this point, any edits were done via double-data entry, with resolution edits for any discrepancies between the first and second data entry. Any invalid or missing (i.e. written on cover, but not bubbled in or bubbled inaccurately) scannable demographic information was corrected via data entry.

Demographic cover pages were then filed by scan batch number and stored securely at MI's Durham warehouse. These pages were not sent to the scoring locations, helping to ensure the anonymity of the students and objective scoring by Scorers.

Invalidated Answer Books

There are two ways that an answer book was invalidated. One occurred at the school level when a teacher invalidated a student's answer book. The second happened when MI applied MSDE rules to invalidate a primary week answer book.

2005 Administration	Answer Books with School Invalidation	Answer Books with MI Invalidation
January05	6,140	447
May05	32,437	1,130
Summer05	419	25

At the school level, answer books were invalidated in one of two ways. If the number of answer books to be invalidated was less than 50 total, each book had a designated section of bubbles on the cover page blackened. If the number of answer books to be invalidated was more than 50, blackening of this section of the answer book cover was not required. Instead, these books were to be securely paper banded and clearly labeled "Invalidated: Do not Score" or banded with red paper bands supplied by the Development Contractor for that purpose.

MI invalidated answer books under two conditions specified by MSDE:

> If MI identified a duplicate student barcode ID that was associated with both a blank primary week answer book and a make-up week answer book that contained responses in the data file, then MI invalidated the blank primary week answer book.
>
> MI invalidated any answer books that had student or school labels or bubbled demographic information, but that were otherwise completely blank (i.e. no student response to any SR, SPR, or CR items).

MI provided a file directly to MSDE containing all invalidated records. Invalidated records were not included in the score files posted for the Development Contractor.

Capturing SR and SPR Item Responses

During OMR scanning, the Optical Mark Reader read and recorded students' bubbled responses to SR and SPR items. The responses were included in the Final Score Files and Field Test Score Files sent to the Development Contractor.

MI reported the student responses to SR and SPR items, but did not score, or apply a key to, SR or SPR items.

After the answer books had been packeted and shipped to the scoring locations, a random sample of answer books for each form and content was selected programmatically. A list of these answer books, along with the SR and SPR responses from the scan file, were provided to management at the scoring locations. The answer books were then visually inspected to ensure that the OMR scanner and its programs were capturing and reporting SR and SPR responses correctly for each content and form.

Packeting Answer Books for Handscoring

Immediately after scanning, the scanned answer book pages, except for the cover pages, were put back in their bins and moved to another area of the warehouse to be manually packeted for handscoring of CR items.

During scanning, booklets were programmatically divided into groups of six or fewer per packet. Packeting materials were generated programmatically and printed for each batch of answer books. These materials included a batch inventory sheet and packet inventory sheets, both of which listed the lithocodes of the answer books in a given packet and showed the year, administration, content, form, batch, and a unique packet number. Scannable handscoring monitor sheets, on which Scorers would later record CR scores, were also generated at the same time.

Answer books in each batch, along with scoring monitors and a packet inventory sheet, were manually inserted into packet envelopes. Packets were packed in boxes that were imprinted with MI's name, logo, and address. Each packet envelope was also imprinted with this information. Boxes were shipped to MI scoring locations via a common carrier with full tracking capabilities. A computer generated packing list listed the packets in each box.

All shipped packets were listed on an "off site shipping folder" that could be accessed from the MI network by any of the MI scoring locations and by project management to monitor shipment and receipt of materials.

At the scoring locations, packets were unpacked, inventoried, and organized. Any discrepancies between the packet inventory and the contents were to be immediately communicated to MI Durham for resolution. However, no such discrepancies occurred.

Handscoring of Constructed Response Items

Rangefinding

The rangefinding process is always the first step in the handscoring of constructed response test items. Committees of Maryland educators have developed content specific generic rubrics. Upon the administration of each new test item, the generic rubric becomes item specific through a process referred to as rangefinding. This process is the foundation of constructed response scoring.

Sample Answer Books for Rangefinding

Sample responses for rangefinding came from three different sources for the 2005 assessments. In January, MI received an early delivery sample of each content area's answer books for supplementing operational responses and for creating new rangefinding sets for field-test responses. On the same day that each January test was administered, sample shipments of completed answer books from schools selected by MSDE were express shipped from the schools to the Measurement Incorporated Central Office in Durham.

The same early delivery process was also used for the new English items administered in May.

In May and summer, there was no early delivery sample for Algebra/Data Analysis, Biology, Geometry, or Government. MI, at MSDE request, used responses from previous field-test administrations to supplement operational responses. (An exception was made for 05 operational items field-tested prior to 2002. As field-test answer books from administrations prior to 2002 had been previously securely disposed of at MSDE request, samples for these items were selected from MSDE-designated schools' materials retrieved on the first pick-up.)

Responses for May field-test rangefinding (Algebra/Data Analysis, Biology, Geometry, and Government) were selected from answer books previously retrieved during regular pick-ups. MSDE supplied a list of selected schools from which these responses were selected.

Selection of Responses for Rangefinding

MI content specialists, in accordance with the generic MSDE rubrics, MSDE answer cues, and MSDE anchor sets, carefully reviewed and selected a variety of responses for the field test items for Maryland educators to evaluate. Using the scoring guides and training materials from previous field test scoring, they also selected additional responses for operational items for Maryland educators to review. In 2005, the MI Content Area Scoring Directors and Team Leaders did much of the preparatory work at the scoring sites, along with specialists at MI's Durham headquarters. This process allowed for rapid selection of responses by personnel with extensive HSA scoring experience. The chosen responses were

assembled in packets that contained an adequate number of responses to show not only the full range of the sample responses, but also a variety of student approaches to each test item. All copying, printing, and shipping functions were carried out by MI, and all materials were kept secure throughout the process.

The rangefinding samples are expected to be representative of the whole assessment. However, whenever a new student approach to a response occurred during the actual scoring sessions, MI always consulted the MSDE Director of Scoring and the MSDE Content Specialists for direction.

Rangefinding

Separate rangefindings took place prior to operational and field-test scoring of the January administration; prior to operational scoring of the May administration; prior to scoring of the summer administration; and, lastly, prior to field-test scoring of the May administration's field test items.

January Rangefinding

Committees composed of educators from Maryland schools and MSDE, along with MI Project Management and MI Content Area Scoring Directors, met prior to the January scoring of constructed responses to pre-score a sample of responses from the current administration. The committees were content specific: Algebra/Data Analysis, Geometry, Government, and Biology. By first training on generic rubrics and established "anchors," or samples from previous administrations, the committee calibrated their scores of student responses to scores from previous administrations. The committees then proceeded to review the existing guides and to score supplemental responses for the operational items. This process was a further opportunity for the members of the committees to calibrate their scores to scores assigned in previous rangefinding meetings. Each new field-test item was then scored, again using the generic rubrics and anchor papers as a means of calibration.

May Rangefinding

In order to supplement training materials for the May operational items in Algebra/Data Analysis, Biology, Geometry, and Government, MI sent the prepared sets of responses for each operational item, along with MI scores, to MSDE Scoring and Content. MSDE then reviewed the sets and noted any discrepancies between their scores and MI scores. Only responses with full consensus from MSDE were selected for use in training materials. A brief rangefinding meeting was held consisting of MI Scoring Directors, MI Project Management, MSDE Scoring and Content Staff.

English rangefinding, with all new items, required a separate rangefinding meeting. As with the other content areas in January, committees composed of educators from Maryland schools and MSDE, along with MI Project Management

and MI Content Area Scoring Directors, met prior to the May scoring of English constructed responses to pre-score a sample of responses from the current administration.

Summer Rangefinding

In order to supplement training materials for the summer operational items in Algebra/Data Analysis, Biology, Geometry, and Government, MI sent the prepared sets of responses for each operational item, along with MI scores, to MSDE Scoring and Content. MSDE then reviewed the sets and noted any discrepancies between their scores and MI scores. Only responses with full consensus from MSDE were selected for use in training materials. (The English summer assessment consisted of forms and items previously scored in the May rangefinding sessions. The summer administration did not include a field-test scoring component.)

May Field Test Rangefinding

Rangefinding of May05 field-test items was accomplished via rangefinding meetings, with committees composed of educators from Maryland schools and MSDE staff, along with MI Project Management and MI Content Area Scoring Directors.

During each of the 2005 rangefinding sessions, as the Maryland educators and/or MSDE staff produced scored responses for each item, these responses became the referenced criteria for the handscoring of those items. Academic discussions of the criteria and the student responses led to a consensus of scores for each rubric score level. The scoring guides and training sets made up of committee-scored papers became the blueprint of the scoring process. All scores assigned throughout the process were based on the foundation laid by these rangefinding sessions.

Team Leader and Scorer Training for Operational Handscoring

Preparation of Training Materials

Upon the completion of rangefinding, the first priority for MI Content Area Scoring Directors was to prepare training materials for operational items. Scoring guides and training sets that were unique to each item contained committee-scored responses from the previous field test administrations. These sets were used, in conjunction with the rubrics, to train Team Leaders and Scorers. Qualifying sets and validity sets, with responses from the current operational administration, were also included in the training and monitoring process for operational scoring.

One guide and two training sets were prepared for each item. Guides typically consisted of two-to-four anchor papers per score point. More examples of each score point were included if a corresponding variety of types of responses had

been found in rangefinding. The number of sample responses for each item varied not only with the complexity of the responses and the extent of the score scale, but also with the variety of student approaches to the item as encountered in rangefinding. Guides included rubrics and, additionally, anchor papers for each score point arranged in the scoring guide in scorepoint order with annotations linking the rubric to the specifics of the individual student responses, thus providing the rationale for the score.

In contrast, student responses in the training sets were in random scorepoint order with no scores or annotations. These sets were given to the Scorers after they trained on the guide. Scorers used the guide and rubric to assign scores to the training set responses.

After completing training on the guides and training sets for every operational item, each Scorer then completed at least two qualifying sets. A minimum standard of perfect agreement with the true scores (consensus scores assigned to the responses by the rangefinding committee) had to be achieved. (Scorers not meeting the standard were dismissed from the project and never scored "live" student responses.) Additionally, for operational scoring, validity check sets were created. These sets, identical to the qualifying sets in structure, were given to each Scorer at least once per week in order to ensure that the Scorer was still assigning accurate scores based on Maryland's criteria.

Any changes in training materials that became necessary as the project evolved were completed with the approval of MSDE Scoring and Content personnel, and any such changes were documented. This included any papers sent to MSDE for scoring decisions. These were documented with the MSDE decision and date. Copies of each scoring guide and each training, qualifying, and validity set (with answer keys) were provided to MSDE. MI also maintains archived copies of the complete training materials.

The following procedures for Team Leader and Scorer training were used for all content areas at all scoring centers.

Team Leader Training

After the guide, training, qualifying, and validity papers had been identified, finalized, and approved, Team Leader training began for each content area. The Content Area Scoring Directors conducted the training of the Team Leaders. Procedures were similar to those for training Scorers (see below) but were slightly more comprehensive, dealing with resolution of discrepant scores, identification of nonscorable responses, unusual prompt treatment (including ESL and dialect), alert responses (e.g., child-in-danger), and other duties performed only by Team Leaders. Team Leaders were required to take careful notes on the training papers in preparation for discussion with the Scorers, and the Content Area Scoring Director counseled Team Leaders on training techniques and application of the rubric.

Scorer Training

After Team Leader training and qualifying was complete, the Content Area Scoring Director conducted the training of Scorers. Training was orchestrated so that Scorers understood how to apply the MSDE rubric and criteria when scoring the papers, learned how to reference the scoring guides, developed the flexibility needed to deal with the variety of responses, and retained the consistency needed to score all papers accurately. In addition to the initial scoring training, a significant amount of time was allotted for demonstrations of paper flow, explanations of "alerts" (see next paragraph) and "flagging" (marking student responses for review by Team Leaders), and instructions about other paper flow and general housekeeping procedures that were necessary for the project.

Alerts were of two types. Packet alerts, which included such things as missing scoring monitors, missing pages, wrong lithocode numbers in packet, etc, were communicated to MI for immediate correction. Student alerts (responses indicating a possible student in danger, possible copying/cheating, etc, were communicated to MSDE for resolution. (See “Decisions and Alerts”.)

All Scorers were trained using the rubrics, and with anchor (guide) papers and training papers scored by the rangefinding committee. Scorers were assigned to a scoring group consisting of one Team Leader and 8 to 10 Scorers. Each Scorer was assigned an individual number for easy identification of his or her scoring work throughout the project.

After the contracts and nondisclosure forms were signed and the introductory remarks given, training began. The Content Area Scoring Director presented and explained the generic rubric. The Content Area Scoring Director then presented the constructed-response item and introduced the guide, then discussed, room wide, each score point and example response. This presentation was followed by practice scoring on the training sets. Each Scorer worked individually to assign scores to the responses in these sets.

Team Leaders collected the monitor sheets after the scoring of each training set and recorded results in a customized log which was examined by the Content Area Scoring Director to determine which papers were giving Scorers difficulty. Scorers broke into teams of eight to ten to score and discuss the papers in the training sets. This gave Scorers an opportunity to discuss any possible points of confusion or problems in understanding the criteria.

The Content Area Scoring Director also “floated” from team to team, listening to the Team Leaders’ explanations and, when necessary, adding additional information. If a particular paper or type of paper seemed to be causing difficulty across teams, the problem was discussed room-wide to ensure that everyone heard the same explanation.

Qualifying

Before they were allowed to read packets of actual student answer books, Team Leaders and Scorers were required to demonstrate their abilities to score accurately by attaining at least the exact agreement percentage established by MSDE. Any Team Leader or Scorer unable to meet the standards set by the MSDE was dismissed. All Team Leaders and Scorers understood this stipulation when they were hired. After reviewing the guide and finishing two training sets for each item, each Team Leader and Scorer then completed two qualifying sets, which incorporated all items from the cluster of items for that form. In order to continue to work on the project, each Team Leader and Scorer had to achieve a minimum percentage of agreement with the "true score" assigned by Maryland rangefinders to each response in the qualifying set.

2005 Minimum Agreement Rates for Qualifying

Content/Item Type	Minimum Exact Agreement Rate	Number of Student Responses Per Set	Number of Items Per Set
Algebra BCR	80%	20	3
Algebra ECR	80%	20	3
Biology BCR	70%	20	3 – 4
English BCR	70%	20	2
English ECR	70%	20	1
Geometry BCR	80%	20	2
Geometry ECR	80%	20	3
Government BCR	70%	20	3 – 4
Government ECR	70%	20	1

Personnel from MSDE Scoring and MI Project Management were on-site during initial Team Leader and Scorer training to monitor the process.

Operational Handscoring

Overview

The following procedures for scoring were used at all scoring centers:

After Scorers had been trained on a given set of items, packets of student answer documents were distributed randomly, by clerical aides under the supervision of the Content Area Scoring Director, to the Scorers by Team. All of the operational items in the Algebra/Data Analysis, Biology, Geometry, and Government packets were read twice. These packets contained two score sheets, one for each reading. The second Scorer used a separate score sheet and was unaware of the scores assigned by the first Scorer.

English responses (May and summer) received 25% double-readings. Before the English score file was posted, MI provided to MSDE a file containing all demographic information for all English test takers. MSDE responded with a file showing any 8th grade students, or students whose grade level could not be determined, to MI. These students, who were required to not only take, but also pass the test, then received additional double-readings so that this group received 100% double readings.

As a Scorer completed a packet of papers, he or she placed it back in the envelope and returned the packet, along with the score sheet, to the Team Leader. The Clerical Aide picked up completed packets and score sheets from Team Leaders. Score sheets collected by clerical staff were visually checked for errors, such as missing bubbles or extra bubbles, then sent to be scanned. The scanner was programmed to automatically reject any score sheet that was incompletely or improperly bubbled. These rejected score sheets were then matched up with the appropriate packet of responses and returned to the Content Area Scoring Director for re-scoring. Aides redistributed the packets designated for second readings. The ID number and team of the first Scorer were recorded on the packet cover. This helped to ensure that that the same Scorer or team did not score the same packet twice. As with the first score sheets, the second score sheets were scanned, and the scores merged into the database. The second score sheet was rejected programmatically if the same Scorer or team was indicated on the first score sheet, and the packet was sent back to be rescored.

Additionally, the MI Content Area Scoring Directors performed resolution readings in the rare cases that the first and second reading scores were not at least numerically adjacent in agreement (e.g., a score of “one” and a score of “three”). First, second, and third reading scores were all reported in the raw score data posted for MSDE and the Development Contractor.

Quality Control of Handscoring

Several procedures ensured quality control on the HSA. The first of these was successful rangefinding. Consistent rangefinding scoring leads to the creation of clear and consistent training materials, which leads to smooth Scorer training which, as a result, enhances the accuracy of handscoring.

A second quality control mechanism was the experience of the MI leadership personnel. MI’s Content Area Scoring Directors were skilled at conducting initial Scorer training and qualifying and were successful in schooling Scorers not only on how to score a variety of responses and still hold to the criteria, but also on how to handle unusual responses. Part of this process was establishing good lines of communication between the Content Area Scoring Directors, Team Leaders, and Scorers.

Third, all Content Area Scoring Directors, all Team Leaders, and usually most of the Scorers at MI's current facilities have had previous experience on HSA and/or large-scale scoring projects. While new Scorers cannot be expected to have had prior scoring experience, all Scorers were trained to implement the scoring criteria and to maintain consistent and reliable scoring throughout the project.

Fourth, unbiased scoring was ensured because the only identifying information on the student responses is an identification number. Unless the students signed their names, wrote about their hometowns, or in some way provided other identifying information, the Scorers had no knowledge of them.

Finally, the quality of each Scorer's work constantly was monitored during the project:

> Content Area Scoring Directors identified scoring trends of individual Scorers during the initial training process and, throughout the scoring of "live" packets, had Team Leaders spot-check Scorers daily. This spot-checking, which consisted of inspecting scores assigned to particular packets, was a major responsibility of Team Leaders throughout the entire course of the project.
>
> Operational constructed response items received a second reading (100% double reads for Algebra/Data Analysis, Biology, Geometry, and Government—25% for English). By matching these scores to those of the first reading, valuable information could be gathered regarding Scorer agreement rates and scoring trends. Scorer status reports were generated and reviewed by the Content Area Scoring Directors and Project Managers, who are experienced in using them to identify Scorers having difficulty, as well as to identify specific items causing problems for the entire room. In the case of a two-point disagreement in scores, a third (resolution) reading was done by the Content Area Scoring Director to ensure the accuracy of the score assigned to the response.
>
> MI's Client Command Center/Project Command Center software program allowed MI Content Area Scoring Directors and Project Management and MSDE to view daily and cumulative reports on score point distribution, agreement rates between Scorers, and numbers of responses scored. Information could be accessed for an individual, team, or the entire group for a specific content area.

Validity Check Sets

Content Area Scoring Directors chose student responses, pre-approved by MSDE, that were placed into unique content-specific validity check set packets of approximately 20 - 25 papers each. These were distributed to each team and

administered daily on a rotating basis. Each Scorer scored at least one of these packets during each week of operational scoring.

Scorers were expected to maintain validity check set scores at least at the minimum qualifying percentage for a particular content. Any Scorer who scored less than this percentage on a validity check set were retrained. If retraining was not successful, the Scorer was dismissed.

Retraining

Spot-checking, validity scores, and status reports provided project management with continuous feedback not only on individual Scorers, but also on room-wide scoring trends. Content Area Scoring Directors met throughout the day with Team Leaders and, using daily status reports, Scorer's questions, and observations from spot-checking, devised retraining strategies to keep Scorers on task with the MSDE criteria. Retraining strategies were geared to the type and degree of scoring difficulty that a Scorer may have been experiencing and were implemented to address scoring problems on an individual basis.

Monitoring

Each Content Area Scoring Director submitted daily progress reports to the MI Project Director. These reports detailed activities during training and scoring, noting any problems or delays encountered. Problems and delays varied by content and by administration. Most were of a minor nature that were resolved without impacting scoring. Examples include weather delays, running out of packets to read, need for MSDE scoring decisions, etc.

Project Management also communicated with the Site Managers, Project Monitors, and the Content Area Scoring Directors via email, phone, fax, or by visiting the scoring centers, as needed.

Decisions and Alerts

Types of responses that were not anticipated and that could not be scored using the rangefinding examples were forwarded to the MI Project Director and Assistant Project Director by the Content Area Scoring Directors. After a brief review, MI Project Management forwarded these responses to MSDE Scoring and Content Specialists for scoring decisions. These decisions and the accompanying explanations from MSDE then were given to the Content Area Scoring Directors. In this way, responses with new and unanticipated approaches to the question, or otherwise aberrant responses, could be scored, and these examples used as scoring tools (guide papers) to score responses with similar strategies. All "decision" responses were documented for the permanent record.

Alerts were handled in a similar fashion. In training, Scorers were advised to flag responses that may indicate teacher interference, plagiarism, suicidal threats or

other threats, or parental/other abuse. They submitted such responses immediately to their Team Leaders or to the Content Area Scoring Directors. At that point, the Content Area Scoring Director submitted a copy of the student response and an accompanying alert form to Project Management in Durham. Project Management then requested identifying student information for the response from the MI Information Technology Department. This information, along with the copy of the response, was then forwarded to MSDE for follow up.

Field Test Scoring

Procedures for the handscoring of field-test constructed response items were similar to the procedures for handscoring of operational constructed response items, including training. However, field test items, which did not have an effect on the students' scores, received 10% double readings. Every tenth answer book scanned was chosen programmatically. Also, qualifying and validity sets were not used for January05 field test scoring, which took place immediately after January05 operational scoring was completed. Scorers were required to maintain satisfactory performance on training sets for each item and to maintain satisfactory agreement rates throughout field test scoring. Qualifying and validity sets were incorporated in May05 field test scoring, which took place well after operational scoring had been completed.

As an adjunct to the statistical evaluation of items, Scorers used an item evaluation sheet daily to record trends and idiosyncrasies observed during scoring. Each Content Area Scoring Director discussed each item with the Scorers, read the evaluation sheets and Team Leader summaries, added his/her own observations, and compiled an anecdotal report on scoring for each item. MI Project Management reviewed this item evaluation information and supplied it to MSDE after the scoring of each test administration.

Preparation and Posting of Data

MI posted five types of score files for each 2005 test administration:

Non-English Final File: this file was posted upon the completion of operational handscoring of Algebra/Data Analysis, Biology, Geometry, and Government. This file included all demographic and header information, all SR and SPR responses, and operational Constructed Response scores for each student record in the file.

English Final File: this file was posted upon the completion of operational handscoring of May05 English. This file included all demographic and header information, all SR responses, and operational Constructed Response scores for each student record in the file.

Field Test File: this file was posted upon the completion of field test handscoring of Algebra/Data Analysis, Biology, Geometry, and Government. It included all

information in the final file, in addition to field test Constructed Response scores for each student record in the file.

Subsequent Files: multiple subsequent files were posted upon the completion of handscoring for materials that arrived late at MI. Since late materials were still arriving from schools after the first subsequent file was posted, additional subsequent files were posted as needed to accommodate all late materials. Subsequent files contained all information that would have been in the final file and field test file for each student record.

Invalidated Records File: this file, sent to MSDE directly from MI, included demographic information, header information, and SR and SPR responses for each invalidated student record. This file included a code showing the reason for invalidation (school-invalidated, MI-invalidated: blank primary week with make-up record present in final file, and MI-invalidated: answer book blank except for demographic information).

Upon completion of handscoring, raw scores assigned to CR items were merged with the data collected from the scanning process, using the same computer program that initially generated and assigned packet numbers and packet positions to the student books. Since these numbers do not rely on any school-produced data, they are an extremely reliable means to ensure that each handscored data record is correctly matched to its student data file captured during the scanning of the actual answer documents.

Data files were generated from the master database server. These data files were made available to the Development Contractor in the desired format. In addition, the final files uploaded to the MSDE server were processed through a quality assurance system developed by MI IT personnel. Each column of data was analyzed based on the type of data valid for that column. The validation requirements were derived from the file layout and descriptions provided by the Development Contractor when they initially transferred the student data files to the database server.

Any questionable data was verified by examining the original data files and/or the original answer document or score sheet. The quality assurance system is, in actuality, a double check, because the provided definition information has already been applied to each data field by the scanning data validation processes prior to the information being stored in the project's master database.

Security Reports

In addition to score files, MI prepared and posted a materials security report for each test administration in 2005. After receipt of test materials, and within 4 weeks of the conclusion of test administration, MI generated an initial Security Report and posted it for MSDE access. The report was posted upon completion

of security scanning of all test books and all used and unused Algebra/Data Analysis and Geometry Answer Books received at MI Durham by the day after the final pickup. The Security Report provided a materials breakdown at both the school and LEA level, indicating for each type of material how many books were received by the school or LEA, how many books were returned to MI, and the percentage of books returned to MI. In addition, the security barcode numbers were listed for any missing materials.

Storage of Materials

As the Scoring Contractor, MI stores used answer documents for the entire contract period. When an entire pallet of storage boxes containing test books was completed, a pallet inventory was produced, detailing the unique bar code numbers of the boxes as well as descriptions of the boxes' contents. This clearly identified materials for storage, retrieval, and eventual recycling. Answer documents were filed in packet order and labeled before being placed in storage. All materials were stored so that retrieval and shipment to Maryland of any documents requested can be accomplished within a 24-hour time frame.

Additionally, test books bearing an accommodations label are stored for the entire contract period.

After each test administration, all test books and all unused answer books are stored securely until MI receives permission from MSDE to destroy these unused materials. All secure materials to be destroyed are shredded on-site at MI headquarters and sent for recycling.

Public Release Items

In 2005 MSDE identified Constructed Response items from each content area to be released on the Maryland State Department of Education web site. MI is currently working in conjunction with MSDE content and scoring staff to prepare annotated student responses for each of these CR items. MSDE Content Leads review and sign off on the final drafts of annotated guides and practice sets before public release.

Staffing

Scoring Project Management

The function of MI Scoring Project Management was to coordinate and execute all handscoring and related activities for the project. The MI Project Director and the Assistant Project Director worked closely with MSDE content and scoring personnel, acted as liaisons between MSDE and the MI Content Area Scoring Directors, and coordinated activities with other contractors. The Project Director

and Assistant Project Director oversaw all MI Content Area Scoring Directors, Team Leaders, Clerical Aides, and Data Processing staff. Scoring Project Management also was responsible for overseeing day-to-day management at all scoring facilities where the HSA scoring took place and for developing all of the scoring guides and other training materials, as well as all the materials used to maintain quality control in training and scoring. Scoring Project Management was also responsible for the training of MI Content Area Scoring Directors. Additionally, Scoring Project Management worked with MI departmental management to oversee activities in the Shipping, Warehousing, and Information Technology departments.

Lead Programmer

The Lead Programmer was responsible for implementing all Information Technology activities related to the project. The Lead Programmer, working under the direction of Scoring Project Management, scheduled and implemented all programming, quality assurance, and data reporting activities as required.

Project Monitors

MI used multiple scoring sites in order to accomplish the large task of scoring the High School Assessment. MI Project Monitors oversaw and administered all scoring projects assigned to their scoring site and communicated daily with MI Content Area Scoring Directors and the Project Director and Assistant Project Director.

Site Managers

Each MI scoring center has an operational supervisor (Site Manager) who recruited Scorers, oversaw the secure receipt, storage, and delivery of all scoring materials and student responses, and supervised on-site warehouse and clerical personnel involved in the scoring project.

Content Area Scoring Directors

Each MI Content Area Scoring Director participated in rangefinding, selected training papers, prepared scoring guides, trained and monitored Scorers and Team Leaders, annotated papers, and directed all operations necessary for conducting a successful project. Additionally, all Content Area Scoring Directors had education and/or experience in the content area to which they were assigned.

MI Content Area Scoring Directors were diligent in adherence to HSA scoring standards and ensured that Team Leaders and Scorers assigned scores to student responses based on these scoring standards. While they competently addressed scoring issues unique to their content areas, they also recognized

issues for which precedent had not been established. They presented these issues to MI Project Management, who conferred with MSDE Scoring and Content specialists for guidance and resolution.

Team Leaders

In selecting HSA Team Leaders, MI's management staff and the Content Area Scoring Directors reviewed the files of all available scoring staff. They looked for people who were experienced Team Leaders with a record of good performance on the HSA, or similar projects, as well as HSA Scorers who had been recommended for promotion to Team Leader.

Effective Scorer training and accurate scoring relies to a great extent on having knowledgeable, flexible Team Leaders. Team Leaders assisted in training Scorers and distributed, collected, and accounted for training packets and sample papers during each training session. During scoring, Team Leaders responded to Scorers' questions and spot-checked scores assigned by Scorers. Team Leaders also monitored the scoring patterns of each Scorer throughout the project, conducted retraining as necessary, and helped to maintain a professional working environment.

In addition to one Team Leader per team of 8 to 12 Scorers, each Content Area Scoring Director had a floating Team Leader. This person directly assisted the Content Area Scoring Director in maintaining paper flow and supervising Team Leaders, and helped other Team Leaders in monitoring Scorer performance during training and scoring.

Scorers

Because MI has been conducting writing and performance assessment scoring for many years, we already had available a pool of qualified, experienced Scorers at our established scoring centers. MI routinely maintains supervisors' evaluations and performance data for each person who works on each scoring project in order to determine employment eligibility for future projects.

As well as employing many of our experienced Scorers for this project, we also recruited new ones. MI procedures for selecting new Scorers are very thorough. After advertising in local newspapers, with the job service, and elsewhere, and receiving applications, staff in our Human Resources Department reviewed the applications and then scheduled interviews for qualified applicants. Qualified applicants were those with a BA or BS in English, education, mathematics, science, social studies, or a related field. Each qualified applicant was required to successfully complete an interview by experienced MI staff, write an acceptable essay, and receive good recommendations from references.

Communication and Planning

Before, during, and after the period of activities for each 2005 test administration, MI maintained communication with MSDE and, through MSDE, the Development Contractor regarding progress and problems encountered. MSDE has also made MI aware of the following changes that will be implemented for the HSA in 2006:

The HSA Geometry test will no longer be administered.

The HSA English test will consist of a January administration, along with May and summer. The May administration will again consist of unique forms with unique items.

MI will provide sample English files, each containing 3,000 students per two forms per file as English scoring is underway.

The second pick-up of materials will include LEA 24 schools.

Rangefinding and scoring of both the January06 and May06 Field Test items for all contents will take place in the Fall.