MME
1050

Michigan Merit Examination

Volume III: Psychometrics & Technical Analyses

2008 Testing Cycle

December 23, 2008

ACT and the Michigan Department of Education

Table of Contents

Preface iii
Chapter 1: Test development analyses 1
Test Specifications 1
Alignment Analyses 1
MME Components 3
Chapter 2: Erasure analyses 5
Description and Purpose 5
Data and Methods 5
Chapter 3: Handscoring analyses 7
Results of Constructed Response Scoring Procedures 7
Rangefinding and Rubric Review 7
Rater Selection 8
Rater Training 9
Rater Statistics and Analyses 10
Chapter 4: Model fit 13
Chapter 5: Scaling and Equating 25
Quality control protocols 25
Results 26
Summary of Comparing the MLE Ability Estimates between PARSCALE and ISE 26
Conclusions 29
Equating/linking/scaling for MME 29
Chapter 6: Reliability 34
SEM/information curves with cuts scores (imposed) 34
Internal Consistency Reliability 34
Empirical IRT Reliability 35
Scale scores (theta): 35
Classification Consistency and Classification Accuracy 35
Chapter 7: Validity 40
Construct Validity Evidence from Content and Curricular Validity 40
Relation to Statewide Content Standards 40
MME Alignment Studies 41
Educator Input 41
Construct Validity Evidence from Criterion Validity 42
Conclusion 42
Chapter 8: Item Analysis 44
Chapter 9: Standard Setting 53
Chapter 10: Adequate Yearly Progress and EducationYES 54
Achievement Status 56
Achievement Change 57
Chapter 11: State Summary Data 63
References 93
Appendices 95
Appendix A: Plots of PARSCALE Information function 96

Appendix B: Data Created for Field-Test Items **100**
Appendix C: Statistics Used on Item Labels for Item Review Committees **143**
Appendix D: Guidelines for Bias Review of Field Test Item Data **146**
Appendix E: Guidelines for Content Review of Field Test Item Data **154**

Preface

This volume documents the technical characteristics of the 2008 Michigan Merit Examination (MME) in light of its intended purposes, and the results of the 2008 operational administration. Analysis results were provided by Michigan's Office of Educational Assessment and Accountability (OEAA), Harcourt Assessment, Inc. (HAI), Pearson Educational Measurement (PEM), and ACT, Inc. The volume is structured around test development analyses (targets, and actual 2008 characteristics), erasure analyses (description of analyses, and actual 2008 results), hand scoring analyses (description of analyses, and 2008 results), model fit (description of analyses, and 2008 results), scaling and equating information related to linking across MME forms, reliability and validity information, item analysis information, standard setting information, and information related to Adequate Yearly Progress and Education YES. The MME is a multi-day examination. Day 1 consists of the ACT Plus Writing assessments. Day 2 consists of two WorkKeys® assessments (*Applied Mathematics* and *Reading for Information*) and an OEAA developed mathematics test. Day 3 (which may be administered on days 2 through 4) consists of OEAA-developed Science and Social Studies tests. The Social Studies assessment includes an essay which is also scored for Writing.

We encourage individuals who want more detailed information on topics that are discussed in this manual, or on related topics, to contact the Michigan Department of Education, Office of Educational Assessment and Accountability.

Office of Educational Assessment & Accountability
Michigan Department of Education
608 W. Allegan Street
P.O. Box 30008
Lansing, MI 48909

Chapter 1: Test development analyses

Test Specifications

Because intact ACT Plus Writing and WorkKeys (*reading for information* and *applied mathematics*) assessments must be included as is in the Michigan Merit Examination (MME), the MME test specifications must start on the foundation of an analysis of the combined alignment of the ACT Plus Writing and WorkKeys assessments. This analysis is the foundation for creating the augmentation needed to assure sufficient alignment of the MME as a whole in each subject to Michigan's high school content standards.

To ensure that the augmented portion of the MME fulfills the requirements for alignment to Michigan's high school content standards, several alignment analyses were conducted. The following section was adapted from the materials submitted to the United States Department of Education for peer review of the MME prior to the first implementation. The evidence referenced in this section is provided as addenda to this technical report.

Alignment Analyses

Three independent alignment studies were conducted on the ACT and WorkKeys against Michigan High School content standards before the pilot of the MME was created.
First, Norman L. Webb, a senior research scientist with the Wisconsin Center for Education Research and the National Institute for Science Education, conducted a preliminary alignment study of the ACT and WorkKeys to the Michigan content standards in December, 2004 as a first step in determining the feasibility of combining a college-entrance exam with a NCLB compliant standards-based exam. The evidence in these reports was used to target augmentation to the ACT and WorkKeys to maximize alignment to the Michigan standards in the pilot of the MME. These reports indicated that of the Michigan ELA standards that are assessable on a large scale, the ACT and WorkKeys combination was well aligned to Michigan's high school standards, with some minor improvements possible. The reader is referred to page 15 of *Alignment Analysis of Language Arts Standards and Assessments: Michigan Grades 9–12. (*Norman L. Webb, 2005). These reports documented some areas of weakness in mathematics and science. The weaknesses in mathematics are summarized on page 13 of *Alignment Analysis of Mathematics Standards and Assessments: Michigan High School.* (Norman L. Webb, 2005). The weaknesses in science are summarized on pages 15-16 of *Alignment Analysis of Science Standards and Assessments: Michigan Grades 9–12.* (Norman L. Webb, 2005). Augmentation was targeted to the weak areas.

Second, John Dossey of Illinois State University evaluated the Mathematics and Science ACT Test items and WorkKeys items in comparison to the Michigan Mathematics and Science content expectations for High School. He identified remarkable consistency between the ACT/WorkKeys and the Michigan content standards, with a few areas of weakness. The weaknesses he identified were in mathematical content coverage of patterns, functions, probability and discrete mathematics, as described on page 14 of *Comparison of the ACT and WorkKeys Assessments with the Mathematics and Science Content Expectations in the Michigan Curriculum Framework.* (John A. Dossey, 2005). Although science was well covered, identified weaknesses in life, physical, and earth science are summarized on page 20 of the same document (John A. Dossey, 2005). Augmentation was targeted to maximize alignment on these areas.

Third, Timothy Shanahan of University of Illinois at Chicago evaluated the ACT and WorkKeys items in comparison to the Michigan English Language Arts (ELA) content standards. In summary, the reviewer clearly states on page 7 of *Review of ACT Coverage of Michigan Language Arts Standards* (Timothy Shanahan, 2005) that the ACT English and Reading assessments are strongly aligned with the Michigan ELA content standards. Although the alignment study suggested no need to further augment the ELA portion of the assessment, OEAA chose to augment the Writing portion. Specifically, in order to resolve a Balance of Representation issue, we added a score for Social Studies Decision Making (constructed response item) to the Writing total score. This addition offset the large number of English Multiple Choice points that were being counted as part of the Writing score.

Post-Hoc Alignment Studies of the Pilot Michigan Merit Exam

Norm Webb from the University of Wisconsin led another alignment study for the Michigan Merit Examination pilot in May, 2006, involving curriculum, instruction and assessment experts from within and outside of the State. For the English Language Arts (ELA) and mathematics portions of the MME, alignment was considered in regard to both the current (2004) Michigan Curriculum Framework Standards and Benchmarks and the soon-to-be-implemented (2006) Content Expectations. For this report, we will only be considering alignment with respect to the existing Standards and Benchmarks.

Members of the alignment teams were solicited from a diverse group of educators who had not previously taken part in developing the assessment instruments, in order to ensure the objectivity of the study.

The alignment studies indicated the following for the individual content areas…

For ELA, seven of the twelve current (2004) standards can be reasonably addressed by an on-demand assessment, as stated on page 10 of *Alignment Analysis of Reading and Language Arts Standards and Michigan Merit Exam: Michigan High School* (Norman L. Webb, 2006). The MME demonstrated Categorical Concurrence for all seven standards (see page 9). Five standards showed Depth-of-Knowledge Consistency and Range of Knowledge, and all but one had an appropriate Balance of Representation.

For mathematics, there are six current (2004) standards, all of which can be addressed in an on-demand assessment. As described in *Alignment Analysis of Mathematics Standards and Michigan Merit Exam: Michigan High School (*Norman L. Webb, 2006), the MME demonstrated Categorical Concurrence on all six standards. Four standards showed Depth-of-Knowledge Consistency, two had an accepTable Range of Knowledge, and all but one had an appropriate Balance of Representation.

For science, the panel concluded that the alignment is reasonable if only the benchmarks that are more suitably assessed by an on-demand assessment are considered. These analyses are described in *Alignment Analysis of Science Standards and Michigan Merit Exam: Michigan High School* (Norman L. Webb, 2006). Of the five current (2004) standards, all but "Reflecting on Scientific Knowledge" demonstrated Categorical Concurrence. This was corrected beginning with the Spring 2007 MME by adding six items assessing Reflecting on Scientific Knowledge. These items were selected to also address depth of knowledge, range of knowledge, and balance

of representation. Of the remaining standards, all showed Depth-of-Knowledge Consistency, three had an accepTable Range of Knowledge, and all had an appropriate Balance of Representation.

The new Michigan Merit Examination (MME) is based on two ACT assessments: the ACT Plus Writing and two WorkKeys assessments (Reading for Information and Applied Mathematics), with Michigan-developed augmented portions designed to address standards not covered by the ACT tests and the WorkKeys assessments. In assembling the Michigan-developed component for MME, the post-hoc alignment studies were used to indicate areas where the ACT and WorkKeys tests need to be augmented.

From the results of the post-hoc alignment studies, it appears that the targeted augmentations of the Mathematics and Science assessments were effective.

MME Components

In accordance with the contents of the ACT Plus Writing and WorkKeys assessments, in accordance with the results of the alignment analyses, and in accordance with legislation. Table 1-1 illustrates that the overall MME is composed of the following components for each subject:

Table 1-1. Components of MME Test Scores

				Components Contributing to MME Scores*				
Day	**Test**	**Subject Session**	**Total ELA**	**Reading**	**Writing**	**Mathematics**	**Science**	**Social Studies**
Day 1	**ACT Plus Writing**	English	X		X			
		Mathematics				X		
		Reading	X	X				
		Science				15 items	X	
		Writing	X		X			
Day 2	**WorkKeys**	Reading for Information	X	X				
		Applied Mathematics				X		
	Michigan Mathematics	Michigan Mathematics				X		
Day 2, 3, or 4	**Michigan Science and Social Studies**	Science					X	
		Social Studies	X		X			X

Note that the ACT Plus Writing was given on day 1 of the assessment, the WorkKeys and Michigan mathematics augmentations were given on day 2, and the remaining Michigan augmentation sections were given on the third day (which can be completed on any one of three days). For each subject (column), students needed to complete each section shown with an "X" to obtain a valid score on the MME.

There are two points of particular interest in this Table. First, note that 15 of the ACT science items count toward MME mathematics. This occurs because the data analysis items on the ACT

science assessment align well with Michigan's high school mathematics content standards. Second, note that the social studies component contributes to the Writing score. This occurs because the social studies extended writing prompt (persuasive civic writing) is scored both for social studies content and for writing in accordance with Michigan's high school writing standards.

The MME ELA is an average of MME Writing and MME Reading. It consists of five components, as shown in Table 1-1.

In developing the augmentation, it was not feasible to employ many of the procedures that the Michigan Department of Education typically employs for test development because the spring 2007 administration of the Michigan Merit Examination (MME) was the first administration of a new assessment using a new scale, and because two components of the MME are pre-designed by ACT. Therefore, there did not exist any Item Response Theory (IRT) item parameter estimates for items to be used on the spring 2007 administration (with the exception of items used to link to the pilot study of spring 2006). Therefore, all analyses used to support test development had to be performed using classical test theory (CTT) statistics. However, for the spring 2008 administration, IRT parameter estimates were available for many items. The inclusion rules were, in order of decreasing importance, the following:

1. Alignment to content standards needing augmentation.
2. Positive corrected point-biserial correlations with either the MME pilot or past MEAP high school scores (preferably above 0.25, but no negatives) where statistics were available.
3. Creation of a reasonable distribution of classical item difficulty where statistics were available, meaning approximately one quarter of the items in each of the following ranges: 0.26-0.50, 0.51-0.75, and 0.76-1.00. Generally, we do not select items in the range of 0.00- 0.25 unless such items are absolutely needed for content alignment.
4. IRT parameter estimates were reviewed when available.

Because classical statistics were gathered from different sources (the MME pilot versus previous assessments) the distributions are not presented as the statistics do not all come from the same population.

For future cycles of the MME, more sophisticated analyses will be run for developing the assessments to ensure that they will be equiTable. These include analyses of the distribution of IRT parameters, projected SEM/Information curves, projected reliability, and projected classification accuracy. The comparison with the baseline (previous year) will be included with current projections to evaluate the overall similarity of each year's assessment to the previous year.

NOTE: Item development for the augmented portion of the MME occurred during the period of the previous High School assessment (the Michigan Educational Assessment Program, or MEAP). The item development protocols and quality assurance checks are detailed in the 2005/06 final MEAP technical report.

Chapter 2: Erasure analyses

Description and Purpose

Erasure analysis (also known as mark darkness analysis) is an analysis of the degree to which certain groups of students tend to mark and then erase those marks on multiple choice items. The purpose is to identify unusually low or unusually high rates of answer changing behavior as circumstantial evidence to support investigations in situations where allegations of widespread cheating have been received and to identify plausible targets for on-site monitoring.

Data and Methods

The data captured to analyze erasure patterns is described here. In a data file with one row per student per subject, the following data are captured:

- DistrictCode (NULL for state rollup)
- BuildingCode (NULL for district rollup)
- Grade (NULL for all grades rollup)
- Subject (NULL for all subjects rollup)
- NW2W (Number of wrong to wrong erasures)
- NW2R (Number of wrong to right erasures)
- NR2W (Number of right to wrong erasures)

Based on the form of the assessment and upon the data already in the file, the following two fields are added to the student-level file:

- Nerase (Total number of erasures, or NW2W+NW2R+NR2W)
- Ntotal (Total number of MC items responses)

From these data, summary data files are created with one row for each district/school/grade/subject combination. Each row of the file contains the following data:

- DistrictCode
- BuildingCode (NULL for district rollups)
- Grade
- Subject
- DistrictCode (NULL for state rollup)
- BuildingCode (NULL for district rollup)
- Grade (NULL for all grades rollup)
- Subject (NULL for all subjects rollup)
- NW2W (sum of wrong to wrong erasures over all students)
- NW2R (Number of wrong to right erasures over all students)
- NR2W (Number of right to wrong erasures over all students)
- Nerase (Total number of erasures, or NW2W+NW2R+NR2W)
- Ntotal (Total number of MC items responses)

From the data in the summary file, two additional fields are created for each row as follows:

R1 (ratio of all erasures to all responses in the combination, or Nerase/Ntotal)
R2 (ratio of wrong-to-right erasures to all erasures in the combination, or NW2R/Nerase)

Based upon the data in this file, four threshold values are calculated for each statistic and each subject at the district level and at the school level. These thresholds are based on the distributions of the ratio statistics at the district and school level. These thresholds may change based on their usefulness in operation, but current plans are that they will be:

1. 3SDlow (3 standard deviations below the mean or zero, whichever is greater)
2. Prcntlow (The 5^{th} percentile)
3. 3SDhigh (3 standard deviations above the mean)
4. Prcnthigh (The 95^{th} percentile)

Based on these thresholds, the following flags are applied in the summary data files:

- R1LowSD (1 if less than 3SDlow, 0 otherwise for R1)
- R1LowPct (1 if less than Prcntlow, 0 otherwise for R1)
- R1HighSD (1 if greater than 3SDhigh, 0 otherwise for R1)
- R1HighPct (1 if greater than Prcnthigh, 0 otherwise for R1)
- R2LowSD (1 if less than 3SDlow, 0 otherwise for R2)
- R2LowPct (1 if less than Prcntlow, 0 otherwise for R2)
- R2HighSD (1 if greater than 3SDhigh, 0 otherwise for R2)
- R2HighPct (1 if greater than Prcnthigh, 0 otherwise for R2)

Based on these flags, district/school/grade/subject combinations with unusually low or unusually high ratios are identified. The criteria for identifying individual combinations will need to be determined through more experience with operational data.

However, there will be at least two uses of the data. First, these data will be used as evidence in investigations following up on allegations of unethical behavior. Second, these data will be used to target individual schools and/or districts for on-site monitoring by MDE and/or contractor staff during the next assessment cycle. It is expected that the erasure data will also be useful in research on erasure patterns as related to item characteristics.

Because the behaviors of these summary statistics are not well known, either in a univariate or bivariate fashion, summary statistics will also be presented to inform OEAA understanding. These summaries will display both graphically and numerically the univariate and bivariate distributions of the ratio statistics, thresholds, and flags where the displays are reasonable. These displays will aid in future construction of erasure analysis indices .

Chapter 3: Handscoring analyses

Results of Constructed Response Scoring Procedures

The MME assessment includes measures in which the examinees must construct their own response for some of the questions. The procedure for scoring these responses is provided.

Outlined below is the scoring process that the PEM Performance Scoring Center (PSC) follows. This procedure is used to score responses to all MME constructed response or written composition items.

Rangefinding and Rubric Review

Pearson Educational Measurement (PEM) Scoring Center Management, Scoring Directors and Supervisors worked in conjunction with OEAA staff to develop the constructed response scoring procedures. In addition to the PEM PSC scoring proposal, PEM staff created a range-finding schedule and work plan sent to OEAA on May 31, 2005.

OEAA staff reviewed the PEM Proposal for Rangefinding to be conducted by PEM Scoring Center Management, Scoring Directors and Supervisors. The rangefinding proposal was accepted by OEAA on June 13, 2005.

PEM conducted an internal rangefinding, supplemented the field test training sets, and submitted these to OEAA for approval. The plan included requirements for each item to be scored including:

1. Rubric
2. Comment Codes (If Applicable)
3. Operational Anchor, Practice, and Qualification papers (with annotations where applicable) from the previous year to use as a guide
4. Sets used to train scorers for the field test
5. Any scoring decisions or scoring notes that come from field test rangefinding and/or scoring.

For pilot and field test items, rangefinding is done as part of the scoring process. Small scoring teams are led by a scoring supervisor, who together with the team reviews the rubrics for a particular form and then reviews a sampling of the books before assigning scores to the books. Problematic issues are discussed with the OEAA. After a consensus has been reached, the teams score all books for that form. Group discussion takes place for problematic papers. At this time the scoring supervisor constructs an exemplar set, with papers for each score point for each item.

Prior to scoring the operational assessment, the PSC's subject teams conduct rangefinding and rubric review activities. In conjunction with OEAA, PEM conducts a review of the rubrics used immediately prior to rangefinding. This establishes a baseline among all the participants. PEM reviews the rubrics with OEAA and the participants on an as-needed basis throughout the course of rangefinding.

Rangefinding materials are chosen from field test materials and in some cases from the archival image banks provided by Measurement Incorporated (the previous MEAP-HST contractor). The PSC staff assembles those materials with enough copies so that all members of the rangefinding committees have working copies at the meetings. The thoughtful selection of papers during rangefinding and the subsequent compilation of anchor papers and other training materials are essential to ensuring that scoring is conducted consistently, reliably and equitably. Teams review a sufficient number of papers from the field tests to select a representative sample of the papers for inclusion in the training sets. Often this number is in excess of 200 papers.

The PSC's scoring team conducts rangefinding meetings and selects exemplar papers for the social studies and writing constructed responses. Items are selected from those given in the field tests. Exemplar papers are selected from field test materials to provide a representative sample from a wide range of Michigan school districts.

The primary task in the selection of training papers is the identification of anchor papers - examples that clearly and unambiguously represent the solid center of a score point as described in the rubric. Those anchor papers form the basis not only of scorer training, but of subsequent discussions as well. The rangefinding team compiles careful notes during its preparation of training sets, and those notes are used to support decisions when replacement responses must be identified.

The goal of the rangefinding meetings is to identify a sufficient pool of student responses which illustrate the full range of student performance in response to the prompt or item, and for which consensus scores can be resolved. This pool of responses will include borderline responses—ones that do not fit neatly into one of the score levels and that, therefore, represent some of the decision-making problems that scorers may face—as well as drawing a line between two score points. As the final step in selecting the exemplar and marker papers, the reviewers will view all the papers that have been assigned the same score point as a check for intra-year consistency of decision-making.

All reasonable steps are taken throughout preparation of the rangefinding materials as well as during the meetings to ensure security, including storing the materials in locked facilities and locking unattended meeting rooms. All rangefinding materials are accounted for at the conclusion of each session.

Following rangefinding and the approval of selected training papers anchor sets are assembled. Drawing from the pool of additional resolved student responses, it constructs the practice sets to be used in scorer training. As those sets are assembled, they are forwarded to the OEAA for review and approval, as further assurance that committee decisions have been accurately enacted.

Rater Selection

Highly qualified scorers are essential to achieving and maintaining a high degree of consistency and reliability in scoring students' responses. The careful selection of professional scorers to evaluate the constructed response items and writing tasks will therefore be essential to scoring the MME. PEM has compiled a personnel database containing the academic training and professional experience of more than 4,500 college graduates who have completed the stringent selection process for scorers. This process requires that each candidate successfully complete a

personal interview, a written essay assignment, and a grammar and editing test or a mathematics and science test when appropriate. Such pre-screening of candidates ensures that only the highest calibers of scorers are selected. Throughout the selection process, PEM actively emphasizes the need for ethnic and racial diversity among professional scorers. Included in this diverse pool is a core group of veteran scorers whose insight, flexibility and dedication have been demonstrated while working on a range of performance assessments.

Scoring supervisors are chosen from that pool of scorers based on demonstrated expertise regarding all facets of the scoring process, including strong organizational abilities and skill in training strategies. Those individuals chosen to perform these assignments possess practical skills, leadership abilities and sensitivity to interpersonal communication requirements. Supervisors also possess the essential capability of assimilating and helping scorers understand the particular scoring requirements of the OEAA.

Upon hiring, scorers sign a confidentiality agreement in which they agree to keep all information and student responses confidential. Scorers and scoring supervisors are trained to internalize the rubric and score according to the scoring guides developed for the specific assessment.

At the beginning of each scoring project, all scoring supervisors and scorers assigned to the project will complete project-specific training.

Rater Training

Thorough training is vital to the successful completion of any scoring. Subject leaders follow a series of prescribed steps to ensure that training is consistent and of the highest quality. The PSC staff develops its training materials to reach all three types of learners: visual, auditory and kinesthetic.

Prior to scorer training, the PSC subject leaders conduct scoring supervisor training. A primary goal of this session is to ensure that scoring supervisors clearly understand the scoring protocols and the training materials. This ensures that all responses are scored in a manner consistent with the scores assigned to the anchor papers and according to the intentions of the OEAA. Scoring supervisors read and discuss the assessment items along with the rubrics which are used to score them. They are expected to carefully read and annotate all training materials so that they can readily assist in scorer training and respond to scorers' questions during training and scoring.

The training agenda includes an introduction to the MME. It is important for scorers to have an understanding of the history and goals of the assessments and the parameters within which students' responses are evaluated. This gives them a better understanding of what types of responses can be expected. The scorers then receive a description of the scoring criteria, which will be applied to the responses. Next, the trainers turn to the first item to be scored and to the scoring rubric itself.

The primary goal of training is to convey to the scorers the decisions made during training paper selection about what type(s) of responses correspond to each score point and to help scorers internalize the scoring protocol so that they may effectively apply those decisions.

Scorers are better able to comprehend the scoring guidelines in context, so the rubric is presented in conjunction with the anchor papers. Anchor papers are the primary points of reference for scorers as they internalize the rubric. There are three anchor papers per item for each score point value. Trainers draw scorers' attention to the score point description from the scoring guide, as well as the illustrative anchor papers encouraging scorers to immediately connect the language of the rubric with actual student performance.

After presentation and discussion of the anchor papers, each scorer is shown a practice set. Practice papers represent each score point and are used during training to help scorers become familiar with applying the rubric. Some papers clearly represent the score point. Others are selected because they represent borderline responses. Use of these practice sets provides guidance to scorers in defining the line between score points.

Training is a continuous process, and scorers are consistently given feedback as they score. With the help of the reliability reports, the scoring lead staff can closely monitor each scorer's performance. In order to document retraining efforts for scorers with low reliabilities, the PSC maintains a Scorer Intervention Log. This form describes the feedback given a scorer regarding his or her problematic scoring and consolidates the interventions taken.

Rater Statistics and Analyses

Calibration

A variety of reports are produced throughout the scoring process to allow scoring supervisory staff to monitor the progress of the project, the reliability of scores assigned and individual scorers' work. Those reports include:

- *Daily and Cumulative Inter-rater Reliability Reports by Item and Scorer*. These reports provide information about how many times scorers were in exact agreement, assigned adjacent scores or required resolutions. The reliability is computed and is monitored daily and cumulatively for the project.
- *Daily and Cumulative Frequency Distributions*. These reports show how many times each score point has been assigned to the item being scored by reader. They are produced both on a daily basis and cumulatively for the entire scoring project. This report allows scoring supervisors and subject leaders to see whether scorers have a tendency to score consistently high or low.

Two types of inter-rater reliabilities are reported at the end of the scoring process: Pearson correlations and scorer percent of agreement which is the sum of exact and adjacent percent of agreement. Both types of inter-rater reliabilities are reported in Table 3-1. The correlations appear to be strong. Inter-rater agreement indices, as expressed by the sum of perfect and adjacent percent agreement, are very high for both scores (99.6 and 98.5).

Rater Monitoring and Retraining

The most immediate method of monitoring a scorer's performance is through backreading by scoring supervisors. If a scoring supervisor discovers that a scorer is consistently assigning scores other than those the scoring supervisor would assign, he or she re-trains that scorer, using the original anchor papers and training materials. This immediate check and remedial correction also provide an effective guard against scorer drift.

Rater Dismissal

Readers are dismissed when, in the opinion of the subject leaders, those readers have been counseled, retrained and given every reasonable opportunity to improve, and are still performing below the accepTable standard.

Score Resolution

In the MME Assessment, every constructed-response item is scored by two scorers. All non-adjacent scores are submitted to scoring directors or scoring supervisors for review, and are resolved by expert scorers appointed by scoring directors.

Inter-Rater Reliability Results

Inter-rater agreement is expressed in terms of exact agreement (Reader Number One's score equals Reader Number Two's score) plus adjacent agreement (+/- 1 point difference). Inter-reader reliability in percent of agreement and Pearson correlations are summarized in Table 3-1.

Rater Validity Checks

An additional set of data, known as validity scoring, are collected daily to check for reader drift and reader consistency in scoring to the established criteria. When scoring supervisors identify ideal student responses, they route these to the scoring directors for preview. Scoring directors review the responses and choose appropriate papers for validity scoring. Validity responses are usually solid score point responses. The scoring directors confirm the true score and enter the response for validity scoring. Readers score a validity response approximately every 30 responses for Social Studies and Writing. Validity scoring is blind; because image based scoring is seamless, scorers do not know when they are scoring a validity response. Results of validity scoring are analyzed regularly by scoring directors, and appropriate measures are initiated as needed, including the retraining or releasing of scorers. Rater validity percent of agreement is reported in Table 3-1.

Table 3-1. Rater Validity Percent of Agreement for Spring 2008

Absolute difference in scores between two raters	ACT essay	Michigan essay for ELA	Social Studies
0	72.91	64.59	64.72
1	26.42	35.34	33.56
2	0.43	0.07	1.49
3 or higher	0.01	0.00	0.03

Chapter 4: Model fit

The MME Writing, Mathematics, Reading, and Science assessments were scaled and are equated using PARSCALE and a three parameter logistic IRT/generalized partial credit model for item calibration. (The methods used for estimating examinee scores is discussed later in this document.) The MME Social Studies assessment was scaled with the Rasch partial credit model using WINSTEPS.

The MME calibration runs for Writing, Mathematics, Reading and Science were conducted using PARSCALE (Muraki & Bock, 1997) under the generalized partial credit model for constructed response items and the three parameter logistic model for dichotomous items. Two model fit indices were used for the dichotomous and polytomous items. They are the Chi-square (χ^2) statistics provided in PARSCALE phase 2 output generated from the calibration runs, and Orlando & Thissen's (2000) S-X^2 statistics. To compute the Chi-square index, the number of ability groups defined was 10, which coincides with the MME item analysis practice of using 10 deciles. Tables 4.1 to 4.4 contain the item fit statistics of all MME scored items on the initial forms for the test subjects of Writing, Reading, Mathematics and Science, respectively.

To test the goodness of fit for each item, a significance level (α) of .05 was used. If the observed p-value associated with the fit indices for an item was lower than .05, the item was considered a "poorly" fitting item. The χ^2 tests of item fit are, however, extremely sensitive to sample size, which is very large for MME.

For all subjects, the Pearson χ^2 statistics tended to be significant. One plausible reason for the observed misfit is the degree of multidimensionality in the assessments that occurs because of the lack of state control over portions of the assessment. A consequence of multidimensionality is that the first principal component being measured on Writing and Mathematics is not as strong as is usually possible to construct when one has complete control over test design and development.

However, this does not invalidate the measure. This simply indicates that beyond the strong overall achievement measured by the MME subject tests, there are also some minor dimensions of achievement that impact the individual item scores of individual students. That the overall dimensions (or principal components) measured by each subject assessment are very strong is demonstrated by both (1) strong Cronbach's alpha internal consistency reliabilities (a Classical Test Theory index of measurement precision of the overall dimension), and (2) strong empirical IRT-model-based reliabilities (a measure of measurement precision of the overall dimension derived from the IRT model). For these measures of reliability, see Chapter 6 where all internal consistency and empirical IRT reliabilities are reported to be 0.89 or higher.

In addition, Yen and Fitzpatrick (2006) indicate that item misfit is typically caused by using an underspecified psychometric model (such as the Rasch or 2-PL model when items provide differing levels of information about the principal component, or when guessing is prevalent).

Yen and Fitzpatrick (2006) describe additional causes of item misfit, including differential item functioning, small sample sizes, poorly estimated item parameters, item stem quality, item miskeys, and item distractor quality. All of these potential causes were carefully investigated and rectified through both ACT and Michigan processes.

Given that other possible sources of item misfit have been carefully addressed, and given that the Generalized Partial Credit Model is the most highly specified psychometric model that has been validated for use in large-scale assessment, the use of that model for MME is the best possible choice available to increase item fit.

Finally, the matrix plots of item characteristic curves resulting from PARSCALE calibration runs are presented in Figures 4-1 to 4-4. In these plots, there are some item characteristic curves (ICCs) that represent serious concerns (e.g. nearly flat ICCs). In these cases, items that exhibited poor ICCs were eliminated from scoring. Note that this tended to occur with the WorkKeys items where there are sufficient items from each content standard covered by WorkKeys to ensure that alignment to Michigan content standards is not degraded by dropping a small number of items.

For MME Social Studies, the mean square fit (MNSQ) statistics obtained from WINSTEPS were used to determine whether items were functioning in a way that is congruent with the assumptions of the Rasch mathematical model. Two types of MNSQ values are presented, OUTFIT and INFIT. MNSQ OUTFIT values are sensitive to outlying observations. MNSQ INFIT values are sensitive to behaviors that affect students' performance on items near their ability estimates. According to the item analysis specification, the model is considered to be moderately misfit if the values are between 1.5 and 2.0, and highly misfit if the values are greater than 2.0. These fit indices are presented in Table 4.5.

Table 4-1. Item Fit Statistics – Writing for Spring 2008

Item	X2	df	p	SX2	df_SX2	p_SX2
AE01	169.08	10	0.00	51.67	62	0.82
AE02	863.71	10	0.00	80.62	56	0.02
AE03	231.73	10	0.00	62.44	66	0.60
AE04	113.41	10	0.00	49.34	61	0.86
AE05	2024.36	10	0.00	118.09	59	0.00
AE06	311.07	10	0.00	69.60	63	0.27
AE07	592.06	10	0.00	55.52	62	0.71
AE08	291.92	10	0.00	61.28	62	0.50
AE09	498.11	10	0.00	48.66	63	0.91
AE10	575.28	10	0.00	71.06	65	0.28
AE11	2750.14	10	0.00	90.57	59	0.01
AE12	316.62	10	0.00	58.82	59	0.48
AE13	1192.41	10	0.00	115.52	64	0.00
AE14	127.95	10	0.00	80.94	64	0.07
AE15	205.70	10	0.00	57.91	63	0.66
AE16	1668.42	10	0.00	59.54	62	0.57
AE17	1293.00	10	0.00	62.46	65	0.57
AE18	1194.86	10	0.00	69.20	63	0.28
AE19	2949.04	10	0.00	108.89	67	0.00
AE20	252.63	10	0.00	74.75	58	0.07
AE21	916.20	10	0.00	99.50	65	0.00
AE22	296.72	10	0.00	59.54	65	0.67
AE23	202.37	10	0.00	70.63	57	0.11
AE24	137.90	10	0.00	51.05	62	0.84
AE25	499.67	10	0.00	91.84	65	0.02
AE26	1473.73	10	0.00	45.69	62	0.94
AE27	74.71	10	0.00	84.16	59	0.02
AE28	230.21	10	0.00	60.57	55	0.28
AE29	629.17	10	0.00	62.35	61	0.43
AE30	1212.33	10	0.00	98.50	60	0.00
AE31	272.53	10	0.00	72.16	64	0.23
AE32	1029.41	10	0.00	78.79	58	0.04
AE33	489.52	10	0.00	61.63	66	0.63
AE34	449.46	10	0.00	84.71	65	0.05
AE35	331.94	10	0.00	54.95	66	0.83
AE36	2132.01	10	0.00	112.18	62	0.00
AE37	33.20	10	0.00	79.61	64	0.09
AE38	437.53	10	0.00	85.58	64	0.04
AE39	416.51	10	0.00	77.15	64	0.13

Item	X2	df	p	SX2	df_SX2	p_SX2
AE41	802.06	10	0.00	54.14	64	0.81
AE42	1857.83	10	0.00	106.49	66	0.00
AE43	1476.76	10	0.00	90.08	62	0.01
AE44	416.37	10	0.00	57.20	57	0.47
AE45	314.32	10	0.00	60.02	58	0.40
AE46	2297.76	10	0.00	106.28	63	0.00
AE47	1069.30	10	0.00	80.69	64	0.08
AE48	741.95	10	0.00	82.28	67	0.10
AE49	612.44	10	0.00	76.06	61	0.09
AE50	769.99	10	0.00	60.88	65	0.62
AE51	182.31	10	0.00	39.10	60	0.98
AE52	1036.88	10	0.00	84.86	61	0.02
AE53	222.27	10	0.00	57.07	60	0.58
AE54	479.91	10	0.00	62.73	61	0.41
AE55	1670.21	10	0.00	88.80	60	0.01
AE56	334.00	10	0.00	69.47	63	0.27
AE57	739.27	10	0.00	63.41	59	0.32
AE58	219.99	10	0.00	66.96	65	0.41
AE59	1141.78	10	0.00	67.64	60	0.23
AE60	3541.56	10	0.00	73.78	61	0.13
AE61	1553.90	10	0.00	85.05	66	0.06
AE62	939.11	10	0.00	58.64	59	0.49
AE63	1176.43	10	0.00	56.91	60	0.59
AE64	383.46	10	0.00	78.37	65	0.12
AE65	560.81	10	0.00	54.81	63	0.76
AE66	1701.72	10	0.00	87.84	63	0.02
AE67	1611.93	10	0.00	61.31	62	0.50
AE68	459.19	10	0.00	89.34	64	0.02
AE69	409.97	10	0.00	53.67	64	0.82
AE70	968.28	10	0.00	64.64	64	0.45
AE71	705.19	10	0.00	88.98	63	0.02
AE72	2127.45	10	0.00	85.80	63	0.03
AE73	1794.73	10	0.00	68.95	63	0.28
AE74	470.31	10	0.00	54.47	66	0.84
AE75	1392.28	10	0.00	72.43	66	0.27
AW01	2044.15	46	0.00	162.68	158	0.38
AW02	1953.65	46	0.00	157.12	158	0.50
MW01	2499.24	48	0.00	126.04	143	0.84
MW02	2460.73	48	0.00	131.41	143	0.75

Table 4-2. Item Fit Statistics – Reading for Spring 2008

ITEM	X2	df	p	SX2	df_SX2	p_SX2
AR01	553.58	10	0.00	36.02	41	0.69
AR02	508.46	10	0.00	84.84	44	0.00
AR03	208.91	10	0.00	43.36	45	0.54
AR04	804.73	10	0.00	42.75	42	0.44
AR05	1098.36	10	0.00	36.99	44	0.76
AR06	319.49	10	0.00	75.96	43	0.00
AR07	1198.28	10	0.00	99.34	44	0.00
AR08	599.61	10	0.00	59.68	45	0.07
AR09	419.16	10	0.00	54.06	44	0.14
AR10	631.25	10	0.00	51.62	45	0.23
AR11	398.04	10	0.00	80.15	44	0.00
AR12	746.67	10	0.00	71.38	44	0.01
AR13	287.87	10	0.00	73.26	45	0.00
AR14	572.58	10	0.00	37.96	45	0.76
AR15	725.42	10	0.00	53.18	43	0.14
AR16	1145.24	10	0.00	55.62	44	0.11
AR17	236.43	10	0.00	43.41	45	0.54
AR18	373.42	10	0.00	47.57	44	0.33
AR19	627.33	10	0.00	76.05	44	0.00
AR20	227.29	10	0.00	48.59	45	0.33
AR21	531.13	10	0.00	39.14	43	0.64
AR22	1097.25	10	0.00	66.96	46	0.02
AR23	667.59	10	0.00	37.57	44	0.74
AR24	655.84	10	0.00	62.58	45	0.04
AR25	168.36	10	0.00	55.89	44	0.11
AR26	613.59	10	0.00	49.02	43	0.24
AR27	460.76	10	0.00	63.25	44	0.03
AR28	1085.13	10	0.00	66.89	43	0.01
AR29	691.54	10	0.00	46.07	44	0.39
AR30	1105.49	10	0.00	81.98	43	0.00
AR31	1302.18	10	0.00	58.25	44	0.07
AR32	2510.82	10	0.00	103.68	43	0.00
AR33	1726.16	10	0.00	66.86	43	0.01
AR34	922.64	10	0.00	74.90	42	0.00
AR35	1450.14	10	0.00	87.75	43	0.00
AR36	1130.28	10	0.00	67.52	45	0.02
AR37	884.25	10	0.00	75.86	44	0.00
AR38	907.55	10	0.00	35.29	44	0.82
AR39	1850.96	10	0.00	59.96	44	0.05
AR40	712.56	10	0.00	62.01	44	0.04
WK01	72.24	9	0.00	32.86	29	0.28
WK02	95.96	10	0.00	18.43	31	0.96
WK03	146.12	9	0.00	16.57	29	0.97
WK04	96.57	10	0.00	22.83	31	0.86
WK05	164.64	10	0.00	26.63	35	0.84
WK06	251.67	10	0.00	17.53	34	0.99
WK07	127.05	10	0.00	33.82	39	0.70
WK08	215.07	10	0.00	31.58	38	0.76
WK09	759.91	10	0.00	33.81	36	0.57
WK10	182.47	9	0.00	37.83	33	0.26
WK11	153.71	10	0.00	41.43	44	0.58
WK12	197.17	10	0.00	37.52	43	0.71
WK13	244.55	10	0.00	31.29	39	0.81
WK14	222.98	10	0.00	41.81	46	0.65
WK15	353.30	10	0.00	33.70	41	0.78
WK16	230.46	10	0.00	23.85	44	0.99
WK17	234.24	10	0.00	35.36	45	0.85
WK18	434.63	10	0.00	41.77	43	0.52
WK19	168.63	10	0.00	56.70	45	0.11
WK20	105.38	10	0.00	42.01	45	0.60
WK21	422.87	10	0.00	43.71	44	0.48
WK22	158.81	10	0.00	56.69	42	0.06
WK23	1281.37	10	0.00	46.57	45	0.41
WK24	1159.70	10	0.00	63.24	44	0.03
WK28	42.21	10	0.00	47.43	46	0.41
WK29	352.04	10	0.00	40.18	44	0.64
WK30	422.35	10	0.00	42.60	44	0.53
WK31	2980.54	10	0.00	178.43	45	0.00
WK32	244.42	10	0.00	58.94	46	0.10
WK33	227.86	10	0.00	60.28	45	0.06

Table 4-3. Item Fit Statistics - Mathematics for Spring 2008

Item	X2	df	p	SX2	df_SX2	p_SX2
AM01	**********	9	0.00	110.51	95	0.13
AM02	2144.66	10	0.00	77.02	97	0.93
AM03	2913.81	9	0.00	115.7	97	0.09
AM04	**********	9	0.00	117.7	93	0.04
AM05	**********	9	0.00	110.48	97	0.17
AM06	**********	9	0.00	116.67	100	0.12
AM07	9638.59	9	0.00	139.44	95	0.00
AM08	**********	9	0.00	105.57	94	0.20
AM09	**********	9	0.00	95.69	97	0.52
AM10	8766.03	10	0.00	179.18	101	0.00
AM11	**********	9	0.00	149.19	97	0.00
AM12	7668.45	9	0.00	121.63	96	0.04
AM13	1540.28	10	0.00	109.46	104	0.34
AM14	**********	10	0.00	145.43	102	0.00
AM15	9389.03	10	0.00	132.45	104	0.03
AM16	7083.64	9	0.00	102.7	94	0.25
AM17	**********	10	0.00	108.24	100	0.27
AM18	**********	9	0.00	148.29	99	0.00
AM19	6133.64	10	0.00	110.59	100	0.22
AM20	**********	9	0.00	110.81	95	0.13
AM21	**********	9	0.00	98.94	95	0.37
AM22	**********	9	0.00	115.16	97	0.10
AM23	**********	9	0.00	116.07	99	0.12
AM24	6507.43	9	0.00	123.17	100	0.06
AM25	9873.51	10	0.00	85.75	100	0.84
AM26	**********	9	0.00	97	100	0.57
AM27	5600.17	10	0.00	116.22	106	0.23
AM28	9853.37	10	0.00	98.75	99	0.49
AM29	5797.33	10	0.00	96.51	107	0.76
AM30	**********	10	0.00	184.2	103	0.00
AM31	**********	9	0.00	117.01	100	0.12
AM32	5664.62	10	0.00	136.94	105	0.02
AM33	**********	9	0.00	78.67	100	0.94
AM34	2882.98	10	0.00	129.62	106	0.06
AM35	**********	9	0.00	140.85	100	0.00
AM36	**********	10	0.00	129.09	105	0.06
AM37	9634.06	10	0.00	99.82	104	0.60
AM38	3906.11	10	0.00	101.7	104	0.55
AM39	**********	9	0.00	127.51	97	0.02
AM40	**********	9	0.00	96.79	97	0.49
AM41	8959.25	9	0.00	107.59	102	0.33
AM42	6809.29	10	0.00	98.97	103	0.59
AM43	**********	9	0.00	89.44	99	0.74
AM44	**********	10	0.00	148.33	104	0.00
AM45	1577.76	10	0.00	126.46	109	0.12
AM46	4782.25	10	0.00	116.04	107	0.26
AM47	6107.61	10	0.00	110.52	102	0.27
AM48	**********	10	0.00	113.72	105	0.26
AM49	3074.04	10	0.00	134.27	107	0.04
AM50	**********	9	0.00	83.98	103	0.91
AM51	3040.63	10	0.00	172.79	107	0.00
AM52	7604.67	10	0.00	100.19	107	0.67
AM53	9629.71	10	0.00	125.67	106	0.09
AM54	9641.52	10	0.00	116.36	106	0.23
AM55	4953.12	10	0.00	120.23	107	0.18
AM56	6384.52	10	0.00	97.41	106	0.71
AM57	**********	10	0.00	101.26	105	0.59
AM58	8367.62	10	0.00	94.89	106	0.77
AM59	2094.07	10	0.00	102.34	109	0.66
AM60	4610.96	10	0.00	112.38	107	0.34
WK01	304.26	7	0.00	30.01	64	1.00
WK02	**********	9	0.00	448.05	81	0.00
WK03	228.45	8	0.00	11.27	65	1.00
WK04	**********	10	0.00	2179.35	111	0.00
WK05	338.60	8	0.00	46.33	72	0.99
WK06	**********	10	0.00	2752.94	112	0.00
WK07	5491.81	9	0.00	79.53	91	0.80
WK08	1373.28	9	0.00	61.58	87	0.98
WK09	924.55	8	0.00	56.4	76	0.96
WK10	2700.94	9	0.00	82.93	87	0.60
WK11	3473.47	9	0.00	84.42	93	0.73
WK12	2537.77	9	0.00	79.4	88	0.73
WK14	5615.61	9	0.00	98.88	92	0.29
WK15	1567.13	10	0.00	94.79	100	0.63
WK16	8519.82	9	0.00	119.05	97	0.06
WK17	2573.48	10	0.00	119.99	102	0.11
WK18	**********	9	0.00	98.87	94	0.35
WK19	6315.81	10	0.00	108.19	103	0.34

Item	X2	df	p	SX2	df_SX2	p_SX2
WK20	**********	9	0.00	583.35	97	0.00
WK21	**********	9	0.00	110.88	101	0.24
WK22	7016.01	10	0.00	111.4	101	0.23
WK23	**********	9	0.00	155.59	93	0.00
WK24	**********	9	0.00	90.86	93	0.54
WK25	**********	8	0.00	77.1	92	0.87
WK27	3289.84	10	0.00	114.51	107	0.29
WK28	**********	10	0.00	126.3	103	0.06
WK29	9972.94	10	0.00	175.24	102	0.00
WK30	4902.88	10	0.00	102.29	107	0.61
WK31	**********	10	0.00	130.22	104	0.04
WK32	6894.90	10	0.00	99.18	107	0.69
AS01	6633.69	10	0.00	100.66	100	0.46
AS02	4945.42	10	0.00	89.87	102	0.80
AS03	3991.87	10	0.00	113.61	107	0.31
AS04	9138.11	10	0.00	113.77	105	0.26
AS05	8033.59	10	0.00	111.46	106	0.34
AS06	6618.22	10	0.00	79.5	99	0.93
AS07	3441.52	10	0.00	138.78	107	0.02
AS08	7790.16	10	0.00	92.2	104	0.79
AS09	1715.24	10	0.00	127.17	109	0.11
AS10	8258.83	10	0.00	107.13	107	0.48
AS11	4737.96	10	0.00	80.95	104	0.95
AS12	6704.04	10	0.00	107.4	106	0.44
AS13	4958.23	10	0.00	100.02	107	0.67
AS14	3808.53	10	0.00	106.83	108	0.51
AS15	2582.09	10	0.00	96.79	109	0.79
MI01	1237.23	10	0.00	88.71	97	0.71
MI02	**********	9	0.00	86.82	96	0.74
MI03	5892.43	9	0.00	89.13	99	0.75
MI04	**********	9	0.00	114.59	102	0.19
MI05	3500.52	10	0.00	115.71	109	0.31
MI06	2566.59	10	0.00	125.01	107	0.11
MI07	**********	10	0.00	98.3	102	0.59
MI08	5667.47	10	0.00	113.22	106	0.30
MI09	9909.59	10	0.00	113.43	104	0.25
MI10	5688.12	10	0.00	82.02	107	0.97
MI11	6304.87	10	0.00	100.32	103	0.56
MI12	5865.21	10	0.00	93.15	106	0.81
MI13	5303.56	10	0.00	100.57	102	0.52
MI14	**********	10	0.00	136.02	105	0.02

Item	X2	df	p	SX2	df_SX2	p_SX2
MI15	6887.88	10	0.00	111.73	102	0.24
MI16	8128.87	10	0.00	76.22	101	0.97
MI17	**********	9	0.00	86.47	98	0.79
MI23	7759.07	9	0.00	92.8	100	0.68
MI24	8920.87	10	0.00	115.13	105	0.23
MI25	3283.31	10	0.00	99.63	104	0.60
MI26	**********	10	0.00	81.36	104	0.95
MI27	7240.84	10	0.00	105.35	104	0.44
MI28	8451.01	9	0.00	89.99	99	0.73
MI29	5904.83	10	0.00	89.52	106	0.88
MI30	8245.65	10	0.00	89.52	101	0.79
MI31	1796.89	10	0.00	101.23	100	0.45
MI32	6505.29	10	0.00	111.94	105	0.30
MI33	6801.86	10	0.00	74.54	106	0.99
MI34	**********	10	0.00	83.55	105	0.94
MI35	**********	10	0.00	78.33	105	0.98
MX01	542.55	10	0.00	92.58	105	0.80
MX02	633.31	10	0.00	134.11	107	0.04
MX03	312.56	10	0.00	120.34	106	0.16
MX04	157.17	10	0.00	109.21	110	0.50
MX05	1091.49	10	0.00	76.06	106	0.99
MX06	601.51	9	0.00	115.59	106	0.25
MX07	838.06	10	0.00	120.53	107	0.18
MX08	298.85	10	0.00	99.05	109	0.74
MX09	2190.49	9	0.00	84.63	103	0.91
MX10	765.95	9	0.00	89.7	101	0.78
MX11	251.96	10	0.00	105.82	108	0.54
MX12	3106.80	9	0.00	133.33	105	0.03
MX13	1298.65	10	0.00	96.63	106	0.73
MX14	1830.03	9	0.00	101.8	105	0.57
MX15	371.86	10	0.00	109.21	105	0.37
MX16	1244.87	10	0.00	92.74	106	0.82
MX17	1292.84	9	0.00	96.75	104	0.68
MX18	1514.86	10	0.00	106.5	105	0.44
MX19	1327.08	10	0.00	111.24	107	0.37
MX20	1616.11	9	0.00	111	101	0.23
MX21	860.37	10	0.00	98.39	109	0.76
MX22	1002.11	9	0.00	93.36	105	0.78

Table 4-4. Item Fit Statistics - Science for Spring 2008

Item	X2	df	p	SX2	df_SX2	p_SX2
AS01	2071.68	10	0.00	69.96	58	0.14
AS02	845.90	10	0.00	53.53	60	0.71
AS03	483.65	10	0.00	80.03	60	0.04
AS04	429.13	10	0.00	74.41	59	0.09
AS05	451.97	10	0.00	77.49	62	0.09
AS06	1235.49	10	0.00	87.13	63	0.02
AS07	1195.53	10	0.00	56.74	62	0.67
AS08	1388.33	10	0.00	83.46	62	0.04
AS09	812.06	10	0.00	86.14	60	0.02
AS10	122.38	10	0.00	40.52	61	0.98
AS11	448.23	10	0.00	86.99	60	0.01
AS12	111.74	10	0.00	75.68	63	0.13
AS13	786.03	10	0.00	95.07	63	0.01
AS14	800.33	10	0.00	117.59	58	0.00
AS15	908.29	10	0.00	52.61	58	0.68
AS16	259.92	10	0.00	41.9	60	0.96
AS17	192.59	10	0.00	64.93	59	0.28
AS18	265.65	10	0.00	103.18	63	0.00
AS19	145.01	10	0.00	58.86	62	0.59
AS20	5241.48	10	0.00	154.96	60	0.00
AS21	487.40	10	0.00	72.19	62	0.18
AS22	344.39	10	0.00	52.24	63	0.83
AS23	655.53	10	0.00	94.71	63	0.01
AS24	1915.41	10	0.00	113.33	63	0.00
AS25	213.25	10	0.00	63.56	61	0.39
AS26	757.06	10	0.00	41.04	60	0.97
AS27	642.77	10	0.00	57.55	63	0.67
AS28	945.04	10	0.00	75.44	63	0.14
AS29	1551.97	10	0.00	88.74	66	0.03
AS30	584.65	10	0.00	64.21	63	0.43
AS31	718.41	10	0.00	55.5	60	0.64
AS32	788.04	10	0.00	85.37	63	0.03
AS33	434.84	10	0.00	58.75	62	0.59
AS34	218.41	10	0.00	84.83	64	0.04
AS35	463.06	10	0.00	74.05	63	0.16
AS36	849.56	10	0.00	65.24	63	0.40
AS37	799.49	10	0.00	77.78	63	0.10
AS38	3823.25	10	0.00	141.45	63	0.00
AS39	204.63	10	0.00	66.85	63	0.35
AS40	483.04	10	0.00	84.89	65	0.05
MS01	312.47	10	0.00	56.67	63	0.70

Item	X2	df	p	SX2	df_SX2	p_SX2
MS02	508.83	10	0.00	62.55	64	0.53
MS03	259.83	10	0.00	55.23	62	0.72
MS04	411.11	10	0.00	60.52	59	0.42
MS05	324.15	10	0.00	75.62	61	0.10
MS06	98.49	10	0.00	40.76	63	0.99
MS07	502.30	10	0.00	67.4	58	0.19
MS08	334.17	10	0.00	61.87	60	0.41
MS09	417.14	10	0.00	47.17	61	0.90
MS10	460.40	10	0.00	76.04	64	0.14
MS11	859.12	10	0.00	61.98	61	0.44
MS12	2005.65	10	0.00	80.16	60	0.04
MS13	142.12	10	0.00	49.11	63	0.90
MS14	387.21	10	0.00	76.74	64	0.13
MS15	189.65	10	0.00	81.12	63	0.06
MS16	136.03	10	0.00	68.16	63	0.31
MS23	159.02	10	0.00	48.28	63	0.91
MS24	167.67	10	0.00	58.79	64	0.66
MS25	427.71	10	0.00	54.55	59	0.64
MS26	211.70	10	0.00	47.69	63	0.92
MS27	312.72	10	0.00	62.28	63	0.50
MS28	420.97	10	0.00	54.84	63	0.76
MS29	1203.56	10	0.00	151.12	62	0.00
MS30	302.72	10	0.00	69.84	62	0.23
MS31	284.39	10	0.00	69.08	63	0.28
MS32	488.61	10	0.00	67.7	62	0.29
MS33	224.03	10	0.00	59.06	62	0.58
MS34	1862.63	10	0.00	96.43	64	0.01
MS41	633.65	10	0.00	83.99	63	0.04
MS42	71.32	10	0.00	62.88	64	0.52
MS43	1515.17	10	0.00	123.1	64	0.00
MS44	134.53	10	0.00	57.68	64	0.70
MS45	522.41	10	0.00	73.63	62	0.15
MS46	708.44	10	0.00	60.79	64	0.59
MS47	878.50	10	0.00	91.24	62	0.01
MS48	332.15	10	0.00	59.15	60	0.51
MS49	198.06	10	0.00	66.75	63	0.35
MS50	158.45	10	0.00	68.08	63	0.31
MS51	262.97	10	0.00	40.21	62	0.99
MS52	319.54	10	0.00	70.94	60	0.16
MS53	400.51	10	0.00	73.53	62	0.15
MS54	392.43	10	0.00	52.21	60	0.75

Figure 4-1. Item Characteristic Curves – Writing

Spring 2008: 75 ACT English MC items + 1 ACT CR item + 1 Michigan CR item

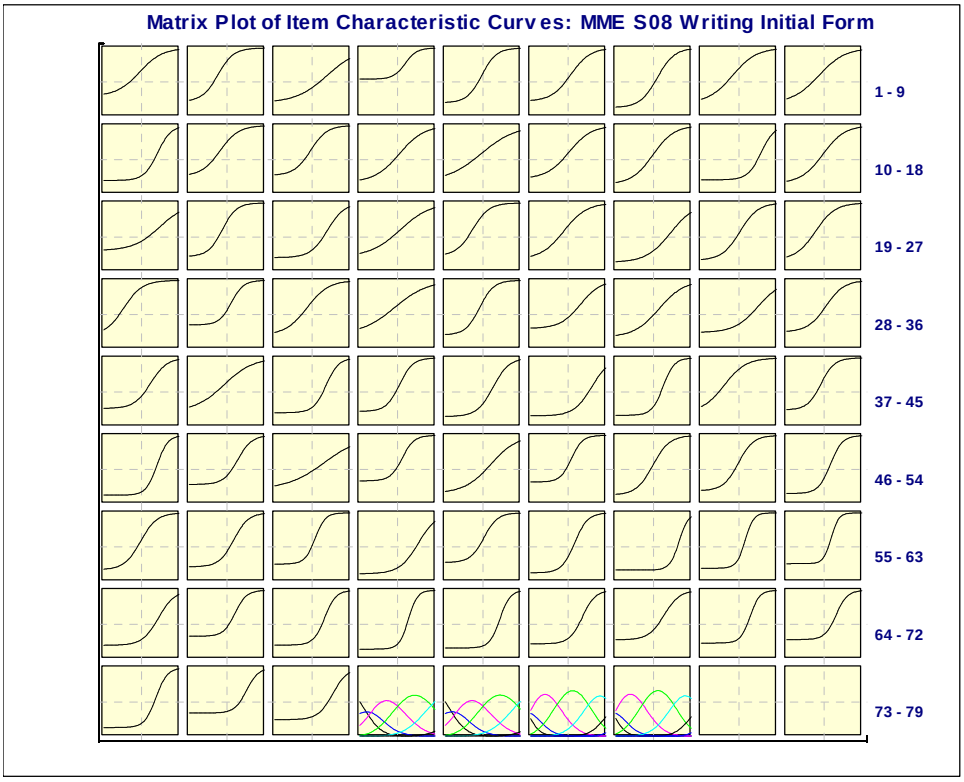


Figure 4-2. Item Characteristic Curves – Reading

Spring 2008: 40 ACT reading items + 30 WK reading items

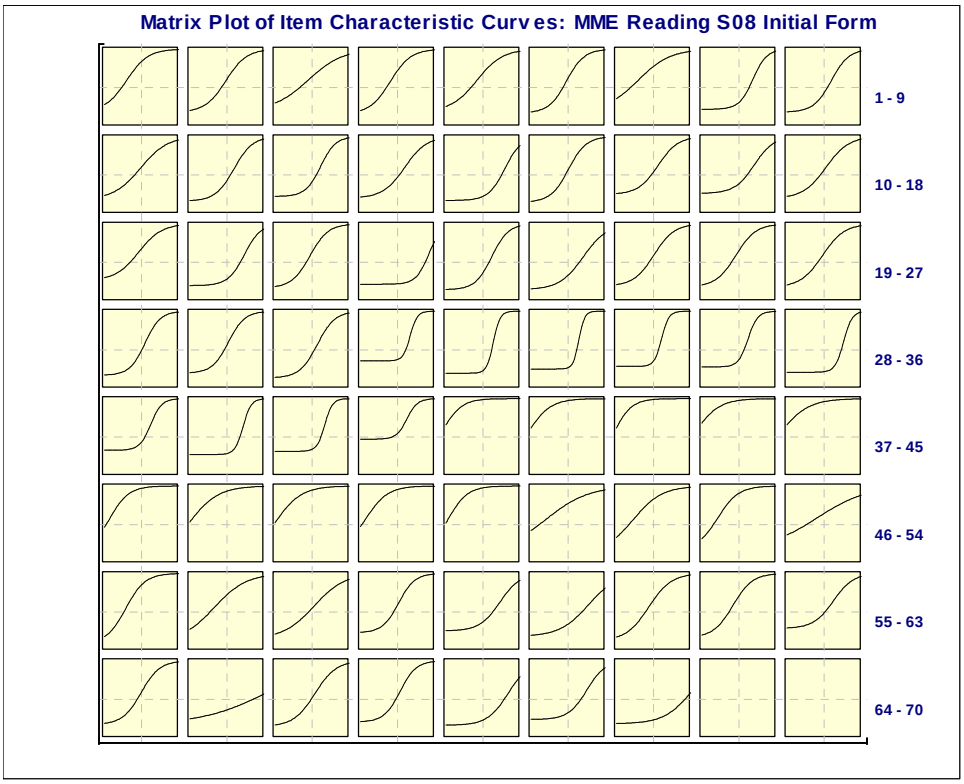


Figure 4-3. Item Characteristic Curves - Mathematics

Spring 2008: 60 ACT math items + 30 WK math items + 15 ACT science items + 52 Michigan math items

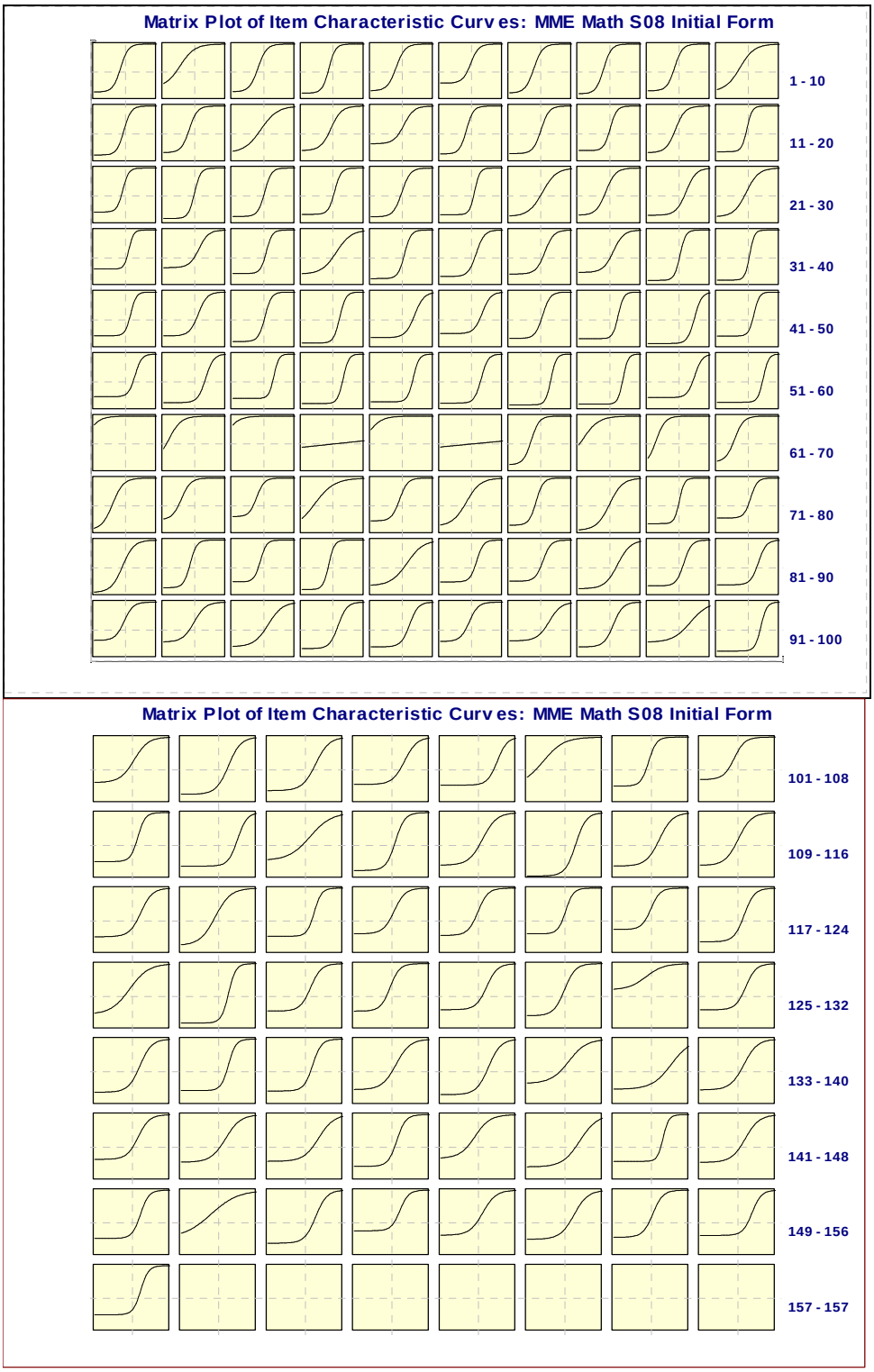


Figure 4-4. Item Characteristic Curves – Science
Spring 2008: 40 ACT science items + 42 Michigan science items

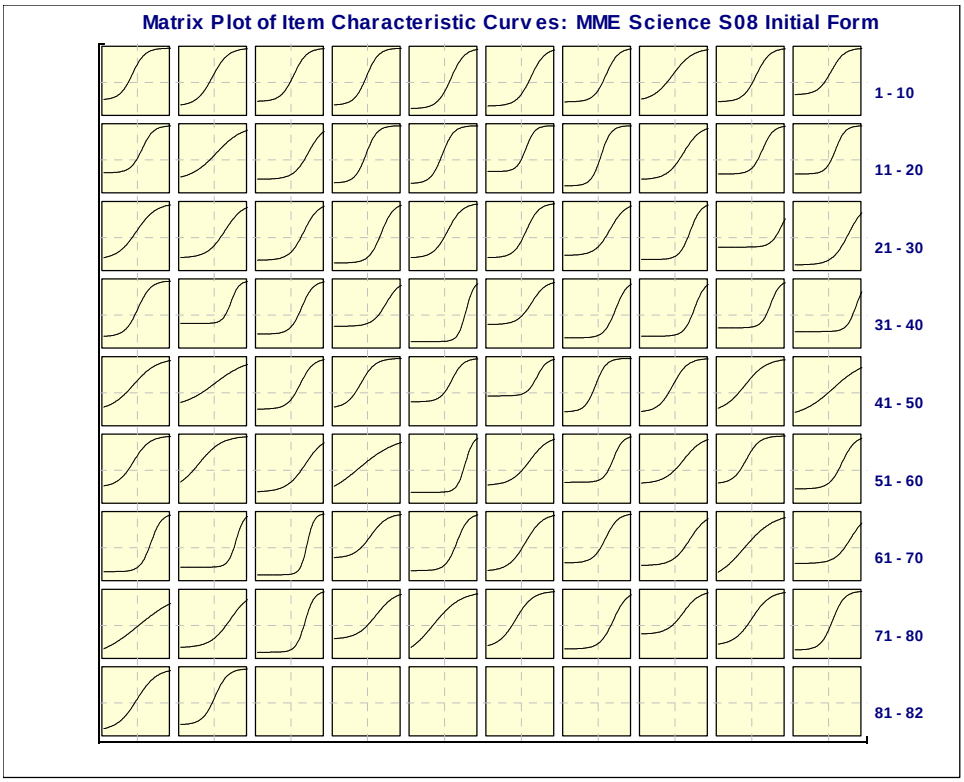


Table 4.5. Item Fit Statistics – Social Studies

Item	INFIT MNSQ	OUTFIT MNSQ
SocS01	1.17	1.50
SocS02	1.04	0.97
SocS03	1.10	1.16
SocS04	0.95	0.76
SocS05	1.06	1.09
SocS06	0.93	0.93
SocS07	1.00	1.01
SocS08	1.01	0.97
SocS09	1.10	1.15
SocS10	1.00	1.01
SocS11	1.06	1.06
SocS12	1.08	1.13
SocS13	0.97	0.95
SocS14	1.05	1.06
SocS15	1.09	1.13
SocS16	1.01	1.11
SocS17	1.08	1.12
SocS18	0.98	1.02
SocS19	1.06	1.07
SocS20	1.03	1.09
SocS21	0.99	1.00
SocS22	0.99	1.16
SocS23	0.97	0.92
SocS24	0.95	0.94
SocS25	1.03	1.02
SocS26	0.95	0.96
SocS27	0.88	0.82
SocS28	1.01	1.03
SocS29	1.06	1.09
SocS30	1.06	1.09
SocS31	1.12	1.24
SocS32	0.90	0.88
SocS33	1.01	1.01
SocS34	0.86	0.79
SocS35	1.09	1.12
SocS36	0.96	0.93
SocS37	1.00	0.99
SocS38	0.95	0.93
SocS39	0.84	0.58
SocS40	1.01	1.02
SocS41	1.07	1.09
SocS42	1.02	1.10
SocS43	0.95	0.92
SocS44	0.89	0.76
SocS45	0.98	0.99
SocS46	0.95	0.91
SocS47	0.90	0.90
SocS48	0.91	0.90

Chapter 5: Scaling and Equating

Quality control protocols

The following quality control (QC) tasks were implemented for MME calibrations. For the MME test subjects of Writing, Mathematics, Reading and Science, the MME calibration runs were conducted using PARSCALE (Muraki & Bock, 1997) under the 3PL model for dichotomous items and the generalized partial credit model for constructed response items. For calibrating MME Social Studies, the Rasch partial credit model was employed.

1. A thorough review of the test maps for Michigan-developed tests and WorkKeys was conducted including the following activities:
   - Cross-checks on fields/variables regarding items (such as item code and item key) provided on the test map.
   - Cross-reference of test positions for scrambled versions.
   - Checks on field test items (e.g., test positions, same field test items occurring on multiple forms).
   - Each updated test map for Michigan-developed tests provided on the PEM/ACT ftp site was reviewed.
2. The linking items were also reviewed and verified. Specifically, the information regarding linking items from the test maps, the new and old test booklets were compared word by word to ensure that there were no differences in linking items from one form to the next.
3. Files containing the item parameter estimates of ACT, WorkKeys, and Michigan linking items were prepared for review. The file naming conventions for such files were developed in advance. The values of the item parameter estimates and the test positions on the new and old forms were checked by test subject and form.
4. To facilitate creation of the PARSCALE and WINSTEPS control files, the 0/1 score data layout was created in advance. The positions for the 0/1 scores in the calibration data files were double-checked.
5. As a preliminary check on the calibration data file, SAS analyses were implemented to produce N-counts, classical item statistics, as well as frequency distributions on form codes, total raw scores, and scores for CR items. These analyses were examined for strange results, outliers, and so forth.
6. To review the calibration results, the following tasks were implemented:
   - Check convergence for each calibration run.
   - Compare classical item statistics produced by PARSCALE runs with those produced from SAS calculations, for an exact match.
   - Check parameter estimates for the discrimination parameter. There should be no negative values.
   - Compute correlation coefficients between p-value and b parameter estimates for reasonableness. The p-values and b parameter estimates should be negatively correlated. Examine the scatter plot of p-values versus b parameter estimates for outliers.
   - Check c parameter estimates for unusually large values, with the understanding that c-parameters interact with a- and b-parameters such that there may be some well-performing items with relatively large c-parameters where the empirical ICCs match the parameterized ICC well.
   - Review ICC plots produced by PARSCALE.

- Plot p-value vs. b parameter estimates.
- Check that fixed item parameter estimates have the correct values.
- Compare p-values for ACT items with those from the history to check that they look reasonably similar.
- Compare p-values for WorkKeys linking items with those from the history to check that they look reasonably similar.
- Compare p-values for Michigan linking items with those from the history to check that they look reasonably similar.
- For constructed response items, compare the item parameter estimates for the two raters to check that they look reasonable. The results indicated that no difficulty, discrimination, or step parameters differed by more than 0.01 across raters. Because the raters are randomly assigned as first and second raters, this is the expected outcome.

Results

Summary of Comparing the MLE Ability Estimates between PARSCALE and ISE

Upon successful PARSCALE calibration and OEAA's approval of item parameter estimates from MME forms (e.g., the initial, makeup or accommodation forms) for each MME test subject (i.e., Writing, Math, Reading or Science), PARSCALE runs with fixed parameter estimates were conducted to compute MLE thetas for MME calibration samples. A file containing IDs and MLE thetas produced by PARSCALE was uploaded to the PEM/ACT ftp cite for PEM's internal checks. For example, for mathematics in the Spring 2007 administration, the n-counts were 106,634, 1,792; and 1,918 for the initial, makeup and accommodation forms, respectively.

This summarizes the comparison of maximum likelihood estimation (MLE) of ability between the PARSCALE and ISE (IRT Score Estimation: developed by the PEM research group) computer programs. PARSCALE assigns values of 999 to a score of 0, a perfect score, and nonestimable score patterns when MLE is used. Note that PARSCALE outputs theta values to the fourth decimal place. ISE classifies the response patterns as normal case (unimodal), zero score, perfect score, mono-increasing case, mono-decreasing case, not converging under Newton-Raphson (NR), flat likelihood curve, or local maximum case. The NR method is used when the log-likelihood curve is unimodal. The user-specified max/min (-6,/6) thetas are assigned to the zero score, perfect score, mono-increasing cases and mono-decreasing cases. For cases such as the log-likelihood curve being flat or the NR method not converging, the grid search method (GS, also known as brute force) is used to find the MLE of the theta value. The GS algorithm divides the theta space into a grid, computes one value for each grid point, and chooses the best theta point (the one with the highest log-likelihood value within the specified range).

Comparison one (small sample size)

ACT provided Pearson Educational Measurement (PEM) with a sample of 500 students for Mathematics (117 items multiple choice, MC, items) and writing (75 multiple choice items and two open-ended, OE, items with two raters for each item) with the item parameters. PEM ran PARSCALE and ISE and compared the theta estimates of both programs. These results are presented in Table 5-1.

Table 5-1. Theta comparison between PARSCALE and ISE: Mathematics and Writing sample (500 students)

MLE Estimates	Mathematics	Writing
Exactly matched	457 (91.4%)	383 (76.6 %)
Parscale:999 (nonestimable) ISE: Estimable (not -6, +6)	26 (5.2%)	101 (20.2%)
Parscale:999 (nonestimable) ISE: +6 or -6	2 (0.4%)	9 (1.8%)
Difference = 0.0001	15 (3%)	7 (1.4%)

For mathematics, 91.4% cases were exactly matched, and 5.2% cases were not estimable by PARSCALE, but could be estimated by ISE. For writing, 76.6% cases were exactly matched, and 20.2% cases were not estimable by PARSCALE but could be estimated by ISE. The descriptive statistics for all the estimable thetas from both programs are given in Table 5-2. For practical purposes, these statistics are nearly equivalent.

Table 5-2. Descriptive Statistics for Theta Estimates between PARSCALE and ISE (sample file from ACT)

Content	Ability Estimates	N	Mean	STD	Min	Max
Mathematics	Parscale	472	-.04168	.08398	-5.9498	0.8777
	ISE	472	-.04168	.08398	-5.9498	0.8777
Writing	Parscale	390	-0.3129	.09088	-4.0362	0.9960
	ISE	390	-0.3129	.09088	-4.0362	0.9960

Comparison two (large sample size)

For an anonymous state, grade 5 Mathematics data with over 50,000 students, and 45 items including 42 MC items (40 3PL and 2 2PL) and 3 OE items with 5 score categories were used to evaluate the performance of the theta estimates for PARSCALE and ISE. As can be seen in Table 5-3, 98.4% of the cases were exactly matched and less than 1% cases were not estimable for PARSCALE. There are five cases that had estimates other than 999 by PARSCALE, but had the minimum theta values (-6) from ISE.

Table 5-3. Theta comparison between PARSCALE and ISE: X state grade 5 Mathematics data

MLE Estimates	Mathematics
Exactly matched	50877(98.4%)
Parscale: estimable ISE: Estimable (-6 or +6)	5 (0.01%)
Parscale:999 (nonestimable) ISE: +6 or -6	150 (0.29%)
Difference = 0.0001	662(1.28%)
Difference = 0.0002	1(0.0025%)

Table 5-4 presents the theta values and associated log-likelihood of these cases. Displayed in Figure 5.1 are the log-likelihood curves of the first four cases in Table 5-4. It can be seen that, except for the last case, those theta values from PARSCALE were local maxima not MLEs. That is, the log-likelihood values from ISE were larger than the log-likelihood from PARSCALE, so the theta estimates from PARSCALE were not MLE. For the last case (case ID 30286), although the PARSCALE theta has a higher log-likelihood value than ISE, it is beyond the specified range (-6, +6). If the theta range were set between -7 and 7, ISE should produce the same theta values as PARSCALE.

Table 5-5 shows that the descriptive statistics for all estimable thetas from both programs were very similar (the mean difference is 0.0002).

Table 5-4. Theta and log-likelihood of PARSCALE and ISE

Student ID	ISE		PARSCALE	
	Theta	Log-likelihood	Theta	Log-likelihood
45294	-6	-26.0683	-2.6526	-26.4811
48265	-6	-19.0879	-3.7871	-19.7349
33893	-6	-21.9597	-3.8493	-21.9604
42061	-6	-19.7264	-3.9583	-19.7349
30286	-6	-18.1164	-6.3719	-18.1138

Figure 5-1. The log-likelihood curves of the first four cases in Table 4

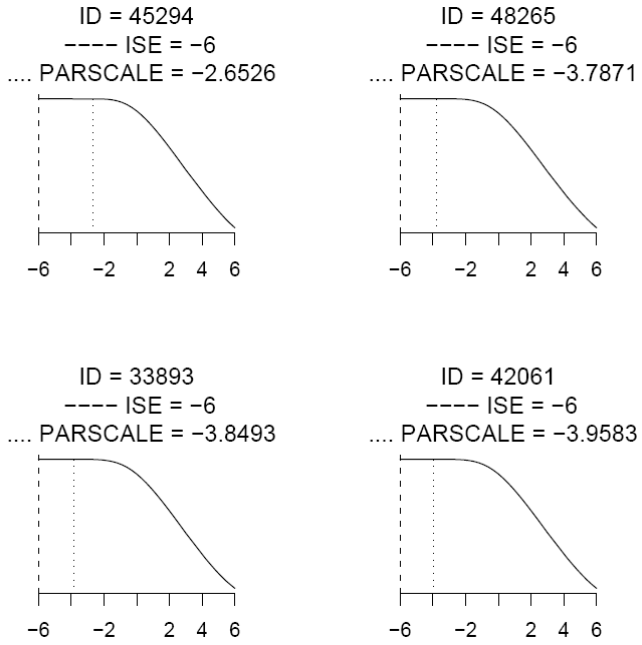


Table 5-5. Descriptive Statistics for Theta Estimates between PARSCALE and ISE: X state grade 5 Mathematics data

Ability Estimates	N	Mean	STD	Min	Max
Parscale	51545	0.1310	1.0055	-6.3719	3.5918
ISE	51545	0.1308	1.0063	-6.0000	3.5918

Conclusions

Overall, ISE and PARSCALE produce identical theta estimates for the normal cases. For the other six cases, the performance of ISE is better than that of PARSCALE in terms of the capability of estimating theta and providing theta estimates that have higher likelihood. The ISE program allows the user to specify the upper/lower boundaries within the range of real numbers. With the upper/lower boundaries specified, every score pattern is estimable by ISE.

Equating/linking/scaling for MME

The MME equatings for Writing, Mathematics, Reading, and Science use national performance data to scale the ACT using the 3-PL model, and fixes the ACT item parameters in calibrating/equating the entire MME. Because the ACT form does not change from the calibration run to the MME run, there should be no item ordering effects from the ACT portion of the MME assessment.

Michigan has chosen to use the National data sample to calibrate the ACT portion of the MME because it provides a highly stable calibration across forms. This high degree of stability may not be possible for MME forms administered to small samples, such as the Braille forms. Michigan is relying on the reasonable assumption that calibration with a larger set of students is more stable and accurate than calibration with a smaller set of students.

The WorkKeys and Michigan-developed components are calibrated (and equated) using the population of MME takers. The common items from the WorkKeys and Michigan-developed portions are included in the MME equating as fixed parameter items, but are also used to determine whether item context effects have occurred for reused items.

Equating for ACT

Several new forms of each of the ACT tests are developed each year. Even though each form is constructed to adhere to the same content and statistical specifications, the forms may differ slightly in difficulty. To control for these differences, subsequent forms are equated, and the scores reported to examinees are scale scores that have the same meaning regardless of the particular form administered to examinees. Thus, scale scores are comparable across test forms and test dates.

A carefully selected sample of examinees from one of the five national test dates each year is used as an equating sample. The examinees in this sample are administered a spiraled set of “n” forms—the new forms (“n – 1” of them) and one anchor form that has already been equated to previous forms. (The base form is the form used initially to establish the score scale.) The use of randomly equivalent

groups is an important feature of the equating procedure and provides a basis for confidence in the continuity of scales. More than 2,000 examinees take each form.

Scores on the new forms are equated to the score scale using an equipercentile equating methodology. In equipercentile equating, a score on Form X of a test and a score on Form Y are considered to be equivalent if they have the same percentile rank in a given group of examinees. The equipercentile equating results are subsequently smoothed using an analytic method described by Kolen (1984) to establish a smooth curve, and the equivalents are rounded to integers. The conversion Tables that result from this process are used to transform raw scores on the new forms to scale scores on the base form scale.

The equipercentile equating technique is applied to the raw scores of each of the four tests for each form separately. The composite score is not directly equated across forms. It is, instead, a rounded arithmetic average of the scale scores for the four equated tests. The subscores are also separately equated using the equipercentile method. Note, in particular, that the equating procedure does *not* lead to a given reported test score being equal to some prespecified arithmetic combination of subscores.

As specified in the *Standards for Educational and Psychological Testing* (APA, 1999), ACT conducts periodic checks on the stability of the ACT scores. The results appear reasonably sTable to date.

Equating for WorkKeys

New forms of the WorkKeys tests are developed as needed. Though each form is constructed to adhere to the same content and statistical specifications, the forms may be slightly different in difficulty. To control for these differences, scores on all forms are equated so that when they are reported to test takers (as either Level Scores or Scale Scores), equated scores have the same meaning regardless of the particular form administered. Thus, Level Scores and Scale Scores are comparable across test forms and test dates. However, they are not comparable across tests. A Level Score of 3 or a Scale Score of 25 in *Reading for Information* does not have the same meaning as a Level Score of 3 or a Scale Score of 25 on any other WorkKeys test (e.g., Applied Mathematics). Two common equating designs are used with the WorkKeys tests (Kolen & Brennan, 1995).

In a *randomly equivalent groups design*, new test forms are administered along with an anchor form that has already been equated to previous forms. A spiraling process is used to distribute test forms to test takers. For example, in each testing room the first person receives Form 1, the next Form 2, and the next Form 3. This pattern is repeated so that each form is given to one-third of the test takers and the forms are given to randomly equivalent groups. When this design is used, the difference in total-group performance on the new and anchor forms is considered a direct indication of the difference in difficulty between the forms. Scores on the new forms are equated to the score scale using various equating methodologies including linear and equipercentile procedures (e.g., see Kolen & Brennan, 1995). When the Level Score and Scale Score conversions are chosen for each form, the equating functions are examined, as are the resulting distributions of the scores and their means, standard deviations, skewnesses, and kurtoses.

A *common-item nonequivalent groups design* has been used when a spiraling technique cannot be implemented in a test administration, when only a single form can be administered per test date, or when some items are changed in a revised form. In a common-item nonequivalent groups design, the new

form and base form have a set of items in common. These common item sets (anchors) are chosen to represent the content and statistical characteristics of the test and are usually interspersed among the other items in the new test form. The different forms are then administered to different groups of test takers. In this design, the groups are not assumed to be equivalent. Observed differences between group performances can result from a combination of (a) test-taker group ability differences and (b) test form difficulty differences. The common items are used to control for group differences, so that adjustments can be made for form differences. Strong statistical assumptions are required to separate these groupd and form differences.

The various equating methods under the common-item nonequivalent groups design are distinguished in terms of their statistical assumptions (Kolen & Brennan, 1995). *Observed-score equating* methods are typically used in equating WorkKeys test forms. For each form, the equating functions are examined, as are the resulting distributions of scale scores and the mean, standard deviation, skewness, and kurtosis of the scale scores. The set of equating conversions chosen for each form is the one that results in scale score distributions and scale score moments that are judged to be reasonable based on the sample sizes, the magnitudes of the form differences and group differences, and the historical statistics for the test.

Equating for MME Social Studies

Social Studies in MME is the only subject using the Rasch Partial Credit Model (RPCM) to derive the scale score system for the MME. The RPCM, an extension of the Rasch model, accommodates the constructed response tasks associated with the multiple-choice items.

The Rasch Partial Credit Model (RPCM) is an extension of the Rasch one-parameter Item-Response Theory model attributed to Georg Rasch (1960), as extended by Wright and Stone (1979), and Wright and Masters (1982). The RPCM is used because of its flexibility in accommodating multiple-response category data and its ability to maintain a one-to-one relationship between the derived (i.e., scale) and the underlying raw score scale. The RPCM is defined via the following mathematical measurement model where, for a given item involving *m* score categories, the probability of person *n* scoring *x* on prompt *i* is given by:

$$P_{xni} = \frac{\exp \sum_{j=0}^{x} (B_n - D_{ij})}{\sum_{k=0}^{m_i} \exp \sum_{j=0}^{k} (B_n - D_{ij})},$$

where x = 0, 1, 2, . . . , *m*, and

$$\sum_{j=0}^{0} (B_n - D_{ij}) = 0.$$

The RPCM provides the probability of a person scoring *x* on the m_i step of task *i* as a function of the person's ability (B_n) and the step difficulties of the *m* steps in task *i*. The item calibration and proficiency estimates are performed using the Rasch Partial Credit Model and procedures implemented in WINSTEPS version 3.33. The statistical elements of the calibrating/scaling process are referred to as Rasch Calibration/Scaling as described in the WINSTEPS manual.

The scaling design is referred to as a common item nonequivalent groups design (Kolen & Brennan, 2004). Each year, new test forms are built based on the test blueprint and available statistical information from field testing in previous years. New field-testing items are embedded in test forms for building and replenishing the item pool. For 2007, there were eight forms for social studies. A sparse matrix that included all the scored items is created and a concurrent calibration was applied. Anchored items in the new forms were then used to scale all items to the MME scale.

ACT follows most calibration and scaling rules/procedures that Pearson used before. However, according to the most recent document of "Attemptedness Table v5.0," the score of the constructed response item will be changed to be sum of the scores from two raters. For a constructed response item, the scores from the two raters are treated like scores from two independent items. The two scores are considered simultaneously when examinees' proficiency is estimated, and they are summed together in the final report.

Specific Steps for Equating of Social Studies:

1. Review test maps and obtain item parameters from item pool for anchored items
2. Review test irregularity reports and clean data for item calibration and equating
3. Check the parameter stability of anchored items
4. Run operational item calibration with fixed anchored items using Winsteps (version 3.63)
5. Review calibration results
6. Create a raw-to-scale score conversion Table for scoring
7. Run FT item calibration using Winsteps
8. Review FT item calibration results for future form construction and linking

Equating for MME Writing, Reading, Mathematics, Science

Depending on the MME test subject (Writing, Reading, Mathematics and Science), an MME test can consist of up to three components: items from one or two of the four ACT tests, one of two WorkKeys tests, and a Michigan-developed test for that subject. To develop the MME scale, an MME base form was administered in the spring 2006 Baseline Study.

The item parameter estimates for all ACT forms administered in the spring 2008 MME were separately calibrated under the three parameter logistic model using the ACT equating samples discussed previously and then placed on the MME scale using the Stocking-Lord characteristic curve method (Stocking & Lord, 1983). To link the WorkKeys and Michigan-developed test forms, respectively, to the MME base form, a set of anchor items that were common to the 2007 or 2006 forms were employed.

The MME calibration runs were conducted using PARSCALE (Muraki & Bock, 1997) under the generalized partial credit model for constructed response items and the three parameter logistic model for dichotomous items. For the spring 2007 MME administration, a concurrent calibration run for the various components was implemented with fixed item parameter estimates for the ACT items, fixed item parameter estimates for the WorkKeys anchor items, and fixed item parameter estimates for the Michigan anchor items with all others being placed on the MME scale by the calibration run. As scrambled versions of the Michigan-developed forms are used for different testing situations, (i.e., initial, makeup and accommodated), the item parameter estimates for Michigan-developed items were obtained from a master initial calibration run using the data for the initial forms for all of the various MME components. These calibration analyses were based on the assumption that the sample size for the master initial run is the largest, and the IRT assumption that item location does not affect item

parameters. Under the IRT assumption of group invariance, these item parameters were fixed for the calibration runs for other form combinations.

Specific steps for equating MME Writing, Mathematics, Reading and Science

1. Review test maps
2. Obtain item parameter estimates from the pool for anchor items
    - For testing forms with small N-counts (e.g., Braille or emergency), item parameter estimates obtained from master initial calibration runs are employed
    - For testing forms that are a scrambled version of the initial form, item parameter estimates of the initial form are used
3. Review test irregularity reports and create data sets for calibration and equating
4. Check anchor item parameter stability
5. Conduct fix-parameter calibration runs using PARSCALE without field test items
6. Evaluate calibration results of operational items and pass item parameter estimates for MME scoring
7. Run PARSCALE to calibrate field test items with item parameter estimates of all operational items being fixed
8. Review calibration results of field test items for future form construction considerations and linking

Equating for MME ELA

MME ELA is not separately equated; it is the average of two separately equated components, MME Writing and MME Reading.

Chapter 6: Reliability

SEM/information curves with cuts scores (imposed)

Appendix A exhibits the plots of SEM/information curves produced by PARSCALE with the MME cut scores imposed for the testing subjects of Writing, Reading, Mathematics and Science, respectively. The vertical lines represent the performance level cut scores. For spring 2008, the performance levels were Not Proficent, Partially Proficient, Proficient, and Advanced. Although the labels were changed, the cut scores were the same.

Internal Consistency Reliability

Based on the raw scores, the alpha coefficients (Cronbach's alpha) are found to be 0.94 for Writing, 0.89 for Reading, 0.93 for Mathematics, 0.92 for Science, and 0.88 for Social Studies for the 2008 spring MME administration. Table 6-1 presents the percentage of agreement of the two raters on the constructed response items. For the spring 2008 administration, over 95,000 examinees were in the reliability analysis dataset, depending on the content area.

Table 6-1. Spring 2008 Rater Reliability Scores of Agreement

Absolute difference in scores between two raters	ACT essay	Michigan essay for ELA	Social Studies
0	72.91	64.59	64.72
1	26.42	35.34	33.56
2	0.43	0.07	1.49
3 or higher	0.01	0.00	0.03

Empirical IRT Reliability

Scale scores (theta):

For the IRT methods, the conditional standard error of measurement (CSEM) is computed as part of the item parameter estimation process, via the test information function. Although these computed CSEMs are on the IRT theta scale, they can be placed on the MME scale score scale. The MME scale score is a linear function of the IRT theta scale. Therefore, the CSEM from the IRT theta scale can be placed on the MME scale by multiplying appropriate constants. Once the mean squared CSEM over examinees is computed, the equation below can be used to compute the reliability. In reference to this equation, $\bar{\sigma}^2(E)$ is the mean squared CSEM and $\sigma^2(S)$ is the observed variance of scale scores for the test taken over examinees. For the 2008 spring MME administration, the values of the empirical IRT reliability estimates were found to be 0.93, 0.90, 0.94, 0.85 and 0.89 for Writing, Reading, Mathematics, Science and Social Studies, respectively.

$$rel = 1 - \frac{\bar{\sigma}^2(E)}{\sigma^2(S)}$$

Classification Consistency and Classification Accuracy

Classification consistency indices quantify the reliability of categorizing examinees into mastery or achievement levels, with respect to specific standards. Several model-based approaches have been developed for estimating classification consistency for a single test administration because repeated testing data are seldom available. An IRT model-based approach (Lee, Hanson, & Brennan, 2002) is used in this technical report to calculate the agreement index, *P*.

Assuming the two raw score random variables X_1 and X_2 from two administrations of a test are independent and identically distributed, the conditional joint distribution of X_1 and X_2 is given by $f(x_1, x_2 \mid \theta) = f(x_1 \mid \theta) f(x_2 \mid \theta)$, where θ denotes true examinee ability. Then, the marginal joint distribution of X_1 and X_2 can be obtained by integrating the conditional probabilities over the distribution of θ as

$f(x_1, x_2) = \int f(x_1, x_2 \mid \theta) g(\theta) d\theta$.

A consistent classification is made if both x_1 and x_2 for an examinee belong to the same category I_h (h=1, 2, …, H). The conditional probability of falling in the same category on the two testing occasions is

$$Pr(X_1 \in I_h, X_2 \in I_h \mid \theta) = \left[\sum_{x_1 = c_{(h-1)}}^{c_h - 1} f(x_1 \mid \theta) \right]^2,$$

where c_1, c_2, …, $c_{(H-1)}$ are raw cutoff scores, c_0 is the lowest raw score, and c_H is a perfect test score. Then, the agreement index *P* conditional on θ is obtained by

$$P(\theta) = \sum_{h=1}^{H} Pr(X_1 \in I_h, X_2 \in I_h \mid \theta),$$

and the marginal values of agreement index can be computed by

$$P = \int P(\theta)g(\theta)d\theta .$$

For each MME assessment, there are three cutoff score points and four categories at the scale-score level. Since there are four categories, examinees are classified into one of the four mutually exclusive categories based on their scale scores and the cutoff points on the MME assessment. To estimate classification consistency, however, 4 × 4 contingency Tables for the MME assessment are created using the psychometric model, with the columns and rows showing the four classification categories. The elements of the 4 × 4 Tables indicate the joint probabilities of examinees being classified in the pairs of the column and row categories; for example, being classified in the Basic level on one occasion (column) and in the Proficient Standards level on the other (row). The sums of the diagonal elements of the 4 × 4 Tables are the indices of classification consistency.

The data used to compute classification consistency reported in the first part of Table 6-2 were obtained from the MME tests administered in spring 2008. The 3 parameter logistic model and the generalized partial credit model are used to estimate classification index. The basic role of these IRT models is to estimate the theta distribution and predict the observed score distribution. Once these distributions are estimated, 4 × 4 contingency Tables can be created, which, in turn, are used as a basis for computing the classification index. Table 6-2 shows the 4 × 4 contingency Tables and indices of classification consistency for the MME assessments.

Table 6-2. The 4 × 4 contingency Table and classification consistency for the MME assessments for the Spring 2008 MME administration

MME Writing

	Not Proficient	Partially Proficient	Proficient	Advanced
Not Proficient	0.03468	0.02844	0.00000	0.00000
Partially Proficient	0.02844	0.40504	0.04795	0.00000
Proficient	0.00000	0.04795	0.36838	0.01012
Advanced	0.00000	0.00000	0.01012	0.01888

MME Reading

	Not Proficient	Partially Proficient	Proficient	Advanced
Not Proficient	0.09194	0.04380	0.00670	0.00000
Partially Proficient	0.04380	0.10081	0.06235	0.00000
Proficient	0.00670	0.06235	0.54166	0.01233
Advanced	0.00000	0.00000	0.01233	0.01521

MME Mathematics

	Not Proficient	Partially Proficient	Proficient	Advanced
Not Proficient	0.27765	0.04226	0.00230	0.00000
Partially Proficient	0.04226	0.09437	0.03829	0.00000
Proficient	0.00230	0.03829	0.31742	0.01769
Advanced	0.00000	0.00000	0.01769	0.10949

MME Science

	Not Proficient	Partially Proficient	Proficient	Advanced
Not Proficient	0.13470	0.04269	0.01277	0.00000
Partially Proficient	0.04269	0.05876	0.04941	0.00000
Proficient	0.01277	0.04941	0.50062	0.01739
Advanced	0.00000	0.00000	0.01739	0.06143

MME Social Studies

	Not Proficient	Partially Proficient	Proficient	Advanced
Not Proficient	0.12506	0.02691	0.00293	0.00000
Partially Proficient	0.02691	0.05937	0.03406	0.00008
Proficient	0.00293	0.03406	0.18284	0.03066

Advanced	0.00000	0.00008	0.03066	0.44344

Table 6-3 provides classification accuracy indices for the MME scales using an index based on estimated thetas and conditional standard errors. Classification accuracy evaluates the degree of accuracy of classifying examinees into score categories based upon observed scores. An expected classification accuracy index (Martineau, 2007) using measurement error is employed in this report. Let κ denote the vector of H+1 cut scores that divide the theta score scale into H categories, or $\kappa = [\kappa_1, \kappa_2, ..., \kappa_{H+1}]$ where $\kappa_1 < \kappa_2 < ... < \kappa_{H+1}$ and $\kappa_1 = -\infty, k_{H+1} = \infty$. For an examinee *i* with observed theta score $\hat{\theta}_i$ and standard error $SE_{\hat{\theta}_i}$, an expected probability that the student falling into the h_i performance level under the assumption of conditional normality of measurement error is defined as the area from κ_h to κ_{h+1} under the normal curve with mean $\hat{\theta}_i$ and standard deviation $SE_{\hat{\theta}_i}$. Let $\hat{p}_{ih_i} = \phi(\kappa_{h_i}, \kappa_{h_i+1}, \hat{\theta}_i, SE_{\hat{\theta}_i})$ represent this expected probability. Then, the expected classification accuracy index, based on measurement error, is equal to $\tau = \sum_{i=1}^{N} \phi(\kappa_{h_i}, \kappa_{h_i+1}, \hat{\theta}_i, SE_{\hat{\theta}_i}) / N$ where *N* is the number of examinees. This index ranges from 0 to 1, with 0 indicating no accuracy in examinee classifications, with 0.5 indicating random accuracy, and 1 indicating perfect expected accuracy in examinee classification.

Table 6-3. Classification accuracy indices for the MME assessments using four classification categories Spring 2008

Assessment	Index Value
Writing	**0.89**
Reading	**0.85**
Math	**0.87**
Science	**0.91**
Social Studies	**0.81**

Chapter 7: Validity

Validity refers to the extent to which a test measures what it is intended to measure and how well it does so. As stated in the *Standards for Educational and Psychological Testing* (1999), validity refers to the "degree to which evidence and theory support the interpretations of test scores entailed by the proposed uses of tests." This statement shows that test validation is an ongoing process, which begins the moment that work on a test begins and continues throughout the life of the test. Validity is the process of continually accumulating and reviewing evidence from various resources to refine the utility of a test for making recommended interpretations consistent with the intended uses and interpretations of the test scores.

Construct Validity Evidence from Content and Curricular Validity

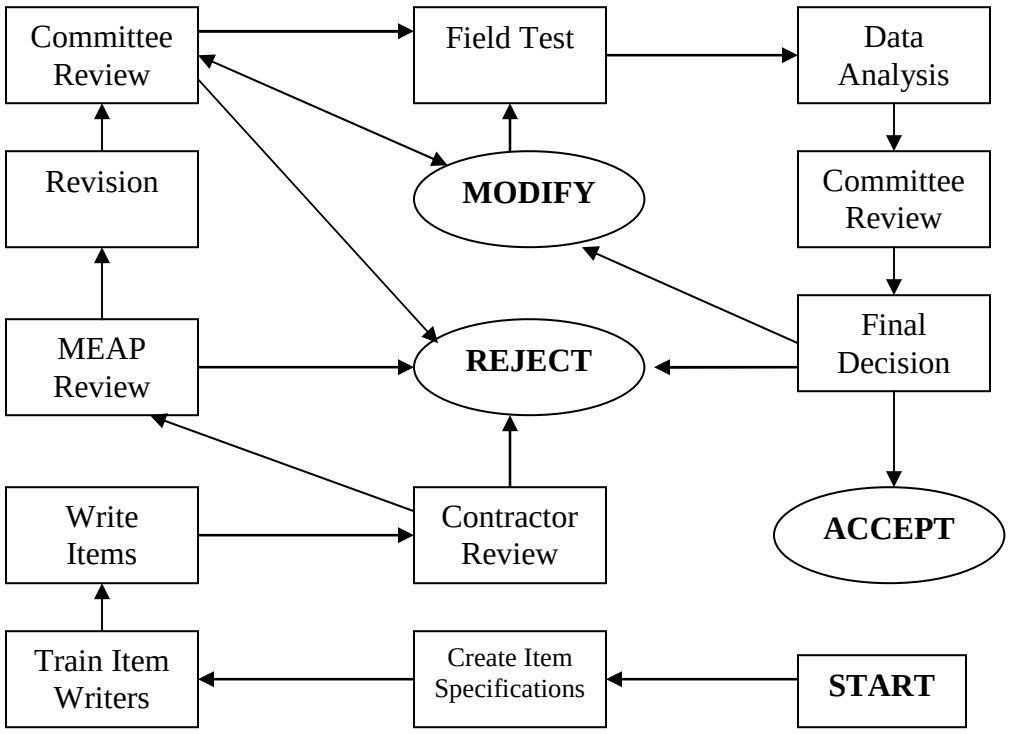
Content validity involves essentially the systematic examination of the test content to determine whether it covers the curricular standards to be measured. As stated in Chapter 1, the MME augmentation is developed to measure what Michigan educators believe all students should know and be able to achieve in the content areas that are not measured on the ACT and WorkKeys assessments. Assessment results paint a picture of how Michigan students and schools are doing when compared with standards established by the State Board of Education. The MME is based on an extensive definition of the content the test is intended to assess and its match to the content standards. Therefore, the MME assessments are content-based and aligned directly to the statewide content standards.

Relation to Statewide Content Standards

From before the inception of the MME, a committee of educators, item development experts, assessment experts, and OEAA staff met annually to review new and field-tested items for use on the MEAP (the old high school assessment) and for use in augmenting the MME. The OEAA has established a sequential review process, as illustrated in Figure 7-1. This process provides many opportunities for these professionals to offer suggestions for improving or eliminating items and to offer insights into the interpretation of the statewide content standards. These review committees participate in this process to ensure test content validity.

In addition to providing information on the difficulty, appropriateness, and fairness of these items, committee members provide a needed check on the alignment between the items and the content standards they are intended to measure. When items are judged to be relevant (i.e., representative of the content defined by the standards), this provides evidence to support the validity of inferences made (regarding knowledge of this content) with MME results. When items are judged to be inappropriate for any reason, the committee can either suggest revisions (e.g., reclassification or rewording) or elect to eliminate the item from the field-test item pool. Items that are approved by the content review committee are later embedded in live MME forms to allow for the collection of performance data. In essence, these committees review and verify the alignment of the test items with the objectives and measurement specifications to ensure that the items measure appropriate content. The nature and specificity of these review procedures provide strong evidence for the content validity of the MME.

Figure 7-1. Item Development/Review Cycle



MME Alignment Studies

As detailed in the chapter on item and test development, two alignment studies have been performed for the MME, documenting alignment of the overall set of items from the ACT, WorkKeys, and Michigan-developed augmentation to Michigan's content standards. These independent alignment studies provide validity evidence which is complementary to the input provided during content reviews. Along with the reliability analyses and other technical analyses, these alignment studies provide strong evidence of the validity of MME.

Educator Input

Michigan educators provide valued input on the MME content and the match between the items and the statewide content standards. In addition, many current and former Michigan educators and some educators from other states work as independent contractors to write items specifically to measure the objectives and specifications of the content standards for the MME. Using a varied source of item writers provides a system of checks and balances for item development and review that reduces single source bias. Because many people with various backgrounds write the items, it is less likely that items will suffer from a bias that might occur if items were written by a single author. This direct input from educators, many of whom serve on the aforementioned committees, offers evidence regarding the content validity of the MME.

Construct Validity Evidence from Criterion Validity

Criterion validity refers to the degree to which a test correlates with other external outcome criteria. Criterion validity addresses how accurately criterion performance can be predicted from test scores. The key to criterion-related evidence is the degree of relationship between the assessment and the outcome criterion. To ensure a good relationship between the assessment and the criterion, the criterion should be relevant to the assessment and reliable. As the ACT and WorkKeys are administered intact as a part of the MME, and there is a large body of evidence concerning their reliability and validity, there is a built in relevance of these criteria to the MME.

There is a large body of evidence from ACT that the ACT successfully predicts success in college, and the WorkKeys successfully predicts workplace success. As a criterion, the WorkKeys and ACT should be strongly correlated with the overall MME scores, indicating that the MME also can be used to predict college and workplace success.

The correlations among the old high school MEAP, the MME, the ACT, and WorkKeys from the Spring 2006 pilot are presented in Table 7-1. The cells reported in bold are the correlations between the ACT and the MME scores and the WorkKeys and MME scores. These correlations are very high correlations, and indicate that the MME should be approximately as effective in predicting workplace and college success as the ACT and WorkKeys assessments.

In addition, the correlations among the MME and old high school MEAP are strong, indicating that as expected, the assessments measure similar constructs.

Conclusion

The evidence from the methods used for item development, item review, augmentation, alignment, and correlation with related measures indicate a strong degree of validity for the MME.

Table 7-1. Correlations between MME and other related measures for the Spring 2006 pilot.

Correlations (based on 3306 students who had valid scores on all MME subjects)																			
			ELA								Mathematics				Science			Social Studies	
			English	Writing			Reading												
Subject			ACT	MME	ACT	MEAP	MME	ACT	WK	MEAP	MME	ACT	WK	MEAP	MME	ACT	MEAP	MME	MEAP
ELA	English	ACT	1.00	**0.96**	0.47	0.51	0.76	0.75	0.62	0.60	0.72	0.72	0.59	0.68	0.75	0.71	0.67	0.67	0.67
	Writing	MME	0.96	1.00	0.59	0.57	0.78	0.74	0.63	0.62	0.73	0.71	0.59	0.69	0.75	0.71	0.67	0.67	0.67
		ACT	0.47	0.59	1.00	0.52	0.44	0.42	0.34	0.39	0.40	0.39	0.29	0.38	0.39	0.41	0.34	0.35	0.35
		MEAP	0.51	0.57	0.52	1.00	0.47	0.44	0.38	0.46	0.43	0.40	0.34	0.44	0.43	0.41	0.40	0.41	0.41
	Reading	MME	0.76	0.78	0.44	0.47	1.00	**0.89**	**0.82**	0.60	0.69	0.64	0.60	0.62	0.74	0.68	0.66	0.68	0.68
		ACT	0.75	0.74	0.42	0.44	0.89	1.00	0.59	0.56	0.61	0.61	0.51	0.57	0.69	0.65	0.62	0.64	0.64
		WK	0.62	0.63	0.34	0.38	0.82	0.59	1.00	0.51	0.63	0.57	0.58	0.57	0.65	0.59	0.58	0.58	0.58
		MEAP	0.60	0.62	0.39	0.46	0.60	0.56	0.51	1.00	0.52	0.49	0.43	0.52	0.58	0.51	0.56	0.59	0.59
Mathematics		MME	0.72	0.73	0.40	0.43	0.69	0.61	0.63	0.52	1.00	**0.90**	**0.88**	0.84	0.81	0.77	0.71	0.66	0.66
		ACT	0.72	0.71	0.39	0.40	0.64	0.61	0.57	0.49	0.90	1.00	0.74	0.82	0.77	0.74	0.69	0.63	0.63
		WK	0.59	0.59	0.29	0.34	0.60	0.51	0.58	0.43	0.88	0.74	1.00	0.72	0.70	0.65	0.63	0.58	0.58
		MEAP	0.68	0.69	0.38	0.44	0.62	0.57	0.57	0.52	0.84	0.82	0.72	1.00	0.76	0.70	0.72	0.66	0.66
Science		MME	0.75	0.75	0.39	0.43	0.74	0.69	0.65	0.58	0.81	0.77	0.70	0.76	1.00	**0.89**	0.88	0.76	0.76
		ACT	0.71	0.71	0.41	0.41	0.68	0.65	0.59	0.51	0.77	0.74	0.65	0.70	0.89	1.00	0.67	0.65	0.65
		MEAP	0.67	0.67	0.34	0.40	0.66	0.62	0.58	0.56	0.71	0.69	0.63	0.72	0.88	0.67	1.00	0.73	0.73
Social Studies		MME	0.67	0.67	0.35	0.41	0.68	0.64	0.58	0.59	0.66	0.63	0.58	0.66	0.76	0.65	0.73	1.00	1.00
		MEAP	0.67	0.67	0.35	0.41	0.68	0.64	0.58	0.59	0.66	0.63	0.58	0.66	0.76	0.65	0.73	1.00	1.00

Chapter 8: Item Analysis

POST-FIELD-TEST ITEM REVIEW

After field-test administration, item analyses were conducted to prepare data for two more rounds of reviews: bias/sensitivity review and content review. For the 2008 MME, the Rasch model was used for item analysis for the social studies portion of the exam. The three parameter logistic item response theory model was used for all other subjects on the exam. This section describes data based on Rasch model analysis for these two post-field-test reviews. A section on item field testing is also in Volume II, and the reader may refer to that section for a presentation that is complementary to this one.

Data

All field-test items were embedded in the live test forms for each test. After the calibration of live test forms, field-test items were calibrated and put onto the same scale as the live operational items. Appendix B lists all the statistics created for the field-tested items. The statistics for each field-test item can be summarized into nine categories.

1. General test information: test name, subject, grade, level;
2. Administration related information: year cycle, administration year, released position;
3. Specific item information: item ID, CID, item type, answer key, maximal score, maturity, item function, character code, number of forms the item appears on, form numbers, test position, n-count (total, male, female, white, and black students), percent for each comment code, percent for each condition code;
4. Content-related information: strand, benchmark, grade level expectation, depth of knowledge, domain, scenario;
5. Option analysis: percent for each option and each score point (total, male, female, white, and black students), p-value or item mean (total, male, female, white, and black students), adjusted p-value, difficulty flag, item standard deviation, item-total correlation, biserial/polyserial correlation, corrected point-serial correlation, item-total correlation flag, option point-biserial correlation, flag for potential miskeying;
6. DIF analysis: Mantel Chi-square, Mantel-Haenszel Delta and its standard error, signed and unsigned SMD, SMD signed effect size, DIF category, and favored group for male vs female comparison and white vs black comparison;
7. IRT parameters: b-parameter and its SE, step parameters and their respective SE, item information at cut points;
8. Fit statistics: mean-square infit, mean-square outfit, mean-square fit flag, misfit level;
9. Data for creating plots: conditional item mean for decile 1 to 10 for each student group (total, male, female, white, and black students) for creating conditional mean plots, 5^{th}, 25^{th}, 50^{th}, 75^{th}, 95^{th} percentile for creating Box & Whisker plot for each student group (total, male, female, white, and black students) for each option and each score point.

These statistics were created by Pearson and sent to Harcourt for creating item labels for bias/sensitivity review and content review.

Statistics and Graphs Prepared for Review Committees

Statistics from item analyses for field-test items were used to create item labels for the post-field-test reviews. Different sets of statistics were prepared for MC and CR items for review committee. Table 8-1 displays all the statistics prepared for MC items for the review committee. These include six categories.

1. General administration information: test name, grade, subject, and administration time;
2. Item general information: CID, maturity, forms and positions;
3. Item specific information: item type, key, p-value, n-count, Rasch difficulty, difficulty flag, point-biserial correlation, point-biserial correlation flag, fit flag, option quality flag;
4. Breakout group descriptives and optional analysis: percent of students selecting each option and omit, option point-biserial correlations, and n-count for all and subgroups: male, female, white, and black students;
5. Differential Item Functioning: flag, and favored group for male vs. female and white vs. black;
6. Review decision;

Table 8-2 displays all the statistics prepared for CR items for the review committee. These include seven categories.

1. General administration information: test name, grade, subject, and administration time;
2. Item general information: CID, maturity, forms and positions;
3. Item specific information: item type, maximal score point, adjusted p-value, item mean, n-count, Rasch difficulty, difficulty flag, item-total correlation, item-total correlation flag, fit flag, score point distribution flag;
4. Breakout group descriptives and score point distribution: percent of students obtaining each score point and omit and n-count for all and subgroups: male, female, white, and black students, omit point-biserial correlation;
5. Invalid code distributions: total invalid scores, frequency of students at each invalid code;
6. Differential Item Functioning: flag, and favored group for male vs. female and white vs. black;
7. Review decision;

All statistics prepared for the review committee for MC and CR items are explained in Appendix C. When the p-value for an MC item, adjusted p-value for a CR item, or Rasch difficulty was out of the desired range, a difficulty flag was shown. When a point-biserial correlation for an MC item or item-total correlation for a CR item was out of range, the appropriate flag was shown. If the mean square infit or outfit was out of desired range, an infit or outfit flag was presented. Similarly, if DIF or improperly functioning options (distracters) were detected, the corresponding flag was activated for the item. The criteria used for flagging an MC or CR item are presented in Table 8-3.

For further psychometric reference, conditional mean plots and Box & Whisker plot for two student group comparison, male vs. female and white vs. black were prepared for the flagged items for the two post-field-test reviews. See Figure 8-1a (for MC items) and 8-1b (for CR items) for conditional mean plots and Figure 8-2a (for MC items) and 8-2b (for CR items) for Box & Whisker plots.

Members of the bias review and content review committees were given specific training in analyzing item quality. Some of the supporting materials for the training sessions are provided in Appendix D (for bias review) and Appendix E (for content review).

Table 8-1. Item Label for a MC Item

MME Grade: 11 Subject: Social Science Admin: Fall 2006

CID: 6688999 **GLCE:** C.2.h.1

Form: 2

Position: 46

Passage:

□ Accept as is
□ Reject
□ Accept with revision

Table 1. Item Information

Type:	MC	**P-value:**	0.37	**Rasch Difficulty:**	0.15	**Difficulty Flag:**	
Key:	B	**N-count:**	860	**PB Correlation:**	0.24	**PB Correlation Flag:**	CL
		Maturity:	FT	**Fit Flag:**		**Option Quality Flag:**	P

Table 2. Breakout Group Descriptives and Option Analysis

		N-count	Percent of Students Selected Option				
			A	B	C	D	Omit
Group	**All**	860	20	37*	21	20	2
	Male	447	21	35	21	20	3
	Female	413	18	40	20	21	1
	White	587	21	35	20	22	2
	Black	207	15	46	20	14	3
Option PB Correlations			-0.13	0.24	-0.14	0.04	

Table 3. Differential Item Functioning

Reference/ Focal Group	Male/ Female	White/ Black
Flag		C
Favored Group		Black

Explanation of DIF Flags
Blank - No or negligible DIF
B - Moderate DIF
C - Large DIF

Table 8-2. Item Label for a CR Item

MME Grade: 11 Subject: Social Science Admin: Fall 2006

ID: 6666666 **Maturity:** FT

Form: 2 5

Position: 27 27

Passage: Government Health Care

□ Accept as is
□ Reject
□ Accept with revision

Table 1. Item Information

Type:	CR	**Adj. P value:**	0.34	**Rasch Difficulty:**	0.22	**Difficulty Flag:**	
Max:	5	**Item Mean:**	1.71	**Item-Total Corr:**	0.55	**Item-Total Corr Flag:**	
		N-count:	1574	**Fit Flag:**		**Score Point Dist. Flag:**	

Table 2. Breakout Group Descriptives and Score Point Distributions

		N-count	Item Mean	Percent of Students at Each Score Point							
				0	1	2	3	4	5	6	Omit
Group	All	1574	1.71	17	34	29	13	7			
	Male	811	1.54	22	36	25	10	7			
	Female	763	1.90	11	32	32	17	8			
	White	1028	1.77	16	33	29	13	9			
	Black	371	1.58	18	34	28	15	5			
Omit PB Correlation											

Table 3. Condition Code Distributions

Frequency of Students at Each Condition Code				
A	B	C	D	E
1		8		

Table 4. Differential Item Functioning

Reference/ Focal Group	Male/ Female	White/ Black
Flag	C	
Favored Group	Female	

Explanation of DIF Flags
Blank - No or negligible DIF
B - Moderate DIF
C - Large DIF

Table 8-3. Flagging Criteria

Statistic	Flag	Flag Definition	Flag Field
PVAL PVAL ADJPVAL	PL PH BL BH	For MC 4 options, if p-value LT .3 (PL) or GT .9 (PH) For CR items, if adj. p-value LT .10 (PL) or GT .9 (PH)	DIFFICFL
BPAR		If b-parameter LT -2.5 (BL) or GT 2.5 (BH)	
ITOT	CL	If item-total correlation LT 0.25 (CL)	ITOTFL
MSQIN MSQOUT	MH MM TP	If msqin or msqout GT 2 (MH) If msqin 1.5 through 2 and msqout LE 2 (MM) If msqout 1.5 through 2 and msqin LE 2 (MM) If msqin LT 0.5 and msqout LT 1.5 (TP) If msqout LT 0.5 and msqin LT 1.5 (TP)	MSQINFL MSQOUTFL
DIF_MF DIF_WB	A B C AA BB CC	For MC items: A: If either |MH Delta| is not significantly GT 0 ($p < 0.05$, using either MH-Chi-Sq or standard error of MH Delta) or if the |MH Delta| is LT 1 B: If |MH Delta| is significantly GT 0 and is either GE 1 and LE 1.5 or is GE 1 but not significantly GT 1 ($p < 0.05$, using standard error of MH Delta) C: If |MH Delta| is both GT 1.5 and significantly GT 1 ($p < 0.05$, using standard error of MH Delta) For CR items: AA: If the Mantel Chi-Sq is not significant ($p > 0.05$) or the |Effect Size| (ES) of SMD LE 0.17 BB: If the Mantel Chi-Sq is significant ($p < 0.05$) and the |ES| is GT 0.17 but LE 0.25 CC: If the Mantel Chi-Sq is significant ($p < 0.05$) and the |ES| is GT 0.25	DIF_MF DIF_WB Categories A and AA are not displayed in flag field
A, B, C, D M, S5, S6, O APB BPB CPB DPB OPB	H L P O N B	For MC items: If the keyed option is not the highest percentage (H) If any option LE 2% (L) If any non-keyed option pb-corr GT 0 (P), or if omit pb-corr GT 0.03 (O) If the keyed option pb-corr LT 0 (N) For CR items: For CR, if omit pb-corr GT 0.03 (O) For CR, if any score point LT 0.5% (L) For CR, if omit GT 20% (B)	MISKFL

Meaning of Flags:

- PL … p-value low
- PH … p-value high
- BL … b-parameter low
- BH … b-parameter high
- CL … correlation low between item and total
- MH … misfit high
- MM … misfit moderate
- TP … too predicTable
- A or AA … no or negligible DIF
- B or BB … moderate DIF
- C or CC … substantial DIF
- H … highest percentage is not a keyed option
- L … low percentage of any option
- P … positive pb-correlation for any non-keyed option
- N … negative pb-correlation for the keyed option
- O … omit has a positive pb-correlation
- B … blanks are over 20%

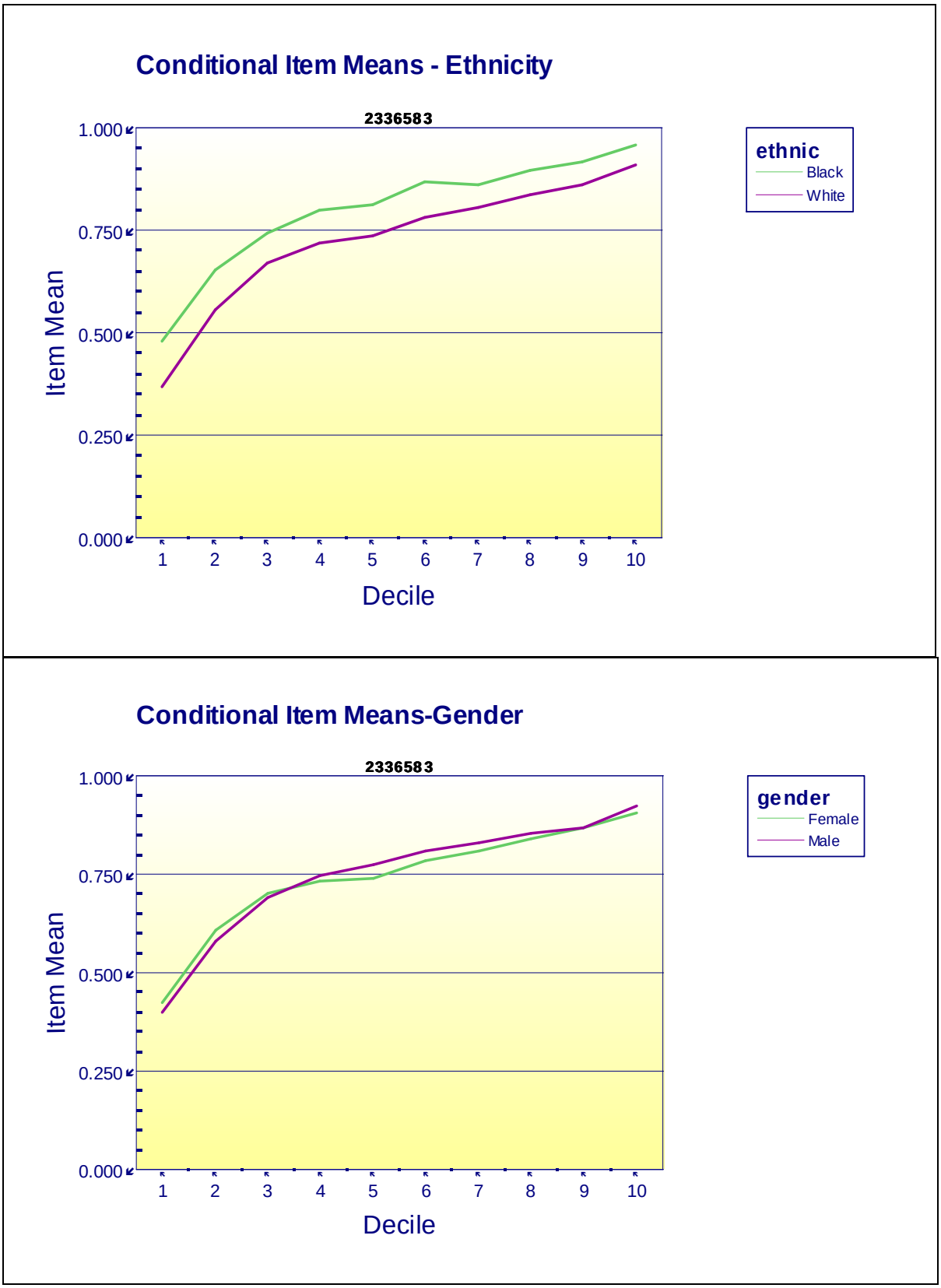


Figure 8-1a. Conditional Item Mean Plots for Ethnicity and Gender for MC Items

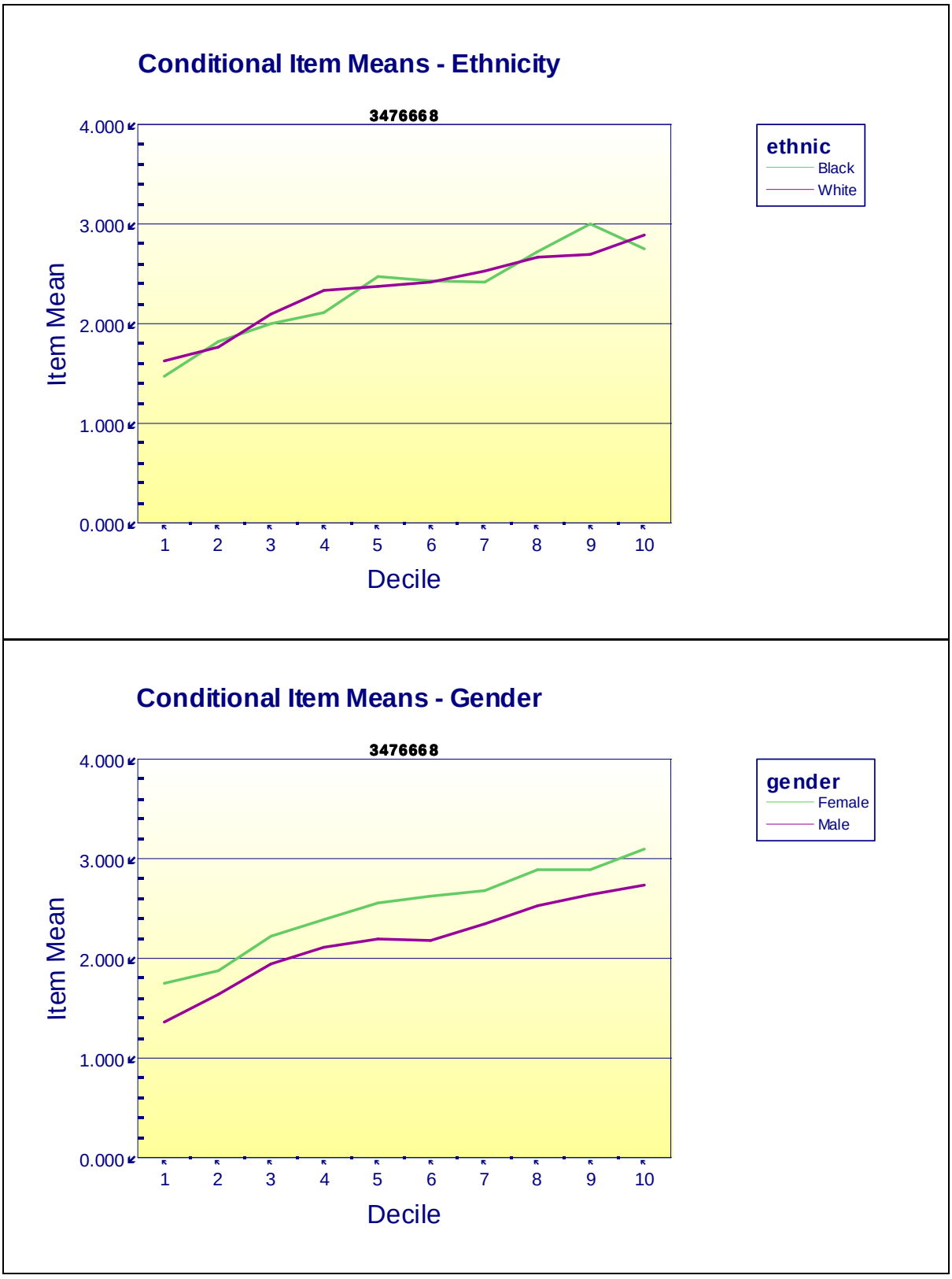


Figure 8-1b. Conditional Item Mean Plots for Ethnicity and Gender for CR Items

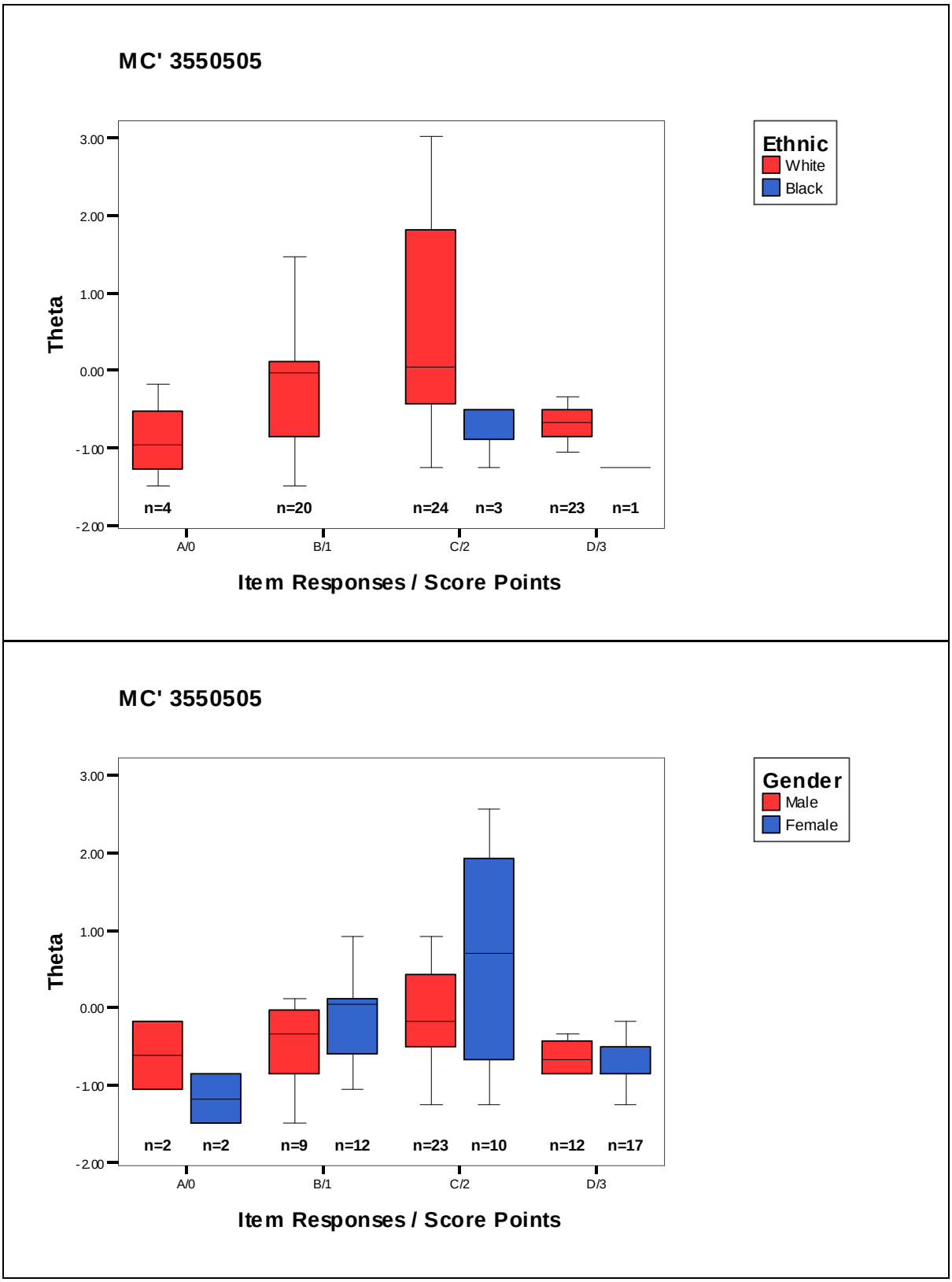


Figure 8-2a. Box & Whisker Plots for Ethnicity and Gender for MC Items

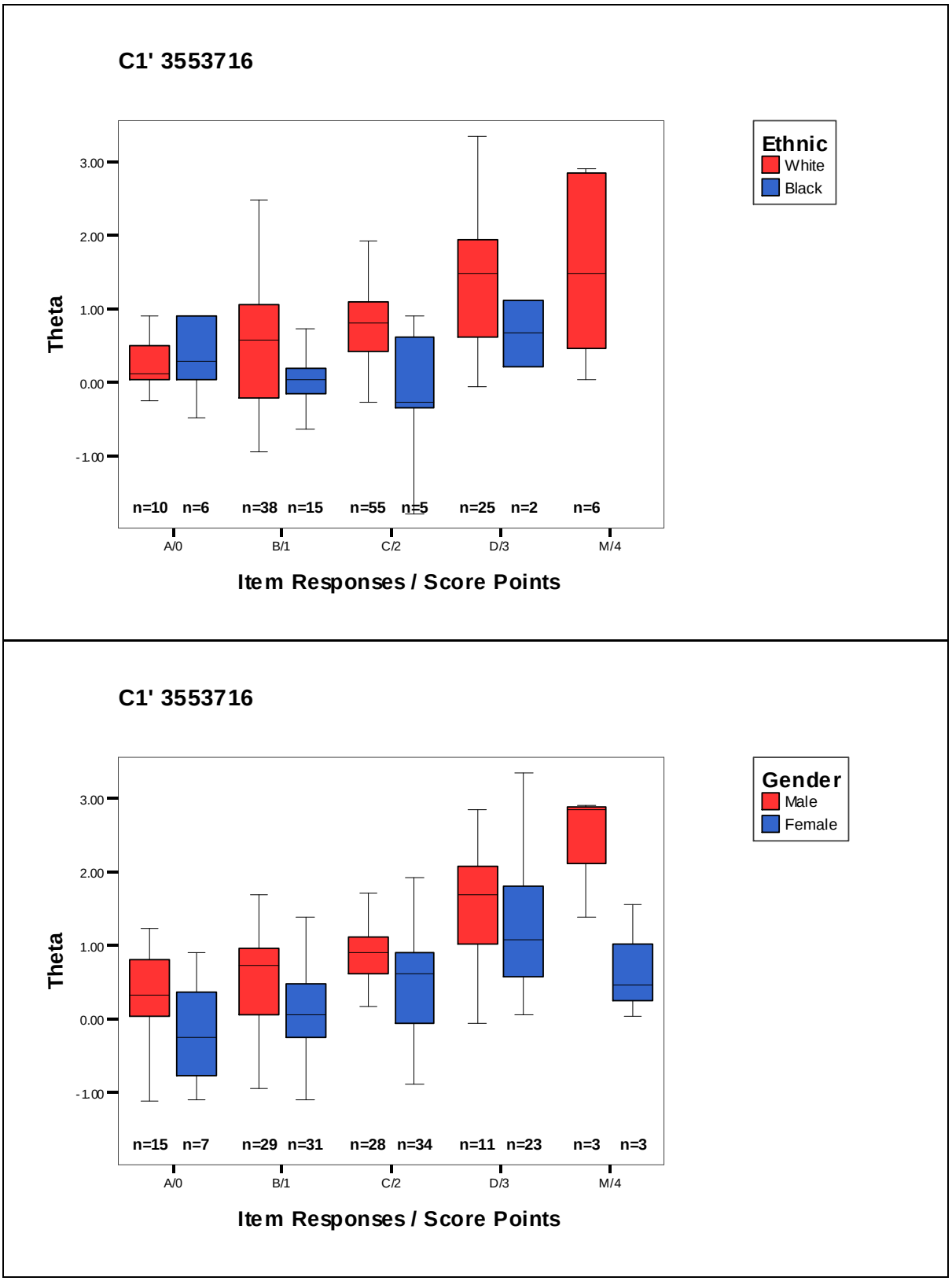


Figure 8-2b. Box & Whisker Plots for Ethnicity and Gender for CR Items

Chapter 9: Standard Setting

Intact documents are used to provide technical understanding of the Standard Setting for the MME rather than being included in the body of this report.

The plan for establishing cut scores for the performance levels is contained in *Standard Setting Plan* (Assessment and Examination Service, 2006). This document described the data collection, methodology (the Bookmark or Item Mapping method) and agenda for conducting the standard setting studies.

The results of a modified item mapping procedure are described in *Standard Setting Report* (Assessment and Examination Service, 2006). The modification to the item mapping method was described as follows. "In the ordered item booklet, three items were flagged as reference items, one for each performance standard(Partially Proficient, Proficient, Advanced). If selected, these items would produce cut-scores such that the percentage of students in each of the four categories would be the same as the results of the Spring 2006 Grade 11 assessments." The data for the standard setting were obtained from panelists who reviewed items ordered with respect to a 2006 field test of the Michigan Merit Examination in Reading, Writing, Mathematics, and Science. The *Standard Setting Report* recommended three cut scores to delineate the four performance levels: Not Proficient, Partially Proficient, Proficient, or Advanced.

A *Michigan Department of Education Memorandum* in October 2006 described four possiblt sets of cut scores for the performance levels, and recommended one. A second *Michigan Department of Education Memorandum* (November 2006) revised the recommendation to a different set of cut scores, and provided a justification based on a change in content specifiactions. The revised recommendation was to adopt MME cut scores based on a linkage to the MEAP.

The formal adoption of MME cut scores is detailed on page 5 of the minutes of the November 2006 State Board of Education meeting (*Minutes of the State Board of Education* November 14, 2006).

Chapter 10: Adequate Yearly Progress and EducationYES

The major policy-based uses of assessment data from the MME, MEAP and MI-Access are for public reporting and school accountability decisions.

Legislative Grounding

- The federal No Child Left Behind Act (NCLB) requires that Adequate Yearly Progress (AYP) be calculated for all public schools, for each school district, and for the state.
- Michigan statute (section 1280 of the Revised School Code) requires the State Board of Education to accredit public elementary and secondary schools. The State Board approved *Education YES – A Yardstick for Excellent Schools!* in 2002, and accepted the report of the Accreditation Advisory Committee in 2003.

NCLB requires that AYP be determined for all public schools, for each school district, and for the state. The school or district must attain the target achievement goal in reading and mathematics or reduce the percentage of students in the non-proficient category (Partially Proficient and Not Proficient) of achievement by 10% ("safe harbor"). A school or district must also test at least 95% of its students enrolled in the grade level tested for the school as a whole and for each required subgroup. In addition, the school and district must meet or exceed the other academic indicators set by the state: graduation rate for high schools and attendance rate for elementary and middle schools. These achievement goals must be reached for each subgroup that has a measurable group of students.

Education YES! uses several components that are interlinked to present a complete picture of performance at the school level. *Education YES!* is a broad set of measures that looks at school performance and student achievement in multiple ways. Measures of student achievement in Michigan's school accreditation system include:

- Achievement status to measure how well a school is doing in educating its students.
- Achievement change to measure whether student achievement is improving or declining.
- Achievement growth (delayed until 2007-2008) to measure whether students are demonstrating at least one year of academic growth for each year of instruction.

In addition, the Indicators of School Performance measure investments that schools are making in improved student achievement, based on indicators that come from research and best practice.

Procedures for Using Assessment Data for Accountability

The school or district must attain the target achievement goal in English language arts (reading and writing) and mathematics or reduce the percentage of students in the non-proficient category (Partially Proficient and Not Proficient) of achievement by 10% ("safe harbor"). A school or district must also assess at least 95% of its students enrolled in the grade level tested for the school as a whole and for each required subgroup. In addition, the school must meet or exceed the other academic indicators set by the state: graduation rate for high schools of 80%, and

attendance rate for elementary and middle schools of 85%. These achievement goals must be reached for each subgroup that has at least the minimum number of students in the group. The group size is the same for the school, school district, and the state as a whole. The subgroups are:

- Major Racial/Ethnic Groups
  - Black or African American
  - American Indian or Alaska Native
  - Asian American, Native Hawaiian or other Pacific Islander
  - Hispanic or Latino
  - White
  - Multiracial
- Students with Disabilities
- Limited English Proficient
- Economically Disadvantaged

Michigan's minimum subgroup size is 30 students. For a district or school that enrolls more than 3,000 students, the minimum subgroup size will be 1% of enrollment, up to a maximum subgroup size of 200 students. An AYP determination will be made for all subgroups of 200 or more students.

It is the policy of the Michigan State Board of Education that all students participate in the state assessment program. The student's status, in terms of enrollment for a full academic year, is not relevant to whether the student should be assessed. The federal No Child Left Behind Act requires that at least 95% of enrolled students be assessed. The number of students to be assessed is determined from the Single Record Student Database (SRSD), collected by the Center for Educational Performance and Information (CEPI). This is taken from the Fall (September) collection for grades 3-8 and from the Spring (February) collection for high schools.

The State Board of Education in Michigan has determined the AYP state targets (Annual Measurable Objectives) for the determination of AYP. These targets are based on assessment data from the 2001-02 administration of the MEAP tests and represent the percentage of proficient students in a public school at the 20th percentile of the State's total enrollment among all schools ranked by the percentage of students at the proficient level.

56% - Elementary Mathematics
48% - Elementary English Language Arts
43% - Middle School Mathematics
43% - Middle School English Language Arts
44% - High School Mathematics
52% - High School English Language Arts

Because valid scores in English language arts and mathematics cannot be ignored, the scores of all tested students must be used in the AYP determination. Michigan has extended the grade range targets with separate targets for each grade, and by basing a school's target on a weighted average of the statewide targets for the grades tested at the school. This procedure accounts for differences in performance standards across grade levels. The method also permits a single AYP

determination for the school, through a comparison between student achievement and the school's target.

Proficiency for AYP is based on the weighted sum of a proficiency index that is computed at each grade (3-11) counted for AYP at the school. Michigan did not change the approved AYP targets that were set previously. A set of grade level targets applicable to the 2005-06 school year has been developed and incorporated into the calculation of a Proficiency Index. The Proficiency Index is used to determine if a school, district, or student group meets the state AYP target.

A school, school district, or subgroup meets the state objective if the proficiency index is equal to or greater than zero (0). MDE will not determine or report AYP by grade. The grade level targets will be used to compute the proficiency index, which is aggregated across grades based on the school's configuration.

It is generally accepted that the SEM varies across the range of student proficiencies and that individual score levels on any particular test could potentially have different degrees of measurement error associated with them. For this reason, it is generally useful to report not only a test level SEM estimate, but individual score level estimate as well. Individual score level estimates of error are commonly referred to as conditional standard errors of measurement (CSEM). The CSEM provides an estimate of error variability, conditional on the proficiency estimate (theta). In other words, it provides an error estimate, at each score point. According to the IRT model, there is typically more information in the middle of the theta score distribution, so the CSEM is usually smallest in this range. Michigan began use of the conditional standard errors of measurement in 2005-06 for its state assessments. Conditional standard errors of measurement are used to improve the accuracy of AYP determinations.

In addition the Indicators of School Performance measure investments that schools are making in improved student achievement, based on indicators that come from research and best practice. Scores on all three components of *Education YES!* have been converted to a common 100 point scale where: 90-100 A; 80-89 B; 70-79 C; 60-69 D; and 50-59 F. Grades of D and F are not used for the school's composite grade, where the labels D/Alert and Unaccredited are used.

Achievement Status

Achievement status is measured in English Language Arts and Mathematics at the elementary level. It includes Science and Social Studies at the middle school and high school levels. Achievement Status uses up to three years of comparable data from the Michigan Educational Assessment Program, the Michigan Merit Examination, or the MI-Access Assessments.

The method of computing achievement status uses students' scale scores on the Michigan assessments, as weighted by the performance level or category (1,2,3, or 4) assigned to each student's score. Scale score values at the chance level are substituted for values below the chance level because values below that point do not have valid information about the student's performance. A template is provided so that a school can paste in their assessment data to see how the values are derived. The weighted index is computed by following these steps:

1. Multiply each student's scale score by the performance level (i.e., 1100*2);

2. Sum of the resulting values resulting in the sum of the index values;
3. Sum of the performance levels or weights;
4. Divide the sum of the index values by the sum of the weights.

The intent of the weighted index is to encourage schools to place priority on improving the achievement of students that attain the lowest scores on the Michigan assessments.

Cut scores for the score ranges in achievement status were set by representative panels that assigned grades to selected schools. The cut scores were reviewed by the Accreditation Advisory Committee and approved by the State Board of Education. The Accreditation Advisory Committee, a group of five national experts, was appointed by the State Board of Education to advise the Board on the implementation of the *Education YES!* school accreditation.

Achievement Change

Achievement change uses up to five years of comparable assessment data to determine if student achievement in a school is improving at a rate fast enough to attain the goal of 100% proficiency in school year 2013-14, as required by the No Child Left Behind Act (NCLB). The change grade is derived from the average of up to three calculations of improvement rates (slopes) using the school's assessment data. Scores from assessments that are not comparable will not be placed on the same trend line. Achievement Change is based on the goal of 100% percent proficient in 2013-14, as set in NCLB. Achievement Change is computed by dividing the computed slope by the target slope, determining the percent of the target that the school has attained.
The linear regression methodology previously used to calculate Achievement Change was not used in 2006-07 for the elementary and middle school levels because scores from assessments that are not comparable cannot be placed on the same slope line. Multiple linear regression was used to predict each school's 2006-07 score based on the school's scores from 2003-04, 2004-05, and 2005-06. A prediction was made for each content area and grade level that was tested in previous years. The prediction was compared to the school's actual 2006-07 percent proficient. The Difference is computed as the (Actual – Predicted). The school's status score for each content area and grade range is adjusted as follows:

- Schools where the actual score exceeds the prediction plus 1.5 times the standard error of the estimate had a 15 point adjustment added to the achievement score for that content area;
- Schools where the actual score exceeds the prediction plus the standard error of the estimate had a 10 point adjustment added to the achievement score for that content area;
- Schools where the actual score is less than the prediction minus 1.5 times the standard error of the estimate had a 15 point deduction applied to the achievement score for that content area; and
- Schools where the actual score is less than the prediction minus the standard error of the estimate had a 10 point deduction applied to the achievement score for that content area.

The Achievement Change adjustment is calculated only if there are at least 10 students tested each year (2002-03, 2003-04, 2005-05 and 2005-06) in the content area and grade level.

A school district has the opportunity to appeal any data that affect its grade or AYP status if it has evidence that the data may be inaccurate. For example, the school district might identify corrected data regarding the number of students that were enrolled and should have been assessed. The Department of Education will do all that it can to correct errors that are brought to its attention. The purpose of the appeal window is to address substantive issues regarding the *Education YES!* grade or AYP status. The school district must cite specific data that are challenged in the appeal. Appeals that have no effect on the *Education YES!* grade or AYP status will not be considered.

The scoring and grading for the Indicators of School Performance are based on the school's self-rating of each component for each indicator. Each school team assigned the school a rating for each component, using the following scale:

- Systematically and Consistently Meeting Criteria;
- Progressing Toward Criteria;
- Starting to Meet Criteria; or
- Not Yet Meeting Criteria.

The ratings were scored on a scale where the number of possible points for each indicator is 36. The number of points possible for each component varies based on the number of components in the indicator. This method equally weights each indicator. For example, an indicator with 3 components receives 12 points per component whereas an indicator with 4 components receives 9 points per component. The possible score for all schools is 396 (11 indicators times 36 points). A single grade is assigned to the group of 11 indicators. The school's grade is based on the percentage of the possible points that the school could score for the total of all 11 indicators.

A "window" to update the School Self Assessments, including updating the self-rating and evidence for the Indicators of School Performance, ends on March 31, 2007. Beginning in 2004-05, the Department published both the school's self-rating and the evidence reported for each component. The school's self-rating for each component, and the evidence provided, is available in the online Report Card at https://oeaa.state.mi.us/ayp/.

The State Board of Education has approved a new School Improvement Framework that is intended to form the basis of revisions to the Indicators of School Performance for 2007-08. Draft rubrics have been developed and a pilot study was done in the spring of 2006.

Scores and grades are calculated for each content area for each school. The content areas remain the same, using only English Language Arts and Mathematics at the elementary level, and adding Science and Social Studies at the middle school and high school levels. The score and grade for each content area is based on the score for achievement status, as adjusted by averaging it with the score for achievement change.

The composite school grade is derived from the school scores and letter grades and the school's status in terms of Adequate Yearly Progress (AYP) under the federal No Child Left Behind Act. The weighting of the components of *Education YES!* in the composite grade has been as follows:

Table 10-1. Education YES! Composite Score Weighting

Component	Point Value	
	Until 2006-07	2007-08 and After
School Performance Indicators	33	33
Achievement Status	34	23
Achievement Change	33	22
Achievement Growth		22
Total	100	100

The scores for each content area are averaged to calculate an achievement score and grade for each school. An achievement score for each content area has been computed by averaging the Status and Change (or adjusted Change) scores for a content area. A preliminary aggregate achievement score is derived by averaging the scores from each content area. The preliminary aggregate achievement score is weighted 67% and the School Self-Assessment (Indicator score) is weighted 33% in calculating the preliminary score and grade for a school.

In 2004-05, the State Board of Education approved a change to the *Education YES!* policy so that the school's indicator score cannot improve the school's composite score and grade by more than one letter grade more than the school's achievement grade. This means that a school that receives an "F" for achievement can receive a composite grade no higher than "D/Alert."

After the computation of a school's composite grade for achievement described above, a final "filter" will be applied, consisting of the question of whether or not a school or district met or did not meet AYP. The answer to this question is an additional determining factor for a school's final composite grade on the report card. A school that does not make AYP shall not be given a grade of "A." A school that makes AYP shall not be listed as unaccredited. A school's composite school grade will be used to prioritize assistance to underperforming schools and to prioritize interventions to improve student achievement.

Table 10-2. Unified Accountability for Michigan Schools

Education YES! Composite Score			
	90-100	B (iv)	A
	80-89	B (iv)	B (iv)
	70-79	C (iii)	C (iii)
	60-69	D/Alert (ii)	C (iii)
	50-59	Unaccredited (i)	D/Alert (ii)
		Did Not Make AYP	***Makes AYP***

(i) – (iv) Priorities for Assistance and Intervention

Schools that are labeled "A", "B", "C", or "D / Alert" will be accredited. Schools that receive an "A" will be summary accredited. Schools that receive a "B", "C", or "D/Alert" will be in interim status. Unaccredited schools will also be labeled as such. Summary accreditation, interim status, and unaccredited are labels from Section 1280 of the Revised School Code.

Results of accountability analyses for 2006-7 are reported in next section. Results of accountability analyses for 2007-08 will be available in August, 2008, and will be included in the 2009 version of this document.

Table 10-3. Results of Accountability Analyses

Report on Michigan School AYP 2007				
	Total Number of Schools	Elementary	Middle School	High Schools
Final Results for 2007				
Total Number of Schools	3,716	1,738	829	1,149
Made AYP	3,011 100.0%	1,637 100.0%	714 100.0%	660 100.0%
Did Not Make AYP	705 23.4%	101 6.2%	115 16.1%	489 74.1%
Final Results for 2006				
Total Number of Schools	3,750	1,729	816	1,205
Made AYP	3,206 100.0%	1,660 100.0%	740 100.0%	806 100.0%
Did Not Make AYP	544 17.0%	69 4.2%	76 10.3%	399 49.5%

Table 10-4. Report on School AYP 2005-2006

			2004-05	2005-06
Total Number of Schools Assigned AYP status			3,748	3,796
Total Number of Schools Not Making AYP			544	666
Percent of Schools Not Making AYP			14.5%	17.5%
Schools that make AYP using Interim Flexibility Option 1 - Students with Disabilities group			360	103
Schools Identified for Improvement			343	380
Schools Identified for Improvement by Phase		1	105	117
		2	102	66
		3	79	87
		4	16	74
		5	15	8
		6	22	15
		7	4	12
		8		1
Schools with Graduation Rates under 80%			157	133
Schools not meeting **Participation** target by group	All Students		145	182
	Black		82	97
	American Indian		2	15
	Asian American		5	3
	Hispanic		15	10
	White		61	79
	Multiracial			1
	Limited English Proficient		3	10
	Students with Disabilities		106	134
	Economically Disadvantaged		142	162
Schools not meeting **Proficiency** target by group	All Students		179	277
	Black		78	108
	Asian American		0	
	Hispanic		9	11
	White		10	14
	Limited English Proficient		11	14
	Students with Disabilities		169	214
	Economically Disadvantaged		73	102

Table 10-5. Report on Michigan District AYP 2007

	Total Number of Districts	Number Met AYP	Percent Met AYP	Number Not Met AYP	Percent Not Met AYP
Final Results for 2007					
All School Districts	551	532	96.6%	19	3.4%
K-12 Districts	493	484	98.2%	9	1.8%
Charters	30	26	86.7%	6	20.0%
ISDs	28	4	14.3%	4	14.3%
Final Results for 2006					
All School Districts	547	539	98.5%	4	0.7%
K-12 Districts	493	490	99.4%	3	0.6%
Charters	26	26	100.0%	0	0.0%
ISDs	28	27	96.4%	1	3.6%

Table 10-6. State Accreditation Letter Grades 2006 and 2007

Grade	2006		2007	
	Number of Schools	Percent of Schools	Number of Schools	Percent of Schools
A	1,186	46.3%	914	31.7%
B	1,319	51.5%	1,333	46.3%
C	672	26.2%	895	31.1%
D-Alert	109	4.3%	234	8.1%
Unaccredited	5	0.2%	5	0.2%
No Grade	457		415	
Total	3,748		3,796	

Chapter 11: State Summary Data

For the spring 2008 administration, the summary data are presented in Table 11-1. For each content area, Table 11-1 presents the average score and the percentages of students falling into each of the four performance levels. Frequency distributions for the MME scale scores are presented in Figures 11-1 through 11-6, and in Tables 11-2 through 11-7. Tables 11-8 through 11-12 present the summary statistics for the item parameter estimates.

Table 11-1. Spring 2008 Michigan State Average Scores and Percentages in each Performance Level

			Percentages within Performance Levels			
Content Area	**N**	**Average**	**Not Proficient**	**Partially Proficient**	**Proficient Standards**	**Advanced Standards**
Reading	130,226	1106	17%	24%	58%	2%
Writing	129,400	1090	10%	50%	38%	2%
ELA	128,818	1099	12%	37%	49%	2%
Mathematics	129,803	1093	38%	16%	37%	10%
Science	129,691	1099	28%	16%	50%	6%
Social Studies	130,957	1123	7%	9%	42%	41%

Figure 11-1. Frequency Plot for MME Spring 2008 English Language Arts Scale Score Total Group -- All Forms Included

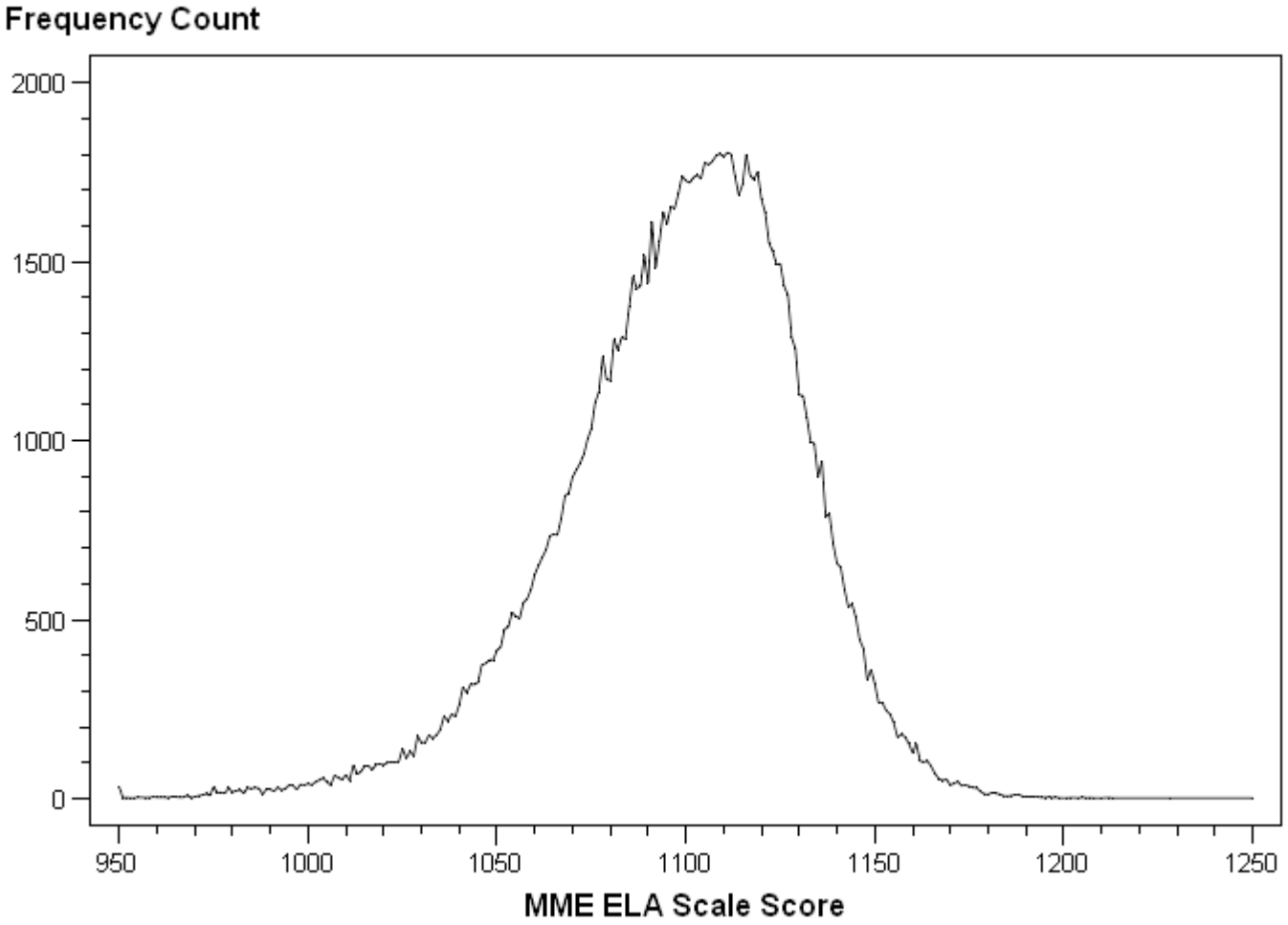


Figure 11-2. Frequency Plot for MME Spring 2008 Mathematics Scale Score Total Group -- All Forms Included

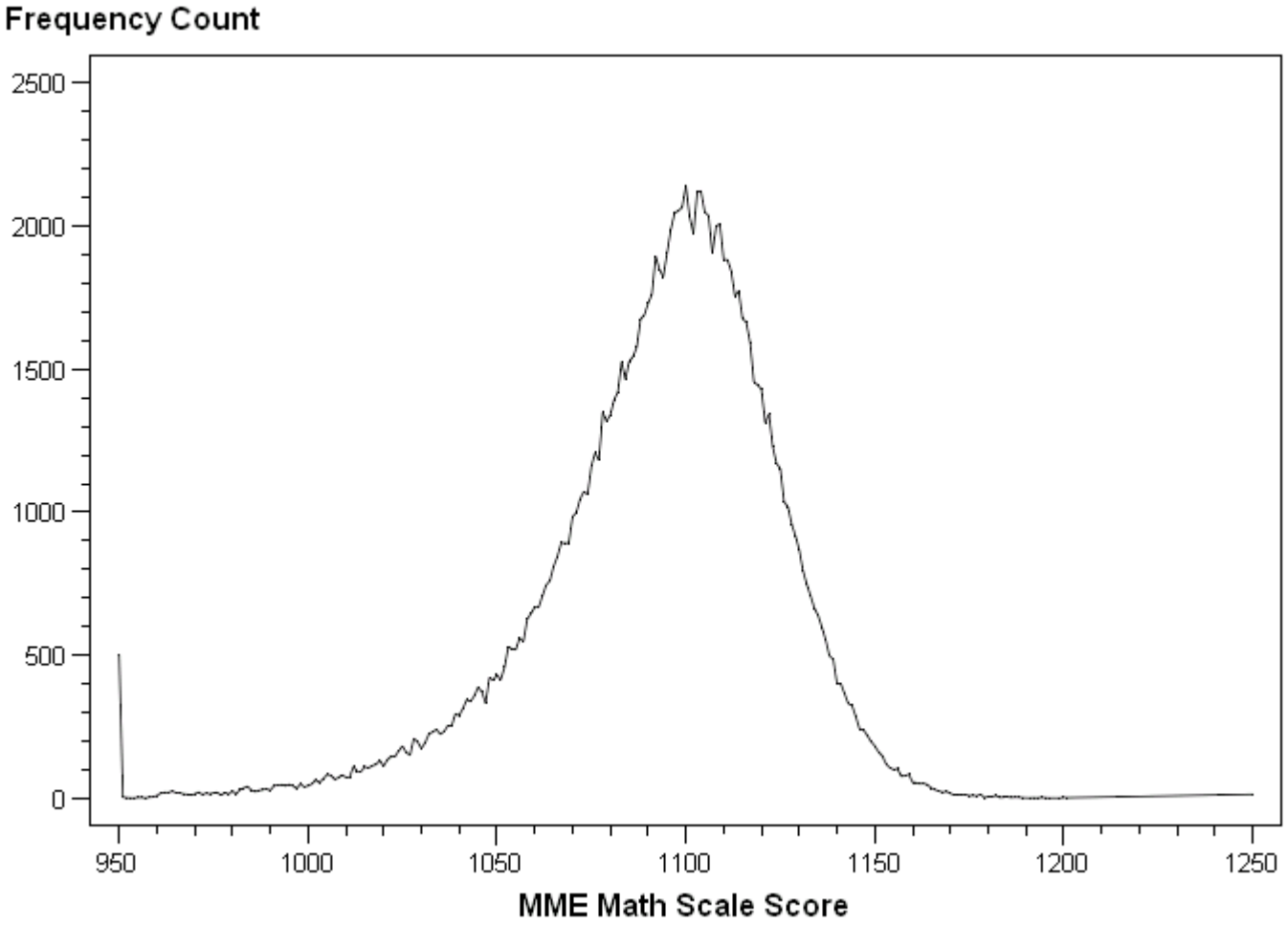


Figure 11-3. Frequency Plot for MME Spring 2008 Reading Scale Score Total Group -- All Forms Included

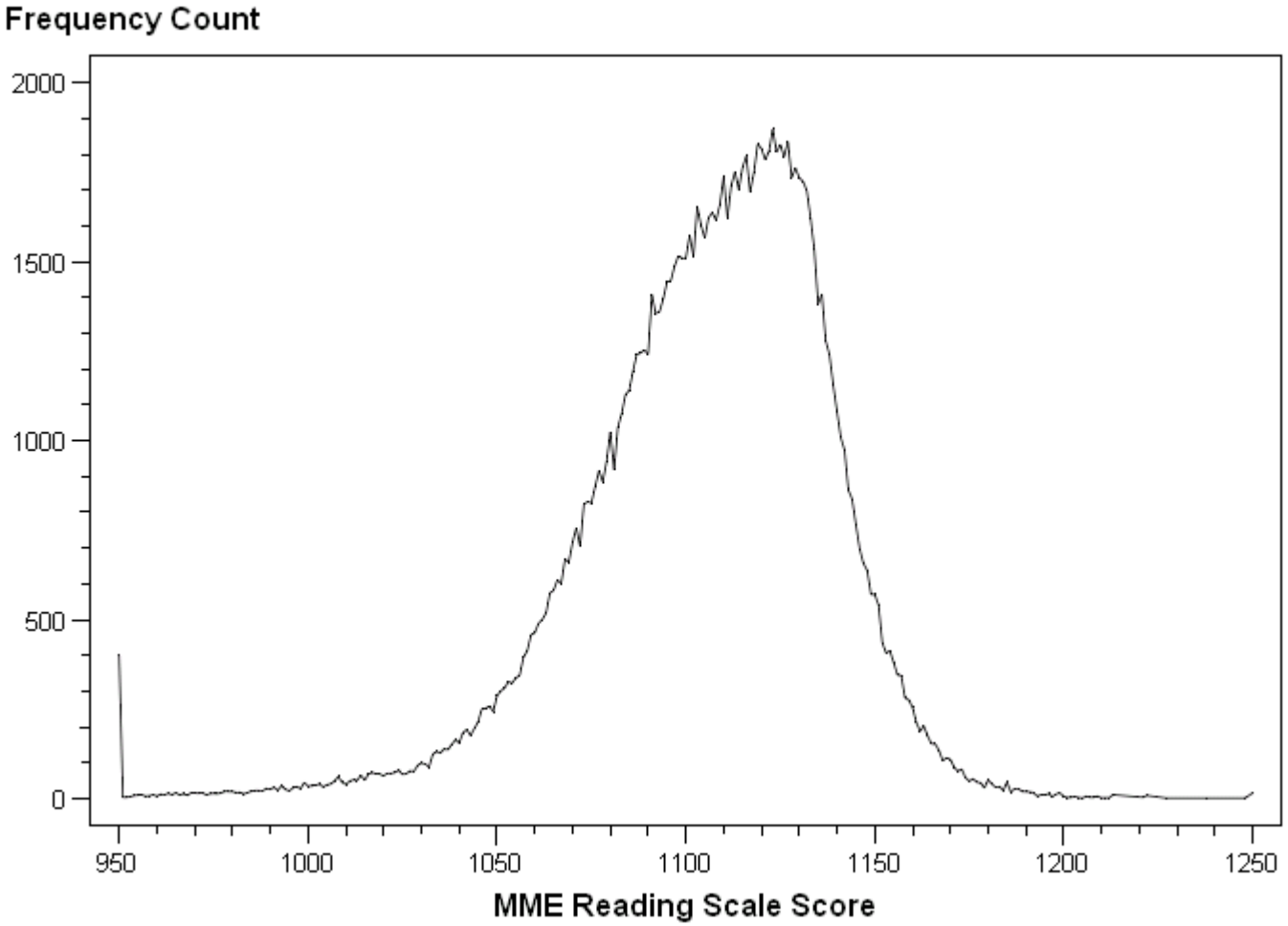


Figure 11-4. Frequency Plot for MME Spring 2008 Science Scale Score Total Group -- All Forms Included

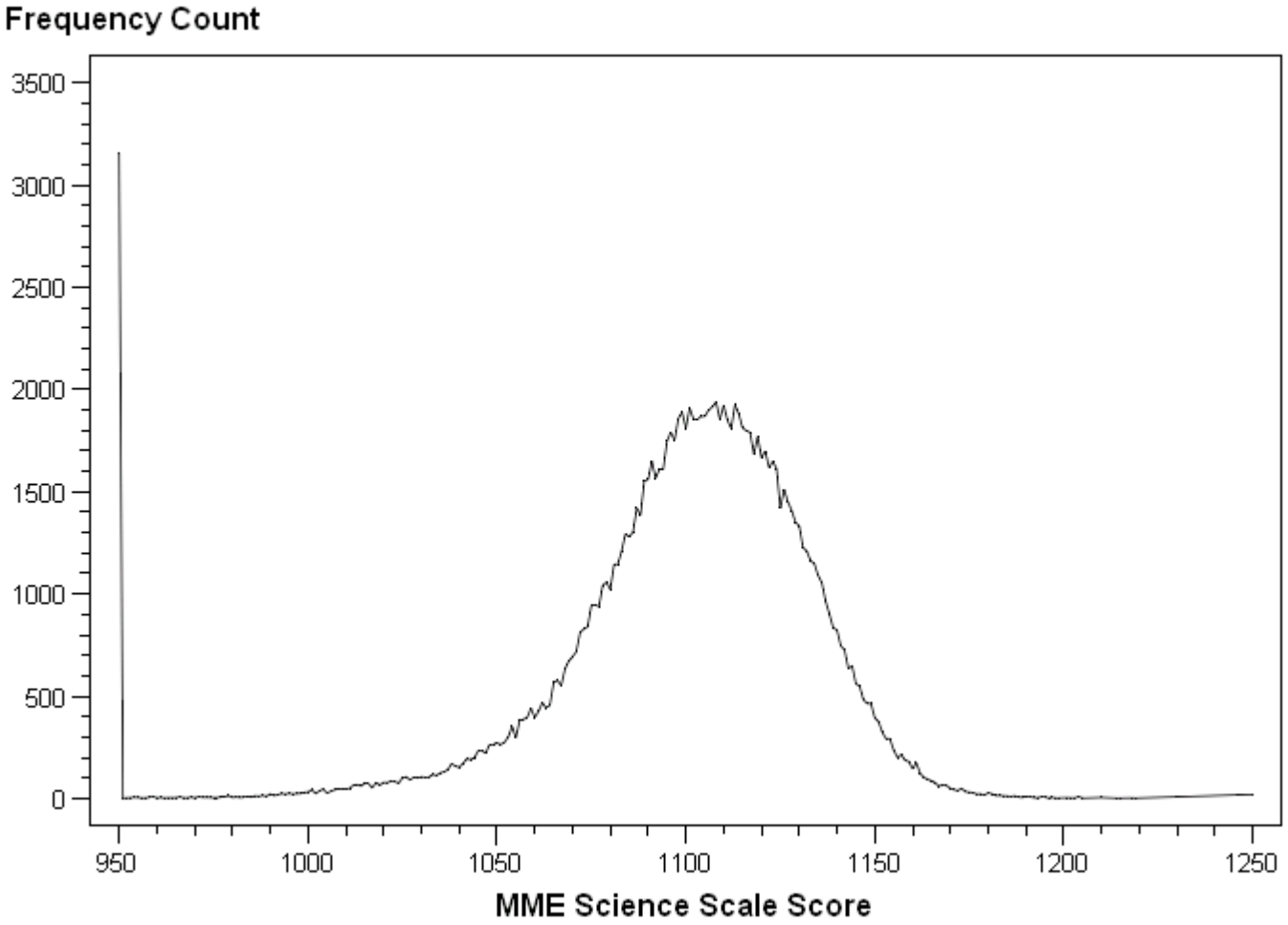


Figure 11-5. Frequency Plot for MME Spring 2008 Social Studies Scale Score Total Group -- All Forms Included

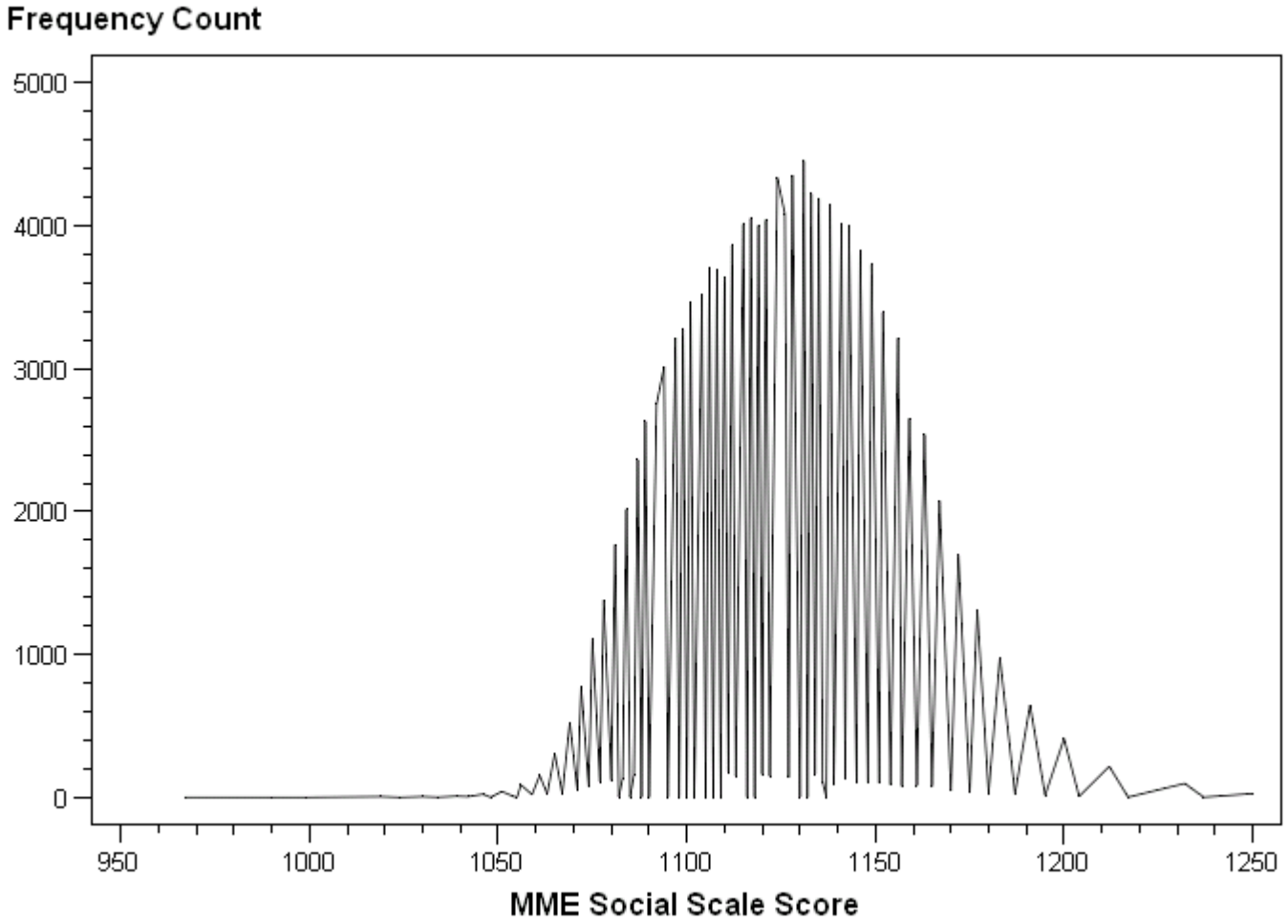


Figure 11-6. Frequency Plot for MME Spring 2008 Writing Scale Score Total Group -- All Forms Included

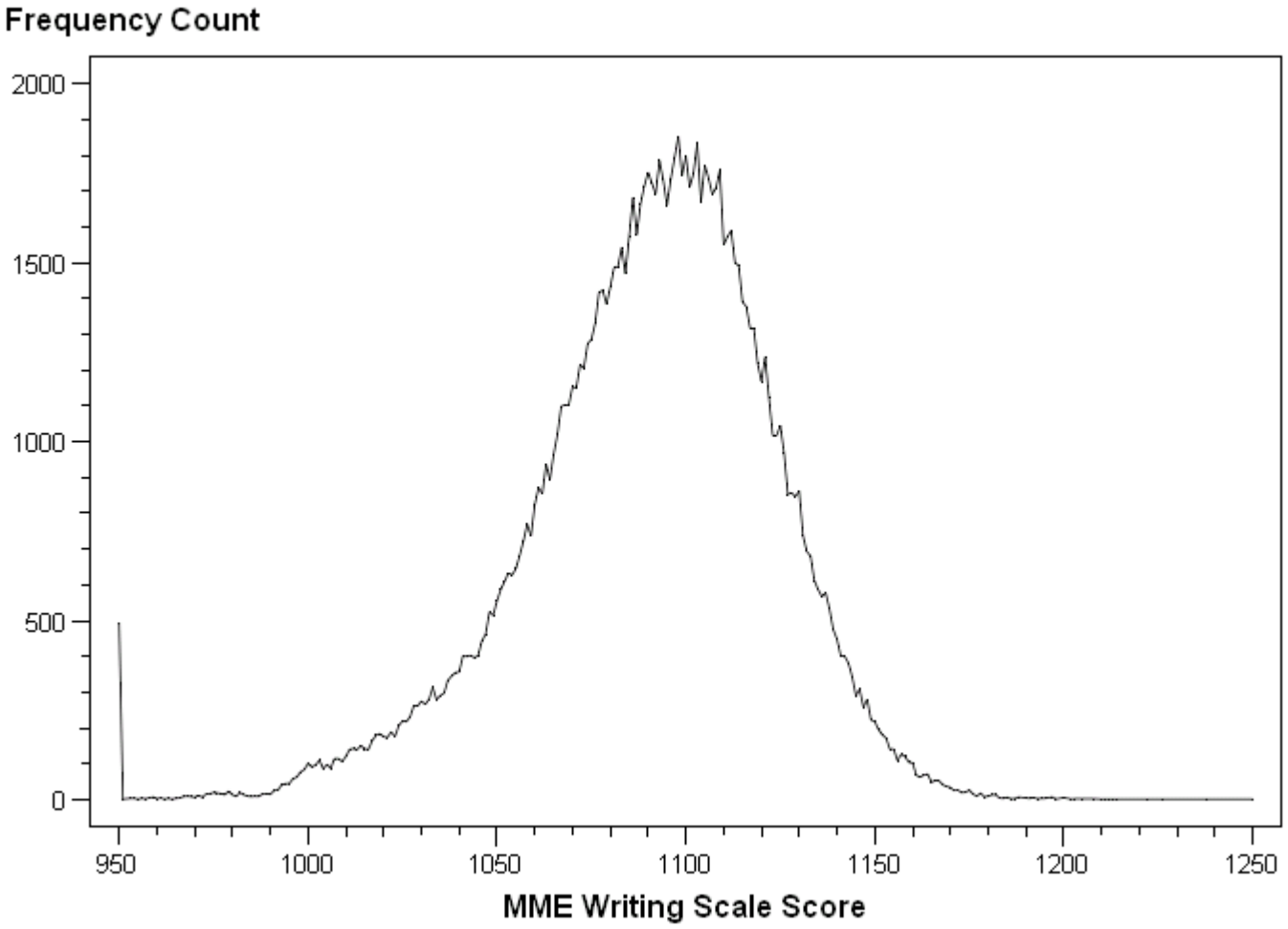


Table 11-2. MME Spring 2008 English Language Arts Frequencies for Total Group -- All Forms Included

ELA Scale Score	Frequency	Percent
950	32	0.02
951	1	0.00
952	2	0.00
953	1	0.00
954	1	0.00
955	3	0.00
957	2	0.00
958	2	0.00
959	4	0.00
960	3	0.00
961	3	0.00
962	3	0.00
963	2	0.00
964	5	0.00
965	4	0.00
966	3	0.00
967	3	0.00
968	8	0.01
969	2	0.00
970	4	0.00
971	7	0.01
972	9	0.01
973	13	0.01
974	8	0.01
975	31	0.02
976	16	0.01
977	17	0.01
978	14	0.01
979	30	0.02
980	17	0.01
981	21	0.02
982	24	0.02
983	14	0.01
984	30	0.02
985	27	0.02
986	30	0.02
987	28	0.02
988	11	0.01
989	26	0.02
990	25	0.02
991	20	0.02

ELA Scale Score	Frequency	Percent
992	31	0.02
993	23	0.02
994	27	0.02
995	37	0.03
996	37	0.03
997	24	0.02
998	39	0.03
999	37	0.03
1000	41	0.03
1001	37	0.03
1002	46	0.04
1003	51	0.04
1004	57	0.04
1005	45	0.03
1006	37	0.03
1007	64	0.05
1008	58	0.05
1009	52	0.04
1010	66	0.05
1011	50	0.04
1012	93	0.07
1013	67	0.05
1014	75	0.06
1015	90	0.07
1016	91	0.07
1017	80	0.06
1018	94	0.07
1019	96	0.07
1020	93	0.07
1021	100	0.08
1022	101	0.08
1023	101	0.08
1024	102	0.08
1025	139	0.11
1026	112	0.09
1027	131	0.10
1028	119	0.09
1029	174	0.14
1030	156	0.12
1031	156	0.12
1032	177	0.14

ELA Scale Score	Frequency	Percent
1033	168	0.13
1034	177	0.14
1035	192	0.15
1036	230	0.18
1037	216	0.17
1038	235	0.18
1039	229	0.18
1040	261	0.20
1041	311	0.24
1042	295	0.23
1043	319	0.25
1044	318	0.25
1045	324	0.25
1046	374	0.29
1047	377	0.29
1048	387	0.30
1049	387	0.30
1050	414	0.32
1051	423	0.33
1052	473	0.37
1053	479	0.37
1054	520	0.40
1055	509	0.40
1056	503	0.39
1057	547	0.42
1058	558	0.43
1059	584	0.45
1060	626	0.49
1061	651	0.51
1062	674	0.52
1063	695	0.54
1064	734	0.57
1065	738	0.57
1066	736	0.57
1067	782	0.61
1068	846	0.66
1069	853	0.66
1070	898	0.70
1071	918	0.71
1072	937	0.73
1073	962	0.75
1074	1005	0.78
1075	1034	0.80

ELA Scale Score	Frequency	Percent
1076	1107	0.86
1077	1137	0.88
1078	1234	0.96
1079	1171	0.91
1080	1169	0.91
1081	1283	1.00
1082	1253	0.97
1083	1289	1.00
1084	1285	1.00
1085	1374	1.07
1086	1459	1.13
1087	1425	1.11
1088	1432	1.11
1089	1518	1.18
1090	1442	1.12
1091	1612	1.25
1092	1484	1.15
1093	1557	1.21
1094	1640	1.27
1095	1606	1.25
1096	1655	1.28
1097	1649	1.28
1098	1690	1.31
1099	1740	1.35
1100	1727	1.34
1101	1721	1.34
1102	1736	1.35
1103	1743	1.35
1104	1732	1.34
1105	1778	1.38
1106	1772	1.38
1107	1781	1.38
1108	1797	1.39
1109	1803	1.40
1110	1794	1.39
1111	1806	1.40
1112	1801	1.40
1113	1740	1.35
1114	1687	1.31
1115	1716	1.33
1116	1801	1.40
1117	1742	1.35
1118	1730	1.34

ELA Scale Score	Frequency	Percent
1119	1752	1.36
1120	1675	1.30
1121	1638	1.27
1122	1553	1.21
1123	1529	1.19
1124	1493	1.16
1125	1495	1.16
1126	1432	1.11
1127	1408	1.09
1128	1288	1.00
1129	1257	0.98
1130	1127	0.87
1131	1126	0.87
1132	1067	0.83
1133	998	0.77
1134	992	0.77
1135	898	0.70
1136	944	0.73
1137	788	0.61
1138	795	0.62
1139	711	0.55
1140	657	0.51
1141	646	0.50
1142	583	0.45
1143	535	0.42
1144	543	0.42
1145	506	0.39
1146	443	0.34
1147	419	0.33
1148	331	0.26
1149	360	0.28
1150	321	0.25
1151	267	0.21
1152	268	0.21
1153	246	0.19
1154	236	0.18
1155	212	0.16
1156	171	0.13
1157	179	0.14
1158	172	0.13
1159	153	0.12
1160	128	0.10
1161	152	0.12

ELA Scale Score	Frequency	Percent
1162	104	0.08
1163	102	0.08
1164	104	0.08
1165	88	0.07
1166	68	0.05
1167	55	0.04
1168	50	0.04
1169	53	0.04
1170	37	0.03
1171	42	0.03
1172	47	0.04
1173	35	0.03
1174	37	0.03
1175	34	0.03
1176	30	0.02
1177	29	0.02
1178	19	0.01
1179	12	0.01
1180	11	0.01
1181	16	0.01
1182	14	0.01
1183	12	0.01
1184	7	0.01
1185	6	0.00
1186	7	0.01
1187	10	0.01
1188	11	0.01
1189	6	0.00
1190	4	0.00
1191	5	0.00
1192	4	0.00
1193	4	0.00
1194	3	0.00
1195	2	0.00
1196	4	0.00
1197	2	0.00
1198	3	0.00
1199	1	0.00
1201	1	0.00
1202	2	0.00
1204	1	0.00
1205	3	0.00
1206	1	0.00

ELA Scale Score	Frequency	Percent
1207	2	0.00
1208	1	0.00
1210	1	0.00
1212	2	0.00

ELA Scale Score	Frequency	Percent
1213	1	0.00
1228	1	0.00
1250	1	0.00

Table 11-3. MME Spring 2008 Mathematics Frequencies for Total Group -- All Forms Included

Mathematics Scale Score	Frequency	Percent
950	503	0.39
951	4	0.00
952	2	0.00
953	1	0.00
954	1	0.00
955	3	0.00
956	3	0.00
957	1	0.00
958	4	0.00
959	5	0.00
960	9	0.01
961	18	0.01
962	21	0.02
963	20	0.02
964	25	0.02
965	20	0.02
966	19	0.01
967	15	0.01
968	14	0.01
969	11	0.01
970	15	0.01
971	22	0.02
972	13	0.01
973	18	0.01
974	15	0.01
975	19	0.01
976	21	0.02
977	10	0.01
978	21	0.02
979	14	0.01
980	26	0.02
981	15	0.01
982	30	0.02
983	36	0.03
984	42	0.03
985	29	0.02
986	23	0.02
987	28	0.02
988	30	0.02
989	34	0.03
990	27	0.02

Mathematics Scale Score	Frequency	Percent
991	44	0.03
992	46	0.04
993	47	0.04
994	44	0.03
995	48	0.04
996	45	0.03
997	34	0.03
998	51	0.04
999	40	0.03
1000	44	0.03
1001	50	0.04
1002	65	0.05
1003	55	0.04
1004	68	0.05
1005	83	0.06
1006	81	0.06
1007	67	0.05
1008	71	0.05
1009	82	0.06
1010	73	0.06
1011	74	0.06
1012	111	0.09
1013	95	0.07
1014	93	0.07
1015	111	0.09
1016	106	0.08
1017	114	0.09
1018	119	0.09
1019	132	0.10
1020	113	0.09
1021	135	0.10
1022	147	0.11
1023	146	0.11
1024	166	0.13
1025	183	0.14
1026	160	0.12
1027	151	0.12
1028	210	0.16
1029	197	0.15
1030	174	0.13
1031	195	0.15

Mathematics Scale Score	Frequency	Percent
1032	225	0.17
1033	232	0.18
1034	241	0.19
1035	224	0.17
1036	233	0.18
1037	254	0.20
1038	253	0.19
1039	295	0.23
1040	290	0.22
1041	314	0.24
1042	345	0.27
1043	338	0.26
1044	358	0.28
1045	388	0.30
1046	372	0.29
1047	333	0.26
1048	422	0.33
1049	413	0.32
1050	432	0.33
1051	415	0.32
1052	458	0.35
1053	530	0.41
1054	521	0.40
1055	520	0.40
1056	560	0.43
1057	548	0.42
1058	626	0.48
1059	646	0.50
1060	668	0.51
1061	667	0.51
1062	706	0.54
1063	745	0.57
1064	760	0.59
1065	811	0.62
1066	845	0.65
1067	893	0.69
1068	892	0.69
1069	892	0.69
1070	984	0.76
1071	998	0.77
1072	1046	0.81
1073	1072	0.83
1074	1062	0.82

Mathematics Scale Score	Frequency	Percent
1075	1162	0.90
1076	1208	0.93
1077	1181	0.91
1078	1350	1.04
1079	1318	1.02
1080	1340	1.03
1081	1392	1.07
1082	1421	1.09
1083	1525	1.17
1084	1465	1.13
1085	1527	1.18
1086	1543	1.19
1087	1577	1.21
1088	1675	1.29
1089	1689	1.30
1090	1730	1.33
1091	1761	1.36
1092	1894	1.46
1093	1849	1.42
1094	1823	1.40
1095	1904	1.47
1096	1987	1.53
1097	2045	1.58
1098	2054	1.58
1099	2065	1.59
1100	2140	1.65
1101	2032	1.57
1102	1974	1.52
1103	2119	1.63
1104	2122	1.63
1105	2048	1.58
1106	2037	1.57
1107	1905	1.47
1108	1998	1.54
1109	2009	1.55
1110	1882	1.45
1111	1883	1.45
1112	1842	1.42
1113	1756	1.35
1114	1773	1.37
1115	1676	1.29
1116	1663	1.28
1117	1593	1.23

Mathematics Scale Score	Frequency	Percent
1118	1455	1.12
1119	1445	1.11
1120	1429	1.10
1121	1313	1.01
1122	1343	1.03
1123	1228	0.95
1124	1170	0.90
1125	1153	0.89
1126	1036	0.80
1127	1019	0.79
1128	959	0.74
1129	917	0.71
1130	869	0.67
1131	796	0.61
1132	748	0.58
1133	708	0.55
1134	665	0.51
1135	639	0.49
1136	598	0.46
1137	554	0.43
1138	498	0.38
1139	487	0.38
1140	400	0.31
1141	401	0.31
1142	366	0.28
1143	331	0.26
1144	324	0.25
1145	285	0.22
1146	241	0.19
1147	239	0.18
1148	220	0.17
1149	199	0.15
1150	183	0.14
1151	162	0.12
1152	148	0.11
1153	120	0.09
1154	107	0.08
1155	100	0.08
1156	104	0.08
1157	77	0.06
1158	79	0.06
1159	84	0.06

Mathematics Scale Score	Frequency	Percent
1160	56	0.04
1161	54	0.04
1162	51	0.04
1163	52	0.04
1164	47	0.04
1165	36	0.03
1166	30	0.02
1167	25	0.02
1168	21	0.02
1169	24	0.02
1170	17	0.01
1171	12	0.01
1172	12	0.01
1173	13	0.01
1174	13	0.01
1175	9	0.01
1176	10	0.01
1177	8	0.01
1178	14	0.01
1179	2	0.00
1180	6	0.00
1181	5	0.00
1182	10	0.01
1183	3	0.00
1184	4	0.00
1185	6	0.00
1186	3	0.00
1187	4	0.00
1188	4	0.00
1189	2	0.00
1191	1	0.00
1192	2	0.00
1193	1	0.00
1194	3	0.00
1195	2	0.00
1196	1	0.00
1199	1	0.00
1200	3	0.00
1201	2	0.00
1250	14	0.01

Table 11-4. MME Spring 2008 Reading Frequencies for Total Group—All Forms Included

Reading Scale Score	Frequency	Percent
950	400	0.31
951	3	0.00
952	3	0.00
953	5	0.00
954	8	0.01
955	9	0.01
956	10	0.01
957	6	0.00
958	7	0.01
959	10	0.01
960	6	0.00
961	12	0.01
962	10	0.01
963	14	0.01
964	10	0.01
965	16	0.01
966	9	0.01
967	13	0.01
968	8	0.01
969	15	0.01
970	15	0.01
971	14	0.01
972	16	0.01
973	10	0.01
974	13	0.01
975	14	0.01
976	13	0.01
977	15	0.01
978	20	0.02
979	19	0.01
980	19	0.01
981	16	0.01
982	17	0.01
983	12	0.01
984	16	0.01
985	20	0.02
986	22	0.02
987	22	0.02
988	21	0.02
989	28	0.02
990	25	0.02

Reading Scale Score	Frequency	Percent
991	31	0.02
992	22	0.02
993	36	0.03
994	25	0.02
995	21	0.02
996	29	0.02
997	33	0.03
998	27	0.02
999	44	0.03
1000	34	0.03
1001	35	0.03
1002	37	0.03
1003	40	0.03
1004	32	0.02
1005	38	0.03
1006	40	0.03
1007	49	0.04
1008	61	0.05
1009	45	0.03
1010	39	0.03
1011	48	0.04
1012	52	0.04
1013	50	0.04
1014	64	0.05
1015	51	0.04
1016	68	0.05
1017	72	0.06
1018	69	0.05
1019	68	0.05
1020	63	0.05
1021	70	0.05
1022	70	0.05
1023	72	0.06
1024	79	0.06
1025	67	0.05
1026	69	0.05
1027	77	0.06
1028	75	0.06
1029	92	0.07
1030	99	0.08
1031	96	0.07

Reading Scale Score	Frequency	Percent
1032	87	0.07
1033	123	0.09
1034	131	0.10
1035	127	0.10
1036	138	0.11
1037	136	0.10
1038	150	0.12
1039	164	0.13
1040	155	0.12
1041	184	0.14
1042	192	0.15
1043	176	0.14
1044	195	0.15
1045	214	0.16
1046	250	0.19
1047	251	0.19
1048	258	0.20
1049	243	0.19
1050	286	0.22
1051	300	0.23
1052	308	0.24
1053	326	0.25
1054	322	0.25
1055	336	0.26
1056	344	0.26
1057	394	0.30
1058	412	0.32
1059	457	0.35
1060	463	0.36
1061	489	0.38
1062	500	0.38
1063	518	0.40
1064	574	0.44
1065	582	0.45
1066	611	0.47
1067	600	0.46
1068	669	0.51
1069	660	0.51
1070	715	0.55
1071	757	0.58
1072	704	0.54
1073	823	0.63
1074	829	0.64

Reading Scale Score	Frequency	Percent
1075	826	0.63
1076	872	0.67
1077	916	0.70
1078	885	0.68
1079	941	0.72
1080	1024	0.79
1081	919	0.71
1082	1038	0.80
1083	1075	0.83
1084	1128	0.87
1085	1141	0.88
1086	1194	0.92
1087	1242	0.95
1088	1245	0.96
1089	1254	0.96
1090	1242	0.95
1091	1408	1.08
1092	1356	1.04
1093	1360	1.04
1094	1396	1.07
1095	1444	1.11
1096	1447	1.11
1097	1488	1.14
1098	1517	1.16
1099	1510	1.16
1100	1511	1.16
1101	1575	1.21
1102	1514	1.16
1103	1654	1.27
1104	1603	1.23
1105	1566	1.20
1106	1624	1.25
1107	1638	1.26
1108	1616	1.24
1109	1660	1.27
1110	1738	1.33
1111	1624	1.25
1112	1713	1.32
1113	1752	1.35
1114	1702	1.31
1115	1764	1.35
1116	1796	1.38
1117	1696	1.30

Reading Scale Score	Frequency	Percent
1118	1749	1.34
1119	1829	1.40
1120	1817	1.40
1121	1786	1.37
1122	1809	1.39
1123	1871	1.44
1124	1807	1.39
1125	1828	1.40
1126	1794	1.38
1127	1835	1.41
1128	1736	1.33
1129	1762	1.35
1130	1736	1.33
1131	1725	1.32
1132	1700	1.31
1133	1623	1.25
1134	1534	1.18
1135	1382	1.06
1136	1410	1.08
1137	1280	0.98
1138	1241	0.95
1139	1156	0.89
1140	1082	0.83
1141	1009	0.77
1142	974	0.75
1143	862	0.66
1144	837	0.64
1145	766	0.59
1146	698	0.54
1147	657	0.50
1148	637	0.49
1149	570	0.44
1150	570	0.44
1151	541	0.42
1152	435	0.33
1153	406	0.31
1154	413	0.32
1155	380	0.29
1156	345	0.26
1157	344	0.26
1158	281	0.22
1159	275	0.21
1160	254	0.20

Reading Scale Score	Frequency	Percent
1161	212	0.16
1162	188	0.14
1163	202	0.16
1164	174	0.13
1165	153	0.12
1166	152	0.12
1167	134	0.10
1168	105	0.08
1169	113	0.09
1170	109	0.08
1171	87	0.07
1172	76	0.06
1173	82	0.06
1174	58	0.04
1175	48	0.04
1176	55	0.04
1177	45	0.03
1178	44	0.03
1179	29	0.02
1180	52	0.04
1181	40	0.03
1182	31	0.02
1183	33	0.03
1184	22	0.02
1185	45	0.03
1186	18	0.01
1187	27	0.02
1188	27	0.02
1189	19	0.01
1190	19	0.01
1191	18	0.01
1192	16	0.01
1193	6	0.00
1194	10	0.01
1195	10	0.01
1196	13	0.01
1197	4	0.00
1198	12	0.01
1199	15	0.01
1200	6	0.00
1201	2	0.00
1202	3	0.00
1203	5	0.00

Reading Scale Score	Frequency	Percent
1204	2	0.00
1205	1	0.00
1206	6	0.00
1207	3	0.00
1208	3	0.00
1209	6	0.00
1210	1	0.00
1211	2	0.00

Reading Scale Score	Frequency	Percent
1212	1	0.00
1213	10	0.01
1221	3	0.00
1222	8	0.01
1227	1	0.00
1238	1	0.00
1248	1	0.00
1250	15	0.01

Table 11-5. MME Spring 2008 Science Frequencies for Total Group—All Forms Included

Science Scale Score	Frequency	Percent
950	3159	2.44
951	1	0.00
952	2	0.00
953	3	0.00
954	5	0.00
955	4	0.00
956	2	0.00
957	2	0.00
958	5	0.00
959	9	0.01
960	2	0.00
961	4	0.00
962	2	0.00
963	3	0.00
964	3	0.00
965	3	0.00
966	7	0.01
967	3	0.00
968	2	0.00
969	6	0.00
970	2	0.00
971	8	0.01
972	5	0.00
973	5	0.00
974	5	0.00
975	3	0.00
976	3	0.00
977	7	0.01
978	8	0.01
979	13	0.01
980	4	0.00
981	7	0.01
982	4	0.00
983	6	0.00
984	9	0.01
985	6	0.00
986	12	0.01
987	10	0.01
988	14	0.01
989	10	0.01
990	22	0.02
991	13	0.01

Science Scale Score	Frequency	Percent
992	18	0.01
993	23	0.02
994	21	0.02
995	24	0.02
996	21	0.02
997	23	0.02
998	24	0.02
999	29	0.02
1000	30	0.02
1001	42	0.03
1002	23	0.02
1003	36	0.03
1004	46	0.04
1005	26	0.02
1006	34	0.03
1007	42	0.03
1008	46	0.04
1009	46	0.04
1010	45	0.03
1011	49	0.04
1012	65	0.05
1013	67	0.05
1014	60	0.05
1015	74	0.06
1016	72	0.06
1017	51	0.04
1018	78	0.06
1019	64	0.05
1020	78	0.06
1021	76	0.06
1022	86	0.07
1023	81	0.06
1024	75	0.06
1025	100	0.08
1026	106	0.08
1027	90	0.07
1028	103	0.08
1029	100	0.08
1030	105	0.08
1031	103	0.08
1032	103	0.08
1033	117	0.09

Science Scale Score	Frequency	Percent
1034	112	0.09
1035	125	0.10
1036	129	0.10
1037	142	0.11
1038	169	0.13
1039	157	0.12
1040	152	0.12
1041	171	0.13
1042	193	0.15
1043	190	0.15
1044	197	0.15
1045	232	0.18
1046	235	0.18
1047	220	0.17
1048	262	0.20
1049	263	0.20
1050	271	0.21
1051	263	0.20
1052	274	0.21
1053	300	0.23
1054	352	0.27
1055	297	0.23
1056	381	0.29
1057	386	0.30
1058	396	0.31
1059	441	0.34
1060	396	0.31
1061	426	0.33
1062	468	0.36
1063	444	0.34
1064	455	0.35
1065	566	0.44
1066	578	0.45
1067	551	0.42
1068	633	0.49
1069	672	0.52
1070	692	0.53
1071	718	0.55
1072	812	0.63
1073	829	0.64
1074	839	0.65
1075	945	0.73
1076	946	0.73
1077	939	0.72

Science Scale Score	Frequency	Percent
1078	1043	0.80
1079	1057	0.82
1080	1022	0.79
1081	1145	0.88
1082	1145	0.88
1083	1211	0.93
1084	1294	1.00
1085	1279	0.99
1086	1303	1.00
1087	1420	1.09
1088	1387	1.07
1089	1554	1.20
1090	1560	1.20
1091	1650	1.27
1092	1567	1.21
1093	1611	1.24
1094	1609	1.24
1095	1748	1.35
1096	1791	1.38
1097	1753	1.35
1098	1859	1.43
1099	1889	1.46
1100	1812	1.40
1101	1910	1.47
1102	1856	1.43
1103	1854	1.43
1104	1869	1.44
1105	1869	1.44
1106	1899	1.46
1107	1916	1.48
1108	1940	1.50
1109	1854	1.43
1110	1923	1.48
1111	1851	1.43
1112	1810	1.40
1113	1925	1.48
1114	1882	1.45
1115	1813	1.40
1116	1799	1.39
1117	1790	1.38
1118	1685	1.30
1119	1770	1.36
1120	1668	1.29
1121	1699	1.31

Science Scale Score	Frequency	Percent
1122	1620	1.25
1123	1648	1.27
1124	1615	1.25
1125	1419	1.09
1126	1512	1.17
1127	1451	1.12
1128	1409	1.09
1129	1352	1.04
1130	1334	1.03
1131	1227	0.95
1132	1210	0.93
1133	1163	0.90
1134	1149	0.89
1135	1091	0.84
1136	1054	0.81
1137	964	0.74
1138	901	0.69
1139	834	0.64
1140	824	0.64
1141	747	0.58
1142	726	0.56
1143	634	0.49
1144	650	0.50
1145	561	0.43
1146	548	0.42
1147	482	0.37
1148	463	0.36
1149	464	0.36
1150	394	0.30
1151	375	0.29
1152	323	0.25
1153	291	0.22
1154	292	0.23
1155	236	0.18
1156	200	0.15
1157	214	0.17
1158	186	0.14
1159	180	0.14
1160	145	0.11
1161	174	0.13
1162	123	0.09
1163	102	0.08
1164	93	0.07
1165	85	0.07

Science Scale Score	Frequency	Percent
1166	76	0.06
1167	59	0.05
1168	67	0.05
1169	62	0.05
1170	49	0.04
1171	42	0.03
1172	38	0.03
1173	49	0.04
1174	32	0.02
1175	30	0.02
1176	28	0.02
1177	20	0.02
1178	22	0.02
1179	16	0.01
1180	26	0.02
1181	22	0.02
1182	13	0.01
1183	14	0.01
1184	12	0.01
1185	11	0.01
1186	9	0.01
1187	12	0.01
1188	7	0.01
1189	4	0.00
1190	9	0.01
1191	9	0.01
1192	4	0.00
1193	2	0.00
1194	7	0.01
1195	7	0.01
1196	1	0.00
1197	5	0.00
1198	1	0.00
1199	2	0.00
1201	3	0.00
1202	2	0.00
1203	2	0.00
1204	8	0.01
1205	1	0.00
1210	4	0.00
1214	1	0.00
1215	1	0.00
1218	2	0.00
1219	1	0.00

Science Scale Score	Frequency	Percent
1250	19	0.01

Table 11-6. MME Spring 2008 Social Studies Frequencies for Total Group—All Forms Included

Social Studies Scale Score	Frequency	Percent
967	1	0.00
990	1	0.00
999	1	0.00
1019	6	0.00
1024	1	0.00
1030	6	0.00
1034	2	0.00
1039	9	0.01
1042	6	0.00
1046	23	0.02
1048	1	0.00
1051	42	0.03
1054	8	0.01
1055	1	0.00
1056	87	0.07
1059	21	0.02
1061	154	0.12
1063	26	0.02
1065	306	0.23
1067	27	0.02
1069	525	0.40
1071	50	0.04
1072	776	0.59
1074	76	0.06
1075	1110	0.85
1077	111	0.08
1078	1381	1.05
1080	120	0.09
1081	1769	1.35
1082	3	0.00
1083	135	0.10
1084	2016	1.54
1085	2	0.00
1086	160	0.12
1087	2363	1.80
1088	3	0.00
1089	2633	2.01
1090	1	0.00
1092	2753	2.10
1094	3015	2.30
1095	3	0.00
1097	3217	2.46
1098	2	0.00
1099	3278	2.50
1100	2	0.00
1101	3472	2.65
1102	2	0.00
1104	3524	2.69
1105	1	0.00
1106	3706	2.83
1107	2	0.00
1108	3690	2.82
1109	1	0.00
1110	3643	2.78
1111	173	0.13
1112	3865	2.95
1113	150	0.11
1115	4012	3.06
1116	2	0.00
1117	4054	3.10
1118	1	0.00
1119	4003	3.06
1120	162	0.12
1121	4042	3.09
1122	149	0.11
1124	4335	3.31
1126	4085	3.12
1127	140	0.11
1128	4355	3.33
1130	4	0.00
1131	4454	3.40
1132	2	0.00
1133	4224	3.23
1134	164	0.13
1135	4188	3.20
1136	112	0.09
1137	1	0.00
1138	4151	3.17
1139	93	0.07
1141	4008	3.06
1142	126	0.10
1143	3997	3.05

Social Studies Scale Score	Frequency	Percent
1145	99	0.08
1146	3829	2.92
1148	107	0.08
1149	3732	2.85
1151	100	0.08
1152	3396	2.59
1154	90	0.07
1156	3215	2.46
1157	80	0.06
1159	2652	2.03
1161	85	0.06
1163	2539	1.94
1165	79	0.06
1167	2076	1.59
1170	47	0.04

Social Studies Scale Score	Frequency	Percent
1172	1698	1.30
1175	33	0.03
1177	1313	1.00
1180	31	0.02
1183	978	0.75
1187	27	0.02
1191	646	0.49
1195	15	0.01
1200	411	0.31
1204	7	0.01
1212	217	0.17
1217	2	0.00
1232	98	0.07
1237	2	0.00
1250	27	0.02

Table 11-7. MME Spring 2008 Writing Frequencies for Total Group—All Forms Included

Writing Scale Score	Frequency	Percent
950	494	0.38
951	1	0.00
953	3	0.00
954	3	0.00
955	1	0.00
956	3	0.00
957	2	0.00
958	4	0.00
959	7	0.01
960	2	0.00
961	4	0.00
962	1	0.00
963	3	0.00
964	1	0.00
965	3	0.00
966	4	0.00
967	8	0.01
968	9	0.01
969	8	0.01
970	6	0.00
971	11	0.01
972	6	0.00
973	15	0.01
974	16	0.01
975	19	0.01
976	18	0.01
977	16	0.01
978	13	0.01
979	23	0.02
980	13	0.01
981	8	0.01
982	19	0.01
983	13	0.01
984	9	0.01
985	8	0.01
986	9	0.01
987	8	0.01
988	16	0.01
989	16	0.01
990	14	0.01
991	25	0.02
992	26	0.02

Writing Scale Score	Frequency	Percent
993	40	0.03
994	44	0.03
995	44	0.03
996	58	0.04
997	62	0.05
998	77	0.06
999	83	0.06
1000	101	0.08
1001	91	0.07
1002	96	0.07
1003	110	0.09
1004	86	0.07
1005	98	0.08
1006	87	0.07
1007	114	0.09
1008	114	0.09
1009	106	0.08
1010	120	0.09
1011	140	0.11
1012	142	0.11
1013	139	0.11
1014	151	0.12
1015	141	0.11
1016	140	0.11
1017	165	0.13
1018	179	0.14
1019	184	0.14
1020	175	0.14
1021	173	0.13
1022	189	0.15
1023	176	0.14
1024	206	0.16
1025	220	0.17
1026	217	0.17
1027	231	0.18
1028	262	0.20
1029	261	0.20
1030	275	0.21
1031	266	0.21
1032	279	0.22
1033	313	0.24
1034	279	0.22

Writing Scale Score	Frequency	Percent
1035	290	0.22
1036	297	0.23
1037	333	0.26
1038	345	0.27
1039	355	0.27
1040	356	0.28
1041	402	0.31
1042	399	0.31
1043	403	0.31
1044	396	0.31
1045	402	0.31
1046	443	0.34
1047	462	0.36
1048	524	0.40
1049	515	0.40
1050	555	0.43
1051	587	0.45
1052	609	0.47
1053	633	0.49
1054	627	0.48
1055	645	0.50
1056	682	0.53
1057	722	0.56
1058	770	0.60
1059	736	0.57
1060	824	0.64
1061	872	0.67
1062	854	0.66
1063	938	0.72
1064	894	0.69
1065	964	0.74
1066	1020	0.79
1067	1099	0.85
1068	1103	0.85
1069	1102	0.85
1070	1154	0.89
1071	1150	0.89
1072	1216	0.94
1073	1205	0.93
1074	1274	0.98
1075	1286	0.99
1076	1331	1.03
1077	1418	1.10
1078	1423	1.10

Writing Scale Score	Frequency	Percent
1079	1385	1.07
1080	1432	1.11
1081	1487	1.15
1082	1489	1.15
1083	1542	1.19
1084	1470	1.14
1085	1575	1.22
1086	1682	1.30
1087	1579	1.22
1088	1665	1.29
1089	1715	1.33
1090	1753	1.35
1091	1723	1.33
1092	1693	1.31
1093	1786	1.38
1094	1732	1.34
1095	1662	1.28
1096	1734	1.34
1097	1794	1.39
1098	1854	1.43
1099	1744	1.35
1100	1799	1.39
1101	1712	1.32
1102	1749	1.35
1103	1834	1.42
1104	1671	1.29
1105	1773	1.37
1106	1737	1.34
1107	1692	1.31
1108	1710	1.32
1109	1759	1.36
1110	1550	1.20
1111	1573	1.22
1112	1590	1.23
1113	1501	1.16
1114	1491	1.15
1115	1390	1.07
1116	1378	1.06
1117	1317	1.02
1118	1316	1.02
1119	1218	0.94
1120	1169	0.90
1121	1236	0.96
1122	1122	0.87

Writing Scale Score	Frequency	Percent
1123	1017	0.79
1124	1017	0.79
1125	1045	0.81
1126	968	0.75
1127	853	0.66
1128	857	0.66
1129	847	0.65
1130	862	0.67
1131	738	0.57
1132	693	0.54
1133	681	0.53
1134	612	0.47
1135	587	0.45
1136	567	0.44
1137	578	0.45
1138	533	0.41
1139	475	0.37
1140	448	0.35
1141	401	0.31
1142	401	0.31
1143	382	0.30
1144	344	0.27
1145	291	0.22
1146	308	0.24
1147	258	0.20
1148	280	0.22
1149	222	0.17
1150	218	0.17
1151	195	0.15
1152	181	0.14
1153	172	0.13
1154	140	0.11
1155	141	0.11
1156	109	0.08
1157	126	0.10
1158	120	0.09
1159	104	0.08
1160	102	0.08
1161	68	0.05
1162	62	0.05
1163	69	0.05
1164	71	0.05
1165	48	0.04
1166	53	0.04

Writing Scale Score	Frequency	Percent
1167	52	0.04
1168	41	0.03
1169	36	0.03
1170	34	0.03
1171	25	0.02
1172	26	0.02
1173	19	0.01
1174	20	0.02
1175	26	0.02
1176	15	0.01
1177	9	0.01
1178	18	0.01
1179	7	0.01
1180	9	0.01
1181	13	0.01
1182	13	0.01
1183	4	0.00
1184	3	0.00
1185	6	0.00
1186	2	0.00
1187	2	0.00
1188	7	0.01
1190	4	0.00
1191	3	0.00
1192	5	0.00
1193	2	0.00
1194	3	0.00
1195	3	0.00
1197	7	0.01
1198	1	0.00
1200	5	0.00
1202	1	0.00
1203	2	0.00
1205	2	0.00
1208	2	0.00
1210	1	0.00
1211	1	0.00
1212	1	0.00
1213	1	0.00
1214	1	0.00
1222	1	0.00
1226	1	0.00
1238	1	0.00
1250	1	0.00

Table 11-8. Mean and SD of Item Parameter Estimates for Math

	2008 Spring Math		
	a	b	c
MC items	Initial Form 1		
mean	1.468	0.064	0.192
SD	0.497	0.954	0.082
MC items	Initial Form 2		
mean	1.471	0.060	0.193
SD	0.493	0.948	0.083
MC items	Initial Form 3		
mean	1.484	0.058	0.192
SD	0.502	0.951	0.083
MC items	Initial Form 4		
mean	1.474	0.052	0.192
SD	0.497	0.949	0.082
MC items	Initial Form 5		
mean	1.473	0.052	0.193
SD	0.496	0.949	0.084
MC items	Initial Form 6		
mean	1.475	0.059	0.194
SD	0.492	0.947	0.084
MC items	Initial Form 7		
mean	1.494	0.063	0.193
SD	0.498	0.951	0.083
MC items	Initial Form 8		
mean	1.480	0.057	0.192
SD	0.491	0.946	0.083
MC items	Initial Form 9		
mean	1.478	0.059	0.193
SD	0.490	0.949	0.083
MC items	Initial Form 10		
mean	1.480	0.054	0.193
SD	0.501	0.947	0.083
	Makeup Form		
	a	b	c
MC items			
mean	1.434	-0.071	0.205
SD	0.480	0.995	0.084
	Accommodated Form		
	a	b	c
MC items			
mean	1.426	-0.063	0.200
SD	0.458	0.907	0.086
	Braille Form		
	a	b	c
MC items			
mean	1.493	-0.019	0.197
SD	0.517	0.921	0.085
	Emergency Form		
	a	b	c
MC items			
mean	1.359	-0.050	0.205
SD	0.516	0.987	0.086

Table 11-9. Mean and SD of Item Parameter Estimates for Reading

	2008 Spring Reading		
	Initial Form		
	a	b	c
MC items			
mean	0.858	-0.102	0.190
SD	0.455	1.583	0.072
	Makeup Form		
	a	b	c
MC items			
mean	1.004	-0.041	0.198
SD	0.585	1.652	0.073
	Accommodated Form		
	a	b	c
MC items			
mean	0.861	-0.041	0.205
SD	0.416	1.428	0.075
	Braille Form		
	a	b	c
MC items			
mean	0.922	0.048	0.196
SD	0.422	1.429	0.078
	Emergency Form		
	a	b	c
MC items			
mean	0.917	-0.138	0.195
SD	0.513	1.588	0.069

Table 11-10. Mean and SD of Item Parameter Estimates for Science

	2008 Spring Science		
	Initial Form		
	a	b	c
MC items			
mean	0.982	0.645	0.218
SD	0.327	0.832	0.088
	Makeup Form		
	a	b	c
MC items			
mean	0.968	0.544	0.212
SD	0.358	0.908	0.091
	Accommodated Form		
	a	b	c
MC items			
mean	0.890	0.299	0.212
SD	0.326	0.941	0.081
	Braille Form		
	a	b	c
MC items			
mean	0.903	0.518	0.218
SD	0.324	0.888	0.086
	Emergency Form		
	a	b	c
MC items			
mean	1.011	0.669	0.223
SD	0.401	0.879	0.086

Table 11-11. Mean and SD of Item Parameter Estimates for Writing

	2008 Spring Writing								
	Initial Form								
	a	b	c	tau1	tau2	tau3	tau4	tau5	tau6
MC items mean SD	0.960 0.335	0.453 0.750	0.214 0.069						
ACT CR item	0.500	0.692		3.192	2.812	0.674	-2.009	-4.669	
Michigan CR item	0.664	0.422		3.646	3.527	1.299	-1.189	-3.160	-4.122
	Makeup Form								
	a	b	c	tau1	tau2	tau3	tau4	tau5	tau6
MC items mean SD	0.971 0.276	0.239 0.751	0.203 0.083						
ACT CR item	0.463	0.561		4.024	2.509	0.143	-2.466	-4.210	
Michigan CR item	0.684	-0.228		5.046	2.781	0.876	-1.426	-3.047	-4.230
	Accommodated Form								
	a	b	c	tau1	tau2	tau3	tau4	tau5	tau6
MC items mean SD	0.895 0.314	0.157 0.875	0.210 0.057						
ACT CR item	0.363	0.958		2.709	2.345	0.479	-1.894	-3.639	
Michigan CR item	0.664	0.422		3.646	3.527	1.299	-1.189	-3.160	-4.122
	Braille Form								
	a	b	c	tau1	tau2	tau3	tau4	tau5	tau6
MC items mean SD	0.836 0.301	0.351 0.840	0.220 0.066						
ACT CR item	0.363	0.958		2.709	2.345	0.479	-1.894	-3.639	
Michigan CR item	0.664	0.422		3.646	3.527	1.299	-1.189	-3.160	-4.122
	Emergency Form								
	a	b	c	tau1	tau2	tau3	tau4	tau5	tau6
MC items mean SD	0.951 0.320	0.259 0.675	0.229 0.098						
ACT CR item	0.372	-0.227		5.714	3.576	0.560	-3.475	-6.376	
Michigan CR item	0.548	0.134		3.929	2.726	0.881	-0.995	-2.808	-3.733

Table 11-12. Mean and SD of Item Parameter Estimates for Social Studies

	2008 Spring Social Studies					
	Initial Form					
	b	tau1	tau2	tau3	tau4	tau5
MC items mean SD	-0.006 0.590					
Michigan CR item	0.714	-2.781	-0.992	0.038	1.171	2.56535
	Makeup Form					
	b	tau1	tau2	tau3	tau4	tau5
MC items mean SD	-0.006 0.590					
Michigan CR item	0.711	-3.983	-0.770	0.146	1.660	2.947
	Accommodated Form (same as above)					
	b	tau1	tau2	tau3	tau4	tau5
MC items mean SD						
Michigan CR item						
	Braille Form					
	b	tau1	tau2	tau3	tau4	tau5
MC items mean SD						
Michigan CR item						
	Emergency Form					
	b	tau1	tau2	tau3	tau4	tau5
MC items mean SD	-0.125 0.747					
Michigan CR item	0.329	-2.888	-0.603	0.166	1.056	2.269

References

American Educational Research Association, American Psychological Association, National Council on Measurement in Education, & Joint Committee on Standards for Educational and Psychological Testing (U.S.). (1999) *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

Assessment and Examination Service. (2006). *Standard Setting Plan.*

Assessment and Examination Service. (2006). *Standard Setting Report.*

Crocker, L. & Algina, J. (1986). *Introduction to classical and modern test theory.* Holt, Rinehart and Winston, Inc.

Dossey, J.A. (2005). *Comparison of the ACT and WorkKeys Assessments with the Mathematics and Science Content Expectations in the Michigan Curriculum Framework.*

Green, B.F., Bock, R.D., Humphreys, L.G. Linn, R.L. & Reckase, M.D. (1984). Technical guidelines for assessing computerized adaptive tests. *Journal of Educational Measurement, 21(4),* pp. 347-360.

Lee, W.-C., Hanson, B. A., & Brennan, R. L. (2002). Estimating consistency and accuracy indices for multiple classifications. *Applied Psychological Measurement, 26(4),* 412-432.

Martineau, J. A. (2007). An extension and practical evaluation of expected classification accuracy. *Applied Psychological Measurement 31(3),* 181-194.

Michigan Department of Education. (2006). *Michigan Department of Education Memorandum.*

Michigan Department of Education. (2006). *Minutes of the State Board of Education.*

Michigan Department of Education Web Site. (2007). http://www.michigan.gov/mde/.

Muraki, E. & Bock R. D. (1997) PARSCALE: IRT item analysis and test scoring for rating-scale data. Scientific Software International: Chicago, Il.

Orlando, M. & Thissen, D. (2000). Likelihood-based item fit indices for dichotomous item response theory models. *Applied Psychological Measurement, 24(1),* 50-64.

Shanahan, T. (1996). *Review of ACT Coverage of Michigan Language Arts Standards.*

Webb, N.L. (2005). *Alignment Analysis of Language Arts Standards and Assessment: Michigan Grades 9–12.*

Webb, N.L. (2005). *Alignment Analysis of Mathematics Standards and Assessments: Michigan High School.*

Webb, N.L. (2006). *Alignment Analysis of Mathematics Standards and Michigan Merit Examination.*

Webb, N.L. (2006). *Alignment Analysis of Reading and Language Arts Standards and Michigan Merit Examination.*

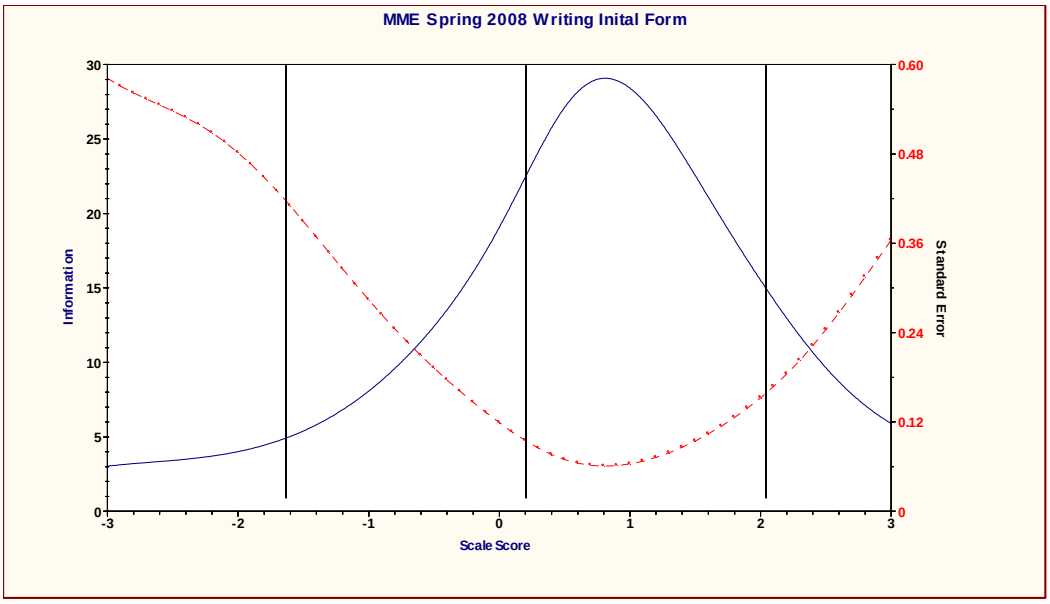
Webb, N.L. (2006). *Alignment Analysis of Science Standards and Michigan Merit Examination.*

Yen, W. M. & Fitzpatrick, A. R. (2006). Item Response Theory. In R. L. Brennan (Ed.) *Educational Measurement* (4th edition, pp. 111-153). Westport, CT: Praeger.

Appendices

Appendix A Plots of PARSCALE Information function

Spring 2008 Writing Initial Form

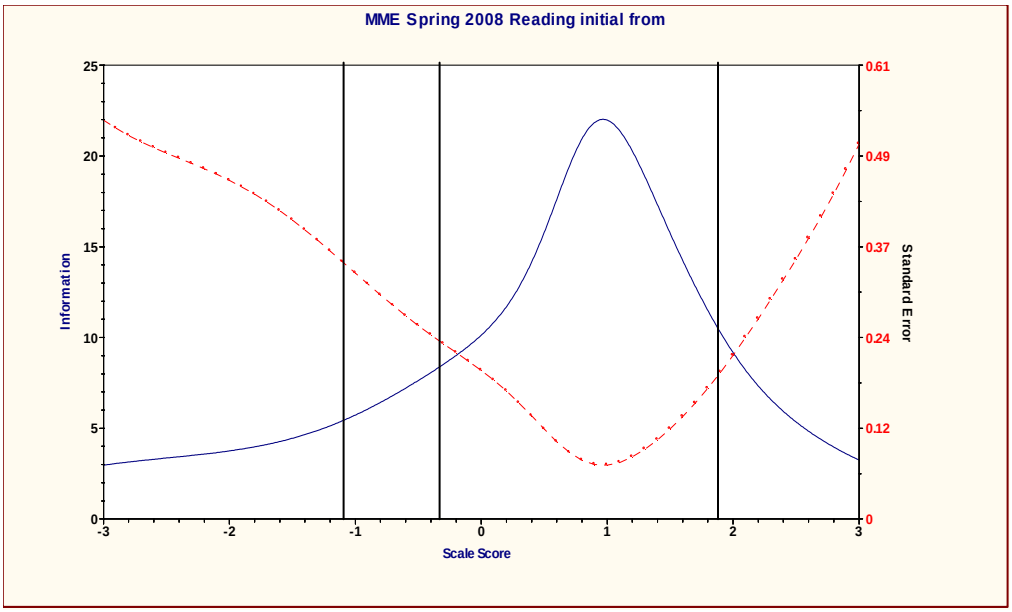


Test information curve: solid line **Standard error curve: dotted line**

The total test information for a specific scale score is read from the left vertical axis.

The standard error for a specific scale score is read from the right vertical axis.

Spring 2008 Reading Initial Form

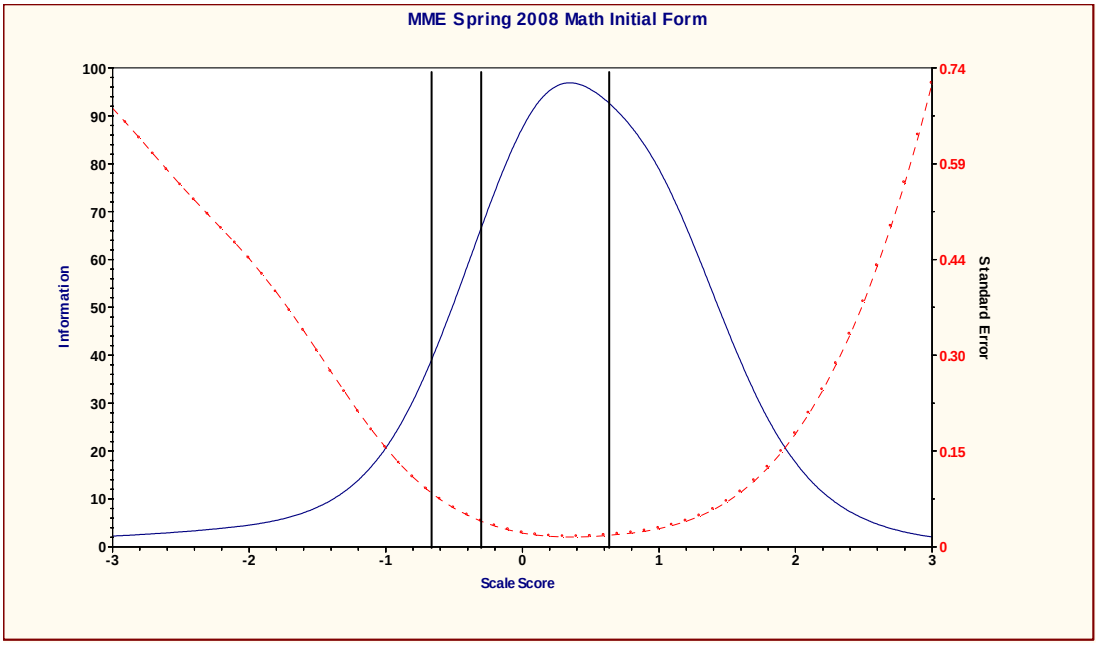


Test information curve: solid line **Standard error curve: dotted line**

The total test information for a specific scale score is read from the left vertical axis.

The standard error for a specific scale score is read from the right vertical axis.

Spring 2008 Mathematics Initial Form

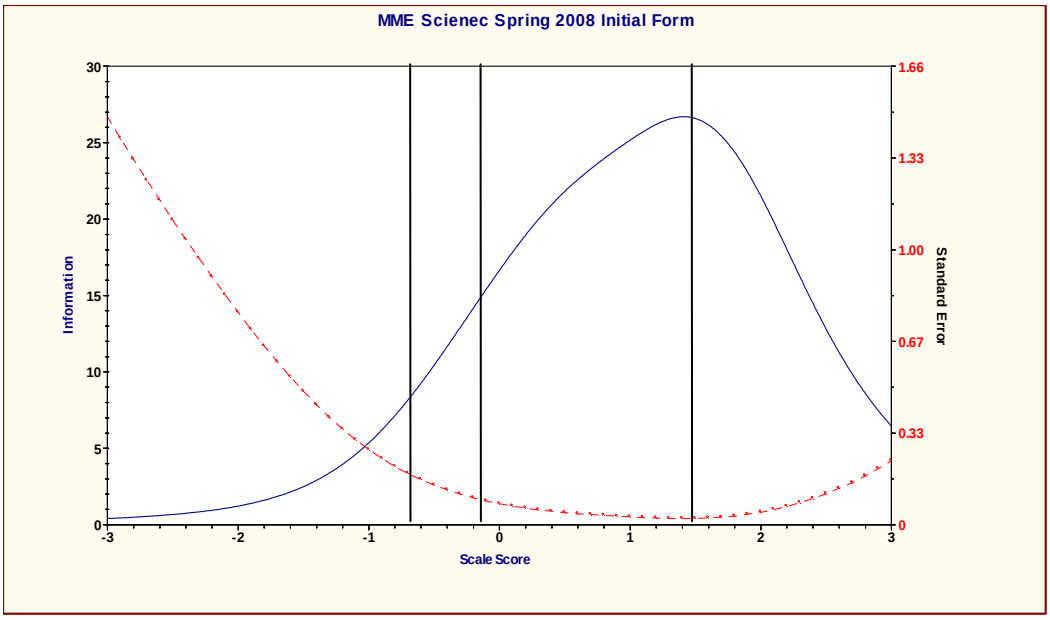


Test information curve: solid line **Standard error curve: dotted line**

The total test information for a specific scale score is read from the left vertical axis.

The standard error for a specific scale score is read from the right vertical axis.

Spring 2008 Science Initial Form



Test information curve: solid line **Standard error curve: dotted line**

The total test information for a specific scale score is read from the left vertical axis.

The standard error for a specific scale score is read from the right vertical axis.

Appendix B: Data Created for Field-Test Items

Field Format	Field Name	Field Description	Notes		Computation Description
A6	TEST	Test Name	MATH11, READ11, SCIE11, SOCS11, WRIT11		(From Test Map-titles)
A2	SUBJ	Subject (RE, MA, SC, SS, WR)	REading, MAthematics, SCience, Social Studies, WRiting		(From Test Map-titles)
A2	GRADE	Grade	Grade in which an item administered		11 in the spring (From Test Map)
A25	MEAP_ID	MEAP Item ID	Michigan item identifyer		(From Test Map)
F12	CID	CID (currently 7 digits used)	Company ID number for an item (HAI or PEM)		(From Test Map)
A2	TYPE	Item Type (MC, CR)	MC - multiple-choice, CR - constructed response		(From Test Map)
A1	KEY	Item Answer Key (A, B, C, D)	For MC items		(From Test Map)
F1	MAX	Item Maximal Score	For CR items		(From Test Map)
A3	STRAND	Item Strand			(From Test Map)
A3	BNCHM	Item Benchmark			(From Test Map)
A10	GLCE	Grade Level Expectation			(From Test Map)
F1	DEPTKN	Depth of Knowledge	**Left blank.**		This is included in the Test Maps folder in "Benchmark-GLCE Descriptors".
A2	CYCLE	Year cycle (2 characters)			Assume 07
A2	DOMAIN	Domain			(From Test Map)
A2	LEVEL	Level			(From Test Map)
A50	SCENARIO	Scenario			(From Test Map)
A4	ADMYEAR	Administration Year	For each administration year a separate line will be provided		Note this is 4 digits here, and 2 digits in Dave's SAS dump. Use 2007.
A3	RELEASED	Released position or N/R			From Test Map, position k. None are released in 2007
A3	MATURITY	PP, PI, FT, OP, RL, EM	Pre-pilot, Pilot, Field-Test, Operational, Released, Emergency		(From Test Map)
A3	FUNC	Item Function in Current Administration	Core, Future core, Extended core, Linking. **Left Blank**		(From Test Map)
A3	CHAR_COD	Character Code	See the spreadsheet 'Codes'		Attached
F2	NFORMS	Number of Forms Item Appears On (1 - 5)	Indicates how many forms a matrix item appears on, ranges 1-5 (not supplied for core items).		NA spring 07
A60	FORMS	Form Numbers (string of 3x20 characters)	Indicates which forms a matrix item appears on, there will be as many form numbers as there are forms that item appears on (not supplied for core items).		NA spring 07

A60	POSITS	Test Positions (string of 3x20 characters)	Indicates positions in the test for each form that a matrix item appears on, there will be as many position numbers as there are forms that item appears on (shows only one number for core items).		NA spring 07
F6	NCOUNT	N-count	Number of calibration cases used to produce statistics		Total number of calibration students who took the item regardless of the number of forms on which that item appears. Inclusion/exclusion rules for calibration students will be defined by OEAA
F6	N_MAL	Ncount Males	N-counts for break-down groups		Total number of calibration male students who took the item regardless of the number of forms on which that item appears
F6	N_FEM	Ncount Females			Total number of calibration female students who took the item regardless of the number of forms on which that item appears
F6	N_WHI	Ncount White			Total number of calibration white students who took the item regardless of the number of forms on which that item appears
F6	N_BLA	Ncount Black			Total number of calibration black students who took the item regardless of the number of forms on which that item appears
F2	COM1	Percent for Comment Code 1			Number of students who were assigned commnet code 1 (see codes sheet for comment code description) divided by the total number of calibration students
F2	COM2	Percent for Comment Code 2			Number of students who were assigned commnet code 2 (see codes sheet for comment code description) divided by the total number of calibration students
F2	COM3	Percent for Comment Code 3			Number of students who were assigned commnet code 3 (see codes sheet for comment code description) divided by the total number of calibration students
F2	COM4	Percent for Comment Code 4			Number of students who were assigned commnet code 4 (see codes sheet for comment code description) divided by the total number of calibration students
F2	COM5	Percent for Comment Code 5			Number of students who were assigned commnet code 5 (see codes sheet for comment code description) divided by the total number of calibration students
F2	COM6	Percent for Comment Code 6			Number of students who were assigned commnet code 6 (see codes sheet for comment code description) divided by the total number of calibration students

F2	COM7	Percent for Comment Code 7			Number of students who were assigned commnet code 7 (see codes sheet for comment code description) divided by the total number of calibration students
F2	COM8	Percent for Comment Code 8			Number of students who were assigned commnet code 8 (see codes sheet for comment code description) divided by the total number of calibration students
F2	COM9	Percent for Comment Code 9			Number of students who were assigned commnet code 9 (see codes sheet for comment code description) divided by the total number of calibration students
F2	COM10	Percent for Comment Code 10 (not used yet)			Number of students who were assigned commnet code 10 (see codes sheet for comment code description) divided by the total number of calibration students
F2	COM11	Percent for Comment Code 11 (not used yet)			Number of students who were assigned commnet code 11 (see codes sheet for comment code description) divided by the total number of calibration students
F2	COM12	Percent for Comment Code 12 (not used yet)			Number of students who were assigned commnet code 12 (see codes sheet for comment code description) divided by the total number of calibration students
F2	COND_A	Percent for Condition Code A	Codition code distribution (for CR items only, see the spreadsheet "Codes")		Number of students who were assigned condition code A (see codes sheet for condition code description) divided by the the total number of calibration students
F2	COND_B	Percent for Condition Code B			Number of students who were assigned condition code B (see codes sheet for condition code description) divided by the the total number of calibration students
F2	COND_C	Percent for Condition Code C			Number of students who were assigned condition code C (see codes sheet for condition code description) divided by the the total number of calibration students
F2	COND_D	Percent for Condition Code D			Number of students who were assigned condition code D (see codes sheet for condition code description) divided by the the total number of calibration students
F2	COND_E	Percent for Condition Code E			Number of students who were assigned condition code E (see codes sheet for condition code description) divided by the the total number of calibration students
F2	COND_F	Percent for Condition Code F (not used yet)			Number of students who were assigned condition code F (see codes sheet for condition code description) divided by the the total number of calibration students

F2	COND_G	Percent for Condition Code G (not used yet)			Number of students who were assigned condition code G (see codes sheet for condition code description) divided by the the total number of calibration students
F2	COND_H	Percent for Condition Code H (not used yet)			Number of students who were assigned condition code H (see codes sheet for condition code description) divided by the the total number of calibration students
F2	A	Percent (option A or scorepoint 0)	Percent of ALL calibration cases		Number of students who chose option A or gained a score point of 0 divided by the the total number of calibration students
F2	B	Percent (option B or scorepoint 1)			Number of students who chose option B or gained a score point of 1 divided by the the total number of calibration students
F2	C	Percent (option C or scorepoint 2)			Number of students who chose option C or gained a score point of 2 divided by the the total number of calibration students
F2	D	Percent (option D or scorepoint 3)			Number of students who chose option D or gained a score point of 3 divided by the the total number of calibration students
F2	M	Percent (mult. marks or scorepoint 4)			Number of students who chose multiple marks or gained a score point of 4 divided by the the total number of calibration students
F2	S5	Percent (scorepoint 5)			Number of students who gained a score point of 5 divided by the the total number of calibration students
F2	S6	Percent (scorepoint 6)			Number of students who gained a score point of 6 divided by the the total number of calibration students
F2	S7	Percent (scorepoint 7)			Number of students who gained a score point of 7 divided by the the total number of calibration students
F2	S8	Percent (scorepoint 8)			Number of students who gained a score point of 8 divided by the the total number of calibration students
F2	S9	Percent (scorepoint 9)			Number of students who gained a score point of 9 divided by the the total number of calibration students
F2	S10	Percent (scorepoint 10)			Number of students who gained a score point of 10 divided by the the total number of calibration students
F2	S11	Percent (scorepoint 11)			Number of students who gained a score point of 11 divided by the the total number of calibration students
F2	S12	Percent (scorepoint 12)			Number of students who gained a score point of 12 divided by the the total number of calibration

					students
F2	O	Percent (Omits)			Number of students who had omits divided by the the total number of calibration students
F2	MAA	Male Percent (A or 0)	Percent for MALE calibration cases		Number of male students who chose option A or gained a score point of 0 divided by the the total number of male calibration students
F2	MAB	Male Percent (B or 1)			Number of male students who chose option B or gained a score point of 1 divided by the the total number of male calibration students
F2	MAC	Male Percent (C or 2)			Number of male students who chose option C or gained a score point of 2 divided by the the total number of male calibration students
F2	MAD	Male Percent (D or 3)			Number of male students who chose option D or gained a score point of 3 divided by the the total number of male calibration students
F2	MAM	Male Percent (MM or 4)			Number of male students who chose multiple marks or gained a score point of 4 divided by the the total number of male calibration students
F2	MAS5	Male Percent (scorepoint 5)			Number of male students who gained a score point of 5 divided by the the total number of male calibration students
F2	MAS6	Male Percent (scorepoint 6)			Number of male students who gained a score point of 6 divided by the the total number of male calibration students
F2	MAS7	Percent (scorepoint 7)			Number of students who gained a score point of 7 divided by the the total number of calibration students
F2	MAS8	Percent (scorepoint 8)			Number of students who gained a score point of 8 divided by the the total number of calibration students
F2	MAS9	Percent (scorepoint 9)			Number of students who gained a score point of 9 divided by the the total number of calibration students
F2	MAS10	Percent (scorepoint 10)			Number of students who gained a score point of 10 divided by the the total number of calibration students
F2	MAS11	Percent (scorepoint 11)			Number of students who gained a score point of 11 divided by the the total number of calibration students
F2	MAS12	Percent (scorepoint 12)			Number of students who gained a score point of 12 divided by the the total number of calibration students
F2	MAO	Male Percent (Omits)			Number of male students who had omits divided by the the total number of male calibration students

F2	FEA	Female Percent (A or 0)	Percent for FEMALE calibration cases		Number of female students who chose option A or gained a score point of 0 divided by the the total number of female calibration students
F2	FEB	Female Percent (B or 1)			Number of female students who chose option B or gained a score point of 1 divided by the the total number of female calibration students
F2	FEC	Female Percent (C or 2)			Number of female students who chose option C or gained a score point of 2 divided by the the total number of female calibration students
F2	FED	Female Percent (D or 3)			Number of female students who chose option D or gained a score point of 3 divided by the the total number of female calibration students
F2	FEM	Female Percent (MM or 4)			Number of female students who chose multiple marks or gained a score point of 4 divided by the the total number of female calibration students
F2	FES5	Female Percent (scorepoint 5)			Number of female students who gained a score point of 5 divided by the the total number of female calibration students
F2	FES6	Female Percent (scorepoint 6)			Number of female students who gained a score point of 6 divided by the the total number of female calibration students
F2	FES7	Percent (scorepoint 7)			Number of students who gained a score point of 7 divided by the the total number of calibration students
F2	FES8	Percent (scorepoint 8)			Number of students who gained a score point of 8 divided by the the total number of calibration students
F2	FES9	Percent (scorepoint 9)			Number of students who gained a score point of 9 divided by the the total number of calibration students
F2	FES10	Percent (scorepoint 10)			Number of students who gained a score point of 10 divided by the the total number of calibration students
F2	FES11	Percent (scorepoint 11)			Number of students who gained a score point of 11 divided by the the total number of calibration students
F2	FES12	Percent (scorepoint 12)			Number of students who gained a score point of 12 divided by the the total number of calibration students
F2	FEO	Female Percent (Omits)			Number of female students who had omits divided by the the total number of female calibration students

F2	WHA	White Percent (A or 0)	Percent for WHITE calibration cases		Number of white students who chose option A or gained a score point of 0 divided by the the total number of white calibration students
F2	WHB	White Percent (B or 1)			Number of white students who chose option B or gained a score point of 1 divided by the the total number of white calibration students
F2	WHC	White Percent (C or 2)			Number of white students who chose option C or gained a score point of 2 divided by the the total number of white calibration students
F2	WHD	White Percent (D or 3)			Number of white students who chose option D or gained a score point of 3 divided by the the total number of white calibration students
F2	WHM	White Percent (MM or 4)			Number of white students who chose multiple marks or gained a score point of 4 divided by the the total number of white calibration students
F2	WHS5	White Percent (scorepoint 5)			Number of white students who gained a score point of 5 divided by the the total number of white calibration students
F2	WHS6	White Percent (scorepoint 6)			Number of white students who gained a score point of 6 divided by the the total number of white calibration students
F2	WHS7	Percent (scorepoint 7)			Number of students who gained a score point of 7 divided by the the total number of calibration students
F2	WHS8	Percent (scorepoint 8)			Number of students who gained a score point of 8 divided by the the total number of calibration students
F2	WHS9	Percent (scorepoint 9)			Number of students who gained a score point of 9 divided by the the total number of calibration students
F2	WHS10	Percent (scorepoint 10)			Number of students who gained a score point of 10 divided by the the total number of calibration students
F2	WHS11	Percent (scorepoint 11)			Number of students who gained a score point of 11 divided by the the total number of calibration students
F2	WHS12	Percent (scorepoint 12)			Number of students who gained a score point of 12 divided by the the total number of calibration students
F2	WHO	White Percent (Omits)			Number of white students who had omits divided by the the total number of white calibration students
F2	BLA	Black Percent (A or 0)	Percent for BLACK calibration cases		Number of black students who chose option A or gained a score point of 0 divided by the the total number of black calibration students

F2	BLB	Black Percent (B or 1)			Number of black students who chose option B or gained a score point of 1 divided by the the total number of black calibration students
F2	BLC	Black Percent (C or 2)			Number of black students who chose option C or gained a score point of 2 divided by the the total number of black calibration students
F2	BLD	Black Percent (D or 3)			Number of black students who chose option D or gained a score point of 3 divided by the the total number of black calibration students
F2	BLM	Black Percent (MM or 4)			Number of black students who chose multiple marks or gained a score point of 4 divided by the the total number of black calibration students
F2	BLS5	Black Percent (scorepoint 5)			Number of black students who gained a score point of 5 divided by the the total number of black calibration students
F2	BLS6	Black Percent (scorepoint 6)			Number of black students who gained a score point of 6 divided by the the total number of black calibration students
F2	BLS7	Percent (scorepoint 7)			Number of students who gained a score point of 7 divided by the the total number of calibration students
F2	BLS8	Percent (scorepoint 8)			Number of students who gained a score point of 8 divided by the the total number of calibration students
F2	BLS9	Percent (scorepoint 9)			Number of students who gained a score point of 9 divided by the the total number of calibration students
F2	BLS10	Percent (scorepoint 10)			Number of students who gained a score point of 10 divided by the the total number of calibration students
F2	BLS11	Percent (scorepoint 11)			Number of students who gained a score point of 11 divided by the the total number of calibration students
F2	BLS12	Percent (scorepoint 12)			Number of students who gained a score point of 12 divided by the the total number of calibration students
F2	BLO	Black Percent (Omits)			Number of black students who had omits divided by the the total number of black calibration students
F8.4	PVAL	P-value or Item Mean	P-value or arithmetic mean of item scores (all cases)		The sum of students' gained score divided by the total number of all students
F8.4	MPVAL	P-value or Item Mean for Male	Impact analysis: item means for break-down groups		The sum of male students' gained score divided by the total number of male students
F8.4	FPVAL	P-value or Item Mean for Female			The sum of female students' gained score divided by the total number of female students

F8.4	WPVAL	P-value or Item Mean for White			The sum of white students' gained score divided by the total number of white students
F8.4	BPVAL	P-value or Item Mean for Black			The sum of black students' gained score divided by the total number of black students
F8.4	ADJPVAL	Adjusted P-value	Adjusted P-value = (Arithmetic mean - MIN item score) / (MAX item score - MIN item score)		Difference between the arithmetic mean and the minimum item score divided by the item score range
A5	DIFFICFL	Difficulty flag	Based on Test Construction Specifications		For MC item p LT .3 or p GT .9. For CR item adj. p LT .1 or adj. p GT .9.
F8.4	SDEV	Item Standard Deviation	Standard deviation of item scores		Standard deviation of item score distribution
F8.4	ITOT	Item-Total Correlation	Pearson product-moment correlation (Point-Biserial correlation for dihotomous items)		Point-biserial correlation for MC items (see Crocker & Algina, 1986, page 317); Pearson product-moment correlation between the item score and the total test score for CR items (see Crocker & Algina, 1986, page 32-33)
F8.4	ITOTBIS	Biserial / Polyserial Correlation	For MC: biserial, for CR:polyserial (optional)		Biserial correlation for MC items (see Crocker & Algina, 1986, page 317); Polyserial correlation for CR items as its optional, we're ignoring it
F8.4	ITOTC	Point-Biserial Correlation (corrected)	For MC items (corrected for maximal possible value)		Corrected point-biserial correlation (see Crocker & Algina, 1986, page 317)
A2	ITOTFL	Item-Total correlation flag	Based on Test Construction Specifications		For MC item if pb LT .25.
F8.4	APB	P-b correlation for option A	Options point-biserial correlations (for CR items only Omits Rpb is supplied)		Point-biserial correlation for option A for a MC item when those students who chose option A is scored as 1
F8.4	BPB	P-b correlation for option B			Point-biserial correlation for option B for a MC item when those students who chose option B is scored as 1
F8.4	CPB	P-b correlation for option C			Point-biserial correlation for option C for a MC item when those students who chose option C is scored as 1
F8.4	DPB	P-b correlation for option D			Point-biserial correlation for option D for a MC item when those students who chose option D is scored as 1
F8.4	OPB	P-b correlation for Omits			Point-biserial correlation for omits for a MC item when those students who omitted the item is scored as 1
A7	MISKFL	Flag for potential miskeying	Based on Test Construction Specifications		For MC, if keyed option not the highest percentage, or any option LT 2% or any non-keyed item pb GT 0, or omit pb GT .03. For CR, if any score percentage LT 5%, or any omit GT 20%, or omit corr GT.03.
F8.4	MCHI_MF	Mantel CHSQ Male-Female	DIF analyses: Mantel chi-square (for both dichotomous and polytomous items), Mantel-Haenszel Delta and corresponding lower and upper 95% confidence interval limits for dichotomous items (not supplied for		Mantel Chi-square for male vs female comparison (See Holland & Wainer, 1993 page 40)
F8.4	MHDL_MF	Lower Limit of 95% Confidence Interval for MHD_MF			

F8.4	MHD_MF	Mantel-Haenszel Delta Male-Female	polytomous items)		Mantel Haenszel delta for male vs female comparison (See Holland & Wainer, 1993 page 41)
F8.4	MHDU_MF	Upper Limit of 95% Confindence Intervalfor MHD_MF			
F8.4	MCHI_WB	Mantel CHSQ White-Black			Mantel Chi-square for white vs black comparison (See Holland & Wainer, 1993 page 40)
F8.4	MHDL_WB	Lower Limit of 95% Confidence Interval for MHD_WB			
F8.4	MHD_WB	Mantel-Haenszel Delta White-Black			Mantel Haenszel delta for white vs black comparison (See Holland & Wainer, 1993 page 41)
F8.4	MHDU_WB	Upper Limit of 95% Confindence Intervalfor MHD_WB			
F8.4	SMDS_MF	SMD signed M-F	DIF analyses: Standardized Mean Difference (signed: mean of algebraic differences; unsigned: mean of absolute differences); Effect size of signed SMD		Standardized mean difference for male vs female comparison. See Zwick & Thayer (1996)
F8.4	SMDES_MF	SMD signed Effect Size for M-F			Signed SMD for male vs. female comparison divided by pooled standard deviation
F8.4	SMDU_MF	SMD unsigned M-F			Mean of absolute difference for male vs female comparison
F8.4	SMDS_WB	SMD signed W-B			Standardized mean difference for white vs black comparison. See Zwick & Thayer (1996)
F8.4	SMDES_WB	SMD signed Effect Size for W-B			Signed SMD for white vs black comparison divided by pooled standard deviation
F8.4	SMDU_WB	SMD unsigned W-B			Mean of absolute difference for white vs black comparison
A2	DIF_MF	DIF category for M-F (A, B, C)	DIF level categorization: A - no or negligible, B - moderate, C - substantial.		Items are classified as A category of DIF if either MH D-DIF is not statistically different from zero (using the 5% significance level) or if the magnitude of the MH D-DIF values is less than one delta unit in absolute value. Items are classified as C category of DIF if MH D-DIF both exceeds 1.5 in absolute value and is statistically significantly larger than 1.0 in absolute value (using the 5% significance level). All other items are classified as category B. The SMD effect size groups each item into one of three categories: negligible DIF (AA), moderate DIF (BB), and large DIF (CC). If the probability is > 0.05, items are classified as AA. Otherwise, items are classified as AA if the effect size of SMD LT 0.17. Items are classified as BB if the effect size = > 0.17 but <= 0.25. Items are classified as CC if the effect size is > 0.25
A2	DIF_WB	DIF category for W-B (A, B, C)			
A6	FG_MF	Favored group for M-F (Male, Female)	Favored group if DIF level equal to B or C		
A6	FG_WB	Favored group for W-B (White,			

		Black)			
F8.5	APAR_R1	A parameter (scaled) for rater 1	For both dichotmous and polytomous items.		Item discrimination parameter from IRT calibration and equaitng
F8.5	ASE_R1	SE for A parameter (scaled) for rater 1	For both dichotmous and polytomous items.		Standard error for item discrimination parameter from IRT calibration and equaitng
F8.5	BPAR_R1	B parameter (scaled) for rater 1	For both dichotmous and polytomous items.		Item difficulty parameter from IRT calibration and equaitng
F8.5	BSE_R1	SE for B parameter (scaled) for rater 1	For both dichotmous and polytomous items.		Standard error for item difficulty parameter from IRT calibration and equaitng
F8.5	APAR_R2	A parameter (scaled) for rater 2	For polytomous item only.		Item discrimination parameter from IRT calibration and equaitng
F8.5	ASE_R2	SE for A parameter (scaled) for rater 2	For polytomous item only.		Standard error for item discrimination parameter from IRT calibration and equaitng
F8.5	BPAR_R2	B parameter (scaled) for rater 2	For polytomous item only.		Item difficulty parameter from IRT calibration and equaitng
F8.5	BSE_R2	SE for B parameter (scaled) for rater 2	For polytomous item only.		Standard error for item difficulty parameter from IRT calibration and equaitng
F8.5	CPAR	C parameter (scaled)			Item pseudo-guessing parameter from IRT calibration and equaitng
F8.5	CSE	SE for C parameter (scaled)			Standard error for item pseudo-guessing parameter from IRT calibration and equaitng
F8.5	D1_R1	D1 category parameter (scaled) for rater 1	For both Writing and Social Studies CR items.		Item step 1 difficulty parameter from IRT calibration and equaitng
F8.5	D1SE_R1	SE for D1 category parameter (scaled) for rater 1	For both Writing and Social Studies CR items.		Standard error for item step 1 difficulty parameter from IRT calibration and equaitng
F8.5	D2_R1	D2 category parameter (scaled) for rater 1	For both Writing and Social Studies CR items.		Item step 2 difficulty parameter from IRT calibration and equaitng
F8.5	D2SE_R1	SE for D2 category parameter (scaled) for rater 1	For both Writing and Social Studies CR items.		Standard error for item step 2 difficulty parameter from IRT calibration and equaitng
F8.5	D3_R1	D3 category parameter (scaled) for rater 1	For both Writing and Social Studies CR items.		Item step 3 difficulty parameter from IRT calibration and equaitng
F8.5	D3SE_R1	SE for D3 category parameter (scaled) for rater 1	For both Writing and Social Studies CR items.		Standard error for item step 3 difficulty parameter from IRT calibration and equaitng
F8.5	D4_R1	D4 category parameter (scaled) for rater 1	For both Writing and Social Studies CR items.		Item step 4 difficulty parameter from IRT calibration and equaitng
F8.5	D4SE_R1	SE for D4 category parameter (scaled) for rater 1	For both Writing and Social Studies CR items.		Standard error for item step 4 difficulty parameter from IRT calibration and equaitng
F8.5	D5_R1	D5 category parameter (scaled) for rater 1	For both Writing and Social Studies CR items.		Item step 5 difficulty parameter from IRT calibration and equaitng
F8.5	D5SE_R1	SE for D5 category parameter (scaled) for rater 1	For both Writing and Social Studies CR items.		Standard error for item step 5 difficulty parameter from IRT calibration and equaitng

F8.5	D6_R1	D6 category parameter (scaled) for rater 1	For Writing CR items only.		Item step 6 difficulty parameter from IRT calibration and equaitng
F8.5	D6SE_R1	SE for D6 category parameter (scaled) for rater 1	For Writing CR items only.		Standard error for item step 6 difficulty parameter from IRT calibration and equaitng
F8.5	D1_R2	D1 category parameter (scaled) for rater 2	For both Writing and Social Studies CR items.		Item step 1 difficulty parameter from IRT calibration and equaitng
F8.5	D1SE_R2	SE for D1 category parameter (scaled) for rater 2	For both Writing and Social Studies CR items.		Standard error for item step 1 difficulty parameter from IRT calibration and equaitng
F8.5	D2_R2	D2 category parameter (scaled) for rater 2	For both Writing and Social Studies CR items.		Item step 2 difficulty parameter from IRT calibration and equaitng
F8.5	D2SE_R2	SE for D2 category parameter (scaled) for rater 2	For both Writing and Social Studies CR items.		Standard error for item step 2 difficulty parameter from IRT calibration and equaitng
F8.5	D3_R2	D3 category parameter (scaled) for rater 2	For both Writing and Social Studies CR items.		Item step 3 difficulty parameter from IRT calibration and equaitng
F8.5	D3SE_R2	SE for D3 category parameter (scaled) for rater 2	For both Writing and Social Studies CR items.		Standard error for item step 3 difficulty parameter from IRT calibration and equaitng
F8.5	D4_R2	D4 category parameter (scaled) for rater 2	For both Writing and Social Studies CR items.		Item step 4 difficulty parameter from IRT calibration and equaitng
F8.5	D4SE_R2	SE for D4 category parameter (scaled) for rater 2	For both Writing and Social Studies CR items.		Standard error for item step 4 difficulty parameter from IRT calibration and equaitng
F8.5	D5_R2	D5 category parameter (scaled) for rater 2	For both Writing and Social Studies CR items.		Item step 5 difficulty parameter from IRT calibration and equaitng
F8.5	D5SE_R2	SE for D5 category parameter (scaled) for rater 2	For both Writing and Social Studies CR items.		Standard error for item step 5 difficulty parameter from IRT calibration and equaitng
F8.5	D6_R2	D6 category parameter (scaled) for rater 2	For Writing CR items only.		Item step 6 difficulty parameter from IRT calibration and equaitng
F8.5	D6SE_R2	SE for D6 category parameter (scaled) for rater 2	For Writing CR items only.		Standard error for item step 6 difficulty parameter from IRT calibration and equaitng
F8.4	MSQIN1	Mean-square infit	Rasch fit index and flag: blank (0.5 < 1.5), MM (misfit moderate: 1.5 < 2.0), MH (misfit high: 2.0 <), TP (too predicTable: < 0.5). Not supplied for 3PL and 2PPC models.		Infit index output from Winsteps calibration
F8.4	MSQOUT1	Mean-square outfit			Outfit index output from Winsteps calibration
A2	MSQFITFL1	Mean-square fit flag (blank, MM, MH, TP)			
F1	FITLEV1	Misfit level (0, 1, 2)			Mean-squares > 2 indicate distorting or degrading the measurement system, flagged as misfit level 2. 1.5 – 2 means unproductive for construction of measurement, but not degrading, flagged as misfit level 1. < 0.5 means less productive for measurement, but not degrading. It may produce misleadingly good reliabilities and separations, flagged as misfit level 1. Otherwise, no flag with a misfit level of 0

F8.4	MSQIN2	Mean-square infit	Rasch fit index and flag: blank (0.5 < 1.5), MM (misfit moderate: 1.5 < 2.0), MH (misfit high: 2.0 <), TP (too predicTable: < 0.5). Not supplied for 3PL and 2PPC models.		Infit index output from Winsteps calibration
F8.4	MSQOUT2	Mean-square outfit			Outfit index output from Winsteps calibration
A2	MSQFITFL2	Mean-square fit flag (blank, MM, MH, TP)			
F1	FITLEV2	Misfit level (0, 1, 2)			Mean-squares > 2 indicate distorting or degrading the measurement system, flagged as misfit level 2. 1.5 – 2 means unproductive for construction of measurement, but not degrading, flagged as misfit level 1. < 0.5 means less productive for measurement, but not degrading. It may produce misleadingly good reliabilities and separations, flagged as misfit level 1. Otherwise, no flag with a misfit level of 0
F10.3	CHISQ	Chi-square statistics for 3PL and GPC fit index computed by PARSCALE	For CR item (rater 1) and dichotomous items.		Use ITEMFIT = 10 to specify the number (10) of frequency score groups to be used for computation of item-fit index in PARSCALE callibration runs. Note 10 deciles are used for other item statistics.
F5.0	DF	Degrees of freedom associated with the Chi-square fit index computed by PARSCALE.	For CR item (rater 1) and dichotomous items.		
F5.3	P_CHISQ	P-value associated with the Chi-square fit index computed by PARSCALE.	For CR item (rater 1) and dichotomous items.		
F10.3	CHISQ_R2	Chi-square statistics for GPC fit index computed by PARSCALE.	For CR item (rater 2) only.		Use ITEMFIT = 10 to specify the number (10) of frequency score groups to be used for computation of item-fit index in PARSCALE callibration runs. Note 10 deciles are used for other item statistics.
F5.0	DF_R2	Degrees of freedom associated with the Chi-square fit index computed by PARSCALE.	For CR item (rater 2) only.		
F5.3	P_CHISQ_R2	P-value associated with the Chi-square fit index computed by PARSCALE.	For CR item (rater 2) only.		
F8.5	INFO1	Item information at cut point 1	Item information at performance level cut-points.		Item information computed at cut score 1 based on Hambleton & Swaminathan (1985, page 106-107)
F8.5	INFO2	Item information at cut point 2			Item information computed at cut score 2 based on Hambleton & Swaminathan (1985, page 106-107)
F8.5	INFO3	Item information at cut point 3			Item information computed at cut score 3 based on Hambleton & Swaminathan (1985, page 106-107)
F8.3	TH01	Theta point 1	Theta points for plotting conditional item means.		Theta point corresponding to decile 1 (lowest 10%)
F8.3	TH02	Theta point 2			Theta point corresponding to decile 2
F8.3	TH03	Theta point 3			Theta point corresponding to decile 3
F8.3	TH04	Theta point 4			Theta point corresponding to decile 4

F8.3	TH05	Theta point 5			Theta point corresponding to decile 5
F8.3	TH06	Theta point 6			Theta point corresponding to decile 6
F8.3	TH07	Theta point 7			Theta point corresponding to decile 7
F8.3	TH08	Theta point 8			Theta point corresponding to decile 8
F8.3	TH09	Theta point 9			Theta point corresponding to decile 9
F8.3	TH10	Theta point 10			Theta point corresponding to decile 10 (highest 10%)
F8.3	AD01	Conditional Item Mean for Decile 1	Conditional item means plot: All		Item mean for decile 1 for all students
F8.3	AD02	Conditional Item Mean for Decile 2			Item mean for decile 2 for all students
F8.3	AD03	Conditional Item Mean for Decile 3			Item mean for decile 3 for all students
F8.3	AD04	Conditional Item Mean for Decile 4			Item mean for decile 4 for all students
F8.3	AD05	Conditional Item Mean for Decile 5			Item mean for decile 5 for all students
F8.3	AD06	Conditional Item Mean for Decile 6			Item mean for decile 6 for all students
F8.3	AD07	Conditional Item Mean for Decile 7			Item mean for decile 7 for all students
F8.3	AD08	Conditional Item Mean for Decile 8			Item mean for decile 8 for all students
F8.3	AD09	Conditional Item Mean for Decile 9			Item mean for decile 9 for all students
F8.3	AD10	Conditional Item Mean for Decile 10			Item mean for decile 10 for all students
F8.3	MD01	Conditional Item Mean for Decile 1	Conditional item means plot: Males		Item mean for decile 1 for male students
F8.3	MD02	Conditional Item Mean for Decile 2			Item mean for decile 2 for male students
F8.3	MD03	Conditional Item Mean for Decile 3			Item mean for decile 3 for male students
F8.3	MD04	Conditional Item Mean for Decile 4			Item mean for decile 4 for male students
F8.3	MD05	Conditional Item Mean for Decile 5			Item mean for decile 5 for male students
F8.3	MD06	Conditional Item Mean for Decile 6			Item mean for decile 6 for male students
F8.3	MD07	Conditional Item Mean for Decile 7			Item mean for decile 7 for male students
F8.3	MD08	Conditional Item Mean for Decile 8			Item mean for decile 8 for male students
F8.3	MD09	Conditional Item Mean for Decile 9			Item mean for decile 9 for male students
F8.3	MD10	Conditional Item Mean for Decile 10			Item mean for decile 10 for male students

F8.3	FD01	Conditional Item Mean for Decile 1	Conditional item means plot: Females		Item mean for decile 1 for female students
F8.3	FD02	Conditional Item Mean for Decile 2			Item mean for decile 2 for female students
F8.3	FD03	Conditional Item Mean for Decile 3			Item mean for decile 3 for female students
F8.3	FD04	Conditional Item Mean for Decile 4			Item mean for decile 4 for female students
F8.3	FD05	Conditional Item Mean for Decile 5			Item mean for decile 5 for female students
F8.3	FD06	Conditional Item Mean for Decile 6			Item mean for decile 6 for female students
F8.3	FD07	Conditional Item Mean for Decile 7			Item mean for decile 7 for female students
F8.3	FD08	Conditional Item Mean for Decile 8			Item mean for decile 8 for female students
F8.3	FD09	Conditional Item Mean for Decile 9			Item mean for decile 9 for female students
F8.3	FD10	Conditional Item Mean for Decile 10			Item mean for decile 10 for female students
F8.3	WD01	Conditional Item Mean for Decile 1	Conditional item means plot: Whites		Item mean for decile 1 for white students
F8.3	WD02	Conditional Item Mean for Decile 2			Item mean for decile 2 for white students
F8.3	WD03	Conditional Item Mean for Decile 3			Item mean for decile 3 for white students
F8.3	WD04	Conditional Item Mean for Decile 4			Item mean for decile 4 for white students
F8.3	WD05	Conditional Item Mean for Decile 5			Item mean for decile 5 for white students
F8.3	WD06	Conditional Item Mean for Decile 6			Item mean for decile 6 for white students
F8.3	WD07	Conditional Item Mean for Decile 7			Item mean for decile 7 for white students
F8.3	WD08	Conditional Item Mean for Decile 8			Item mean for decile 8 for white students
F8.3	WD09	Conditional Item Mean for Decile 9			Item mean for decile 9 for white students
F8.3	WD10	Conditional Item Mean for Decile 10			Item mean for decile 10 for white students
F8.3	BD01	Conditional Item Mean for Decile 1	Conditional item means plot: Blacks		Item mean for decile 1 for black students
F8.3	BD02	Conditional Item Mean for Decile 2			Item mean for decile 2 for black students
F8.3	BD03	Conditional Item Mean for Decile 3			Item mean for decile 3 for black students
F8.3	BD04	Conditional Item Mean for Decile 4			Item mean for decile 4 for black students
F8.3	BD05	Conditional Item Mean for Decile 5			Item mean for decile 5 for black students

F8.3	BD06	Conditional Item Mean for Decile 6			Item mean for decile 6 for black students
F8.3	BD07	Conditional Item Mean for Decile 7			Item mean for decile 7 for black students
F8.3	BD08	Conditional Item Mean for Decile 8			Item mean for decile 8 for black students
F8.3	BD09	Conditional Item Mean for Decile 9			Item mean for decile 9 for black students
F8.3	BD10	Conditional Item Mean for Decile 10			Item mean for decile 10 for black students
F8.3	A95_A0	95th percentile	Box & whisker plot: All	Option A /Score 0	95th percentile of theta for all students for Option A or Score 0
F8.3	A75_A0	75th percentile			75th percentile of theta for all students for Option A or Score 0
F8.3	A50_A0	50th percentile			50th percentile of theta for all students for Option A or Score 0
F8.3	A25_A0	25th percentile			25th percentile of theta for all students for Option A or Score 0
F8.3	A05_A0	5th percentile			5th percentile of theta for all students for Option A or Score 0
F8.3	M95_A0	95th percentile	Box & whisker plot: Males		95th percentile of theta for male students for Option A or Score 0
F8.3	M75_A0	75th percentile			75th percentile of theta for male students for Option A or Score 0
F8.3	M50_A0	50th percentile			50th percentile of theta for male students for Option A or Score 0
F8.3	M25_A0	25th percentile			25th percentile of theta for male students for Option A or Score 0
F8.3	M05_A0	5th percentile			5th percentile of theta for male students for Option A or Score 0
F8.3	F95_A0	95th percentile	Box & whisker plot: Females		95th percentile of theta for female students for Option A or Score 0
F8.3	F75_A0	75th percentile			75th percentile of theta for female students for Option A or Score 0
F8.3	F50_A0	50th percentile			50th percentile of theta for female students for Option A or Score 0
F8.3	F25_A0	25th percentile			25th percentile of theta for female students for Option A or Score 0
F8.3	F05_A0	5th percentile			5th percentile of theta for female students for Option A or Score 0
F8.3	W95_A0	95th percentile	Box & whisker plot: Whites		95th percentile of theta for white students for Option A or Score 0

F8.3	W75_A0	75th percentile			75th percentile of theta for white students for Option A or Score 0
F8.3	W50_A0	50th percentile			50th percentile of theta for white students for Option A or Score 0
F8.3	W25_A0	25th percentile			25th percentile of theta for white students for Option A or Score 0
F8.3	W05_A0	5th percentile			5th percentile of theta for white students for Option A or Score 0
F8.3	B95_A0	95th percentile	Box & whisker plot: Blacks		95th percentile of theta for black students for Option A or Score 0
F8.3	B75_A0	75th percentile			75th percentile of theta for black students for Option A or Score 0
F8.3	B50_A0	50th percentile			50th percentile of theta for black students for Option A or Score 0
F8.3	B25_A0	25th percentile			25th percentile of theta for black students for Option A or Score 0
F8.3	B05_A0	5th percentile			5th percentile of theta for black students for Option A or Score 0
F8.3	A95_B1	95th percentile	Box & whisker plot: All	Option B /Score 1	95th percentile of theta for all students for Option B or Score 1
F8.3	A75_B1	75th percentile			75th percentile of theta for all students for Option B or Score 1
F8.3	A50_B1	50th percentile			50th percentile of theta for all students for Option B or Score 1
F8.3	A25_B1	25th percentile			25th percentile of theta for all students for Option B or Score 1
F8.3	A05_B1	5th percentile			5th percentile of theta for all students for Option B or Score 1
F8.3	M95_B1	95th percentile	Box & whisker plot: Males		95th percentile of theta for male students for Option B or Score 1
F8.3	M75_B1	75th percentile			75th percentile of theta for male students for Option B or Score 1
F8.3	M50_B1	50th percentile			50th percentile of theta for male students for Option B or Score 1
F8.3	M25_B1	25th percentile			25th percentile of theta for male students for Option B or Score 1
F8.3	M05_B1	5th percentile			5th percentile of theta for male students for Option B or Score 1
F8.3	F95_B1	95th percentile	Box & whisker plot: Females		95th percentile of theta for female students for Option B or Score 1

F8.3	F75_B1	75th percentile			75th percentile of theta for female students for Option B or Score 1
F8.3	F50_B1	50th percentile			50th percentile of theta for female students for Option B or Score 1
F8.3	F25_B1	25th percentile			25th percentile of theta for female students for Option B or Score 1
F8.3	F05_B1	5th percentile			5th percentile of theta for female students for Option B or Score 1
F8.3	W95_B1	95th percentile	Box & whisker plot: Whites		95th percentile of theta for white students for Option B or Score 1
F8.3	W75_B1	75th percentile			75th percentile of theta for white students for Option B or Score 1
F8.3	W50_B1	50th percentile			50th percentile of theta for white students for Option B or Score 1
F8.3	W25_B1	25th percentile			25th percentile of theta for white students for Option B or Score 1
F8.3	W05_B1	5th percentile			5th percentile of theta for white students for Option B or Score 1
F8.3	B95_B1	95th percentile	Box & whisker plot: Blacks		95th percentile of theta for black students for Option B or Score 1
F8.3	B75_B1	75th percentile			75th percentile of theta for black students for Option B or Score 1
F8.3	B50_B1	50th percentile			50th percentile of theta for black students for Option B or Score 1
F8.3	B25_B1	25th percentile			25th percentile of theta for black students for Option B or Score 1
F8.3	B05_B1	5th percentile			5th percentile of theta for black students for Option B or Score 1
F8.3	A95_C2	95th percentile	Box & whisker plot: All	Option C /Score 2	95th percentile of theta for all students for Option C or Score 2
F8.3	A75_C2	75th percentile			75th percentile of theta for all students for Option C or Score 2
F8.3	A50_C2	50th percentile			50th percentile of theta for all students for Option C or Score 2
F8.3	A25_C2	25th percentile			25th percentile of theta for all students for Option C or Score 2
F8.3	A05_C2	5th percentile			5th percentile of theta for all students for Option C or Score 2
F8.3	M95_C2	95th percentile	Box & whisker plot: Males		95th percentile of theta for male students for Option C or Score 2

F8.3	M75_C2	75th percentile			75th percentile of theta for male students for Option C or Score 2
F8.3	M50_C2	50th percentile			50th percentile of theta for male students for Option C or Score 2
F8.3	M25_C2	25th percentile			25th percentile of theta for male students for Option C or Score 2
F8.3	M05_C2	5th percentile			5th percentile of theta for male students for Option C or Score 2
F8.3	F95_C2	95th percentile	Box & whisker plot: Females		95th percentile of theta for female students for Option C or Score 2
F8.3	F75_C2	75th percentile			75th percentile of theta for female students for Option C or Score 2
F8.3	F50_C2	50th percentile			50th percentile of theta for female students for Option C or Score 2
F8.3	F25_C2	25th percentile			25th percentile of theta for female students for Option C or Score 2
F8.3	F05_C2	5th percentile			5th percentile of theta for female students for Option C or Score 2
F8.3	W95_C2	95th percentile	Box & whisker plot: Whites		95th percentile of theta for white students for Option C or Score 2
F8.3	W75_C2	75th percentile			75th percentile of theta for white students for Option C or Score 2
F8.3	W50_C2	50th percentile			50th percentile of theta for white students for Option C or Score 2
F8.3	W25_C2	25th percentile			25th percentile of theta for white students for Option C or Score 2
F8.3	W05_C2	5th percentile			5th percentile of theta for white students for Option C or Score 2
F8.3	B95_C2	95th percentile	Box & whisker plot: Blacks		95th percentile of theta for black students for Option C or Score 2
F8.3	B75_C2	75th percentile			75th percentile of theta for black students for Option C or Score 2
F8.3	B50_C2	50th percentile			50th percentile of theta for black students for Option C or Score 2
F8.3	B25_C2	25th percentile			25th percentile of theta for black students for Option C or Score 2
F8.3	B05_C2	5th percentile			5th percentile of theta for black students for Option C or Score 2
F8.3	A95_D3	95th percentile	Box & whisker plot: All	Option D	95th percentile of theta for all students for Option D or Score 3

F8.3	A75_D3	75th percentile		/Score 3	75th percentile of theta for all students for Option D or Score 3
F8.3	A50_D3	50th percentile			50th percentile of theta for all students for Option D or Score 3
F8.3	A25_D3	25th percentile			25th percentile of theta for all students for Option D or Score 3
F8.3	A05_D3	5th percentile			5th percentile of theta for all students for Option D or Score 3
F8.3	M95_D3	95th percentile	Box & whisker plot: Males		95th percentile of theta for male students for Option D or Score 3
F8.3	M75_D3	75th percentile			75th percentile of theta for male students for Option D or Score 3
F8.3	M50_D3	50th percentile			50th percentile of theta for male students for Option D or Score 3
F8.3	M25_D3	25th percentile			25th percentile of theta for male students for Option D or Score 3
F8.3	M05_D3	5th percentile			5th percentile of theta for male students for Option D or Score 3
F8.3	F95_D3	95th percentile	Box & whisker plot: Females		95th percentile of theta for female students for Option D or Score 3
F8.3	F75_D3	75th percentile			75th percentile of theta for female students for Option D or Score 3
F8.3	F50_D3	50th percentile			50th percentile of theta for female students for Option D or Score 3
F8.3	F25_D3	25th percentile			25th percentile of theta for female students for Option D or Score 3
F8.3	F05_D3	5th percentile			5th percentile of theta for female students for Option D or Score 3
F8.3	W95_D3	95th percentile	Box & whisker plot: Whites		95th percentile of theta for white students for Option D or Score 3
F8.3	W75_D3	75th percentile			75th percentile of theta for white students for Option D or Score 3
F8.3	W50_D3	50th percentile			50th percentile of theta for white students for Option D or Score 3
F8.3	W25_D3	25th percentile			25th percentile of theta for white students for Option D or Score 3
F8.3	W05_D3	5th percentile			5th percentile of theta for white students for Option D or Score 3
F8.3	B95_D3	95th percentile	Box & whisker plot: Blacks		95th percentile of theta for black students for Option D or Score 3

F8.3	B75_D3	75th percentile			75th percentile of theta for black students for Option D or Score 3
F8.3	B50_D3	50th percentile			50th percentile of theta for black students for Option D or Score 3
F8.3	B25_D3	25th percentile			25th percentile of theta for black students for Option D or Score 3
F8.3	B05_D3	5th percentile			5th percentile of theta for black students for Option D or Score 3
F8.3	A95_4	95th percentile	Box & whisker plot: All	Score 4	95th percentile of theta for all students for Score 4
F8.3	A75_4	75th percentile			75th percentile of theta for all students for Score 4
F8.3	A50_4	50th percentile			50th percentile of theta for all students for Score 4
F8.3	A25_4	25th percentile			25th percentile of theta for all students for Score 4
F8.3	A05_4	5th percentile			5th percentile of theta for all students for Score 4
F8.3	M95_4	95th percentile	Box & whisker plot: Males		95th percentile of theta for male students for Score 4
F8.3	M75_4	75th percentile			75th percentile of theta for male students for Score 4
F8.3	M50_4	50th percentile			50th percentile of theta for male students for Score 4
F8.3	M25_4	25th percentile			25th percentile of theta for male students for Score 4
F8.3	M05_4	5th percentile			5th percentile of theta for male students for Score 4
F8.3	F95_4	95th percentile	Box & whisker plot: Females		95th percentile of theta for female students for Score 4
F8.3	F75_4	75th percentile			75th percentile of theta for female students for Score 4
F8.3	F50_4	50th percentile			50th percentile of theta for female students for Score 4
F8.3	F25_4	25th percentile			25th percentile of theta for female students for Score 4
F8.3	F05_4	5th percentile			5th percentile of theta for female students for Score 4
F8.3	W95_4	95th percentile	Box & whisker plot: Whites		95th percentile of theta for white students for Score 4
F8.3	W75_4	75th percentile			75th percentile of theta for white students for Score 4
F8.3	W50_4	50th percentile			50th percentile of theta for white students for Score 4
F8.3	W25_4	25th percentile			25th percentile of theta for white students for Score 4
F8.3	W05_4	5th percentile			5th percentile of theta for white students for Score 4
F8.3	B95_4	95th percentile	Box & whisker plot: Blacks		95th percentile of theta for black students for Score 4
F8.3	B75_4	75th percentile			75th percentile of theta for black students for Score

					4
F8.3	B50_4	50th percentile			50th percentile of theta for black students for Score 4
F8.3	B25_4	25th percentile			25th percentile of theta for black students for Score 4
F8.3	B05_4	5th percentile			5th percentile of theta for black students for Score 4
F8.3	A95_5	95th percentile	Box & whisker plot: All	Score 5	95th percentile of theta for all students for Score 5
F8.3	A75_5	75th percentile			75th percentile of theta for all students for Score 5
F8.3	A50_5	50th percentile			50th percentile of theta for all students for Score 5
F8.3	A25_5	25th percentile			25th percentile of theta for all students for Score 5
F8.3	A05_5	5th percentile			5th percentile of theta for all students for Score 5
F8.3	M95_5	95th percentile	Box & whisker plot: Males		95th percentile of theta for male students for Score 5
F8.3	M75_5	75th percentile			75th percentile of theta for male students for Score 5
F8.3	M50_5	50th percentile			50th percentile of theta for male students for Score 5
F8.3	M25_5	25th percentile			25th percentile of theta for male students for Score 5
F8.3	M05_5	5th percentile			5th percentile of theta for male students for Score 5
F8.3	F95_5	95th percentile	Box & whisker plot: Females		95th percentile of theta for female students for Score 5
F8.3	F75_5	75th percentile			75th percentile of theta for female students for Score 5
F8.3	F50_5	50th percentile			50th percentile of theta for female students for Score 5
F8.3	F25_5	25th percentile			25th percentile of theta for female students for Score 5
F8.3	F05_5	5th percentile			5th percentile of theta for female students for Score 5
F8.3	W95_5	95th percentile	Box & whisker plot: Whites		95th percentile of theta for white students for Score 5
F8.3	W75_5	75th percentile			75th percentile of theta for white students for Score 5
F8.3	W50_5	50th percentile			50th percentile of theta for white students for Score 5
F8.3	W25_5	25th percentile			25th percentile of theta for white students for Score 5
F8.3	W05_5	5th percentile			5th percentile of theta for white students for Score 5
F8.3	B95_5	95th percentile	Box & whisker plot: Blacks		95th percentile of theta for black students for Score 5
F8.3	B75_5	75th percentile			75th percentile of theta for black students for Score 5
F8.3	B50_5	50th percentile			50th percentile of theta for black students for Score 5

F8.3	B25_5	25th percentile			25th percentile of theta for black students for Score 5
F8.3	B05_5	5th percentile			5th percentile of theta for black students for Score 5
F8.3	A95_6	95th percentile	Box & whisker plot: All	Score 6	95th percentile of theta for all students for Score 6
F8.3	A75_6	75th percentile			75th percentile of theta for all students for Score 6
F8.3	A50_6	50th percentile			50th percentile of theta for all students for Score 6
F8.3	A25_6	25th percentile			25th percentile of theta for all students for Score 6
F8.3	A05_6	5th percentile			5th percentile of theta for all students for Score 6
F8.3	M95_6	95th percentile	Box & whisker plot: Males		95th percentile of theta for male students for Score 6
F8.3	M75_6	75th percentile			75th percentile of theta for male students for Score 6
F8.3	M50_6	50th percentile			50th percentile of theta for male students for Score 6
F8.3	M25_6	25th percentile			25th percentile of theta for male students for Score 6
F8.3	M05_6	5th percentile			5th percentile of theta for male students for Score 6
F8.3	F95_6	95th percentile	Box & whisker plot: Females		95th percentile of theta for female students for Score 6
F8.3	F75_6	75th percentile			75th percentile of theta for female students for Score 6
F8.3	F50_6	50th percentile			50th percentile of theta for female students for Score 6
F8.3	F25_6	25th percentile			25th percentile of theta for female students for Score 6
F8.3	F05_6	5th percentile			5th percentile of theta for female students for Score 6
F8.3	W95_6	95th percentile	Box & whisker plot: Whites		95th percentile of theta for white students for Score 6
F8.3	W75_6	75th percentile			75th percentile of theta for white students for Score 6
F8.3	W50_6	50th percentile			50th percentile of theta for white students for Score 6
F8.3	W25_6	25th percentile			25th percentile of theta for white students for Score 6
F8.3	W05_6	5th percentile			5th percentile of theta for white students for Score 6
F8.3	B95_6	95th percentile	Box & whisker plot: Blacks		95th percentile of theta for black students for Score 6
F8.3	B75_6	75th percentile			75th percentile of theta for black students for Score 6
F8.3	B50_6	50th percentile			50th percentile of theta for black students for Score 6
F8.3	B25_6	25th percentile			25th percentile of theta for black students for Score 6
F8.3	B05_6	5th percentile			5th percentile of theta for black students for Score 6

F8.3	A95_7	95th percentile	Box & whisker plot: All	Score 7	95th percentile of theta for all students for Score 7
F8.3	A75_7	75th percentile			75th percentile of theta for all students for Score 7
F8.3	A50_7	50th percentile			50th percentile of theta for all students for Score 7
F8.3	A25_7	25th percentile			25th percentile of theta for all students for Score 7
F8.3	A05_7	5th percentile			5th percentile of theta for all students for Score 7
F8.3	M95_7	95th percentile	Box & whisker plot: Males		95th percentile of theta for male students for Score 7
F8.3	M75_7	75th percentile			75th percentile of theta for male students for Score 7
F8.3	M50_7	50th percentile			50th percentile of theta for male students for Score 7
F8.3	M25_7	25th percentile			25th percentile of theta for male students for Score 7
F8.3	M05_7	5th percentile			5th percentile of theta for male students for Score 7
F8.3	F95_7	95th percentile	Box & whisker plot: Females		95th percentile of theta for female students for Score 7
F8.3	F75_7	75th percentile			75th percentile of theta for female students for Score 7
F8.3	F50_7	50th percentile			50th percentile of theta for female students for Score 7
F8.3	F25_7	25th percentile			25th percentile of theta for female students for Score 7
F8.3	F05_7	5th percentile			5th percentile of theta for female students for Score 7
F8.3	W95_7	95th percentile	Box & whisker plot: Whites		95th percentile of theta for white students for Score 7
F8.3	W75_7	75th percentile			75th percentile of theta for white students for Score 7
F8.3	W50_7	50th percentile			50th percentile of theta for white students for Score 7
F8.3	W25_7	25th percentile			25th percentile of theta for white students for Score 7
F8.3	W05_7	5th percentile			5th percentile of theta for white students for Score 7
F8.3	B95_7	95th percentile	Box & whisker plot: Blacks		95th percentile of theta for black students for Score 7
F8.3	B75_7	75th percentile			75th percentile of theta for black students for Score 7
F8.3	B50_7	50th percentile			50th percentile of theta for black students for Score 7
F8.3	B25_7	25th percentile			25th percentile of theta for black students for Score 7
F8.3	B05_7	5th percentile			5th percentile of theta for black students for Score 7
F8.3	A95_8	95th percentile	Box & whisker plot: All	Score 8	95th percentile of theta for all students for Score 8
F8.3	A75_8	75th percentile			75th percentile of theta for all students for Score 8
F8.3	A50_8	50th percentile			50th percentile of theta for all students for Score 8

F8.3	A25_8	25th percentile			25th percentile of theta for all students for Score 8
F8.3	A05_8	5th percentile			5th percentile of theta for all students for Score 8
F8.3	M95_8	95th percentile	Box & whisker plot: Males		95th percentile of theta for male students for Score 8
F8.3	M75_8	75th percentile			75th percentile of theta for male students for Score 8
F8.3	M50_8	50th percentile			50th percentile of theta for male students for Score 8
F8.3	M25_8	25th percentile			25th percentile of theta for male students for Score 8
F8.3	M05_8	5th percentile			5th percentile of theta for male students for Score 8
F8.3	F95_8	95th percentile	Box & whisker plot: Females		95th percentile of theta for female students for Score 8
F8.3	F75_8	75th percentile			75th percentile of theta for female students for Score 8
F8.3	F50_8	50th percentile			50th percentile of theta for female students for Score 8
F8.3	F25_8	25th percentile			25th percentile of theta for female students for Score 8
F8.3	F05_8	5th percentile			5th percentile of theta for female students for Score 8
F8.3	W95_8	95th percentile	Box & whisker plot: Whites		95th percentile of theta for white students for Score 8
F8.3	W75_8	75th percentile			75th percentile of theta for white students for Score 8
F8.3	W50_8	50th percentile			50th percentile of theta for white students for Score 8
F8.3	W25_8	25th percentile			25th percentile of theta for white students for Score 8
F8.3	W05_8	5th percentile			5th percentile of theta for white students for Score 8
F8.3	B95_8	95th percentile	Box & whisker plot: Blacks		95th percentile of theta for black students for Score 8
F8.3	B75_8	75th percentile			75th percentile of theta for black students for Score 8
F8.3	B50_8	50th percentile			50th percentile of theta for black students for Score 8
F8.3	B25_8	25th percentile			25th percentile of theta for black students for Score 8
F8.3	B05_8	5th percentile			5th percentile of theta for black students for Score 8
F8.3	A95_9	95th percentile	Box & whisker plot: All	Score 9	95th percentile of theta for all students for Score 9
F8.3	A75_9	75th percentile			75th percentile of theta for all students for Score 9
F8.3	A50_9	50th percentile			50th percentile of theta for all students for Score 9
F8.3	A25_9	25th percentile			25th percentile of theta for all students for Score 9
F8.3	A05_9	5th percentile			5th percentile of theta for all students for Score 9
F8.3	M95_9	95th percentile	Box & whisker plot: Males		95th percentile of theta for male students for Score

					9
F8.3	M75_9	75th percentile			75th percentile of theta for male students for Score 9
F8.3	M50_9	50th percentile			50th percentile of theta for male students for Score 9
F8.3	M25_9	25th percentile			25th percentile of theta for male students for Score 9
F8.3	M05_9	5th percentile			5th percentile of theta for male students for Score 9
F8.3	F95_9	95th percentile	Box & whisker plot: Females		95th percentile of theta for female students for Score 9
F8.3	F75_9	75th percentile			75th percentile of theta for female students for Score 9
F8.3	F50_9	50th percentile			50th percentile of theta for female students for Score 9
F8.3	F25_9	25th percentile			25th percentile of theta for female students for Score 9
F8.3	F05_9	5th percentile			5th percentile of theta for female students for Score 9
F8.3	W95_9	95th percentile	Box & whisker plot: Whites		95th percentile of theta for white students for Score 9
F8.3	W75_9	75th percentile			75th percentile of theta for white students for Score 9
F8.3	W50_9	50th percentile			50th percentile of theta for white students for Score 9
F8.3	W25_9	25th percentile			25th percentile of theta for white students for Score 9
F8.3	W05_9	5th percentile			5th percentile of theta for white students for Score 9
F8.3	B95_9	95th percentile	Box & whisker plot: Blacks		95th percentile of theta for black students for Score 9
F8.3	B75_9	75th percentile			75th percentile of theta for black students for Score 9
F8.3	B50_9	50th percentile			50th percentile of theta for black students for Score 9
F8.3	B25_9	25th percentile			25th percentile of theta for black students for Score 9
F8.3	B05_9	5th percentile			5th percentile of theta for black students for Score 9
F8.3	A95_10	95th percentile	Box & whisker plot: All	Score 10	95th percentile of theta for all students for Score 10
F8.3	A75_10	75th percentile			75th percentile of theta for all students for Score 10
F8.3	A50_10	50th percentile			50th percentile of theta for all students for Score 10
F8.3	A25_10	25th percentile			25th percentile of theta for all students for Score 10
F8.3	A05_10	5th percentile			5th percentile of theta for all students for Score 10
F8.3	M95_10	95th percentile	Box & whisker plot: Males		95th percentile of theta for male students for Score 10
F8.3	M75_10	75th percentile			75th percentile of theta for male students for Score 10

F8.3	M50_10	50th percentile			50th percentile of theta for male students for Score 10
F8.3	M25_10	25th percentile			25th percentile of theta for male students for Score 10
F8.3	M05_10	5th percentile			5th percentile of theta for male students for Score 10
F8.3	F95_10	95th percentile	Box & whisker plot: Females		95th percentile of theta for female students for Score 10
F8.3	F75_10	75th percentile			75th percentile of theta for female students for Score 10
F8.3	F50_10	50th percentile			50th percentile of theta for female students for Score 10
F8.3	F25_10	25th percentile			25th percentile of theta for female students for Score 10
F8.3	F05_10	5th percentile			5th percentile of theta for female students for Score 10
F8.3	W95_10	95th percentile	Box & whisker plot: Whites		95th percentile of theta for white students for Score 10
F8.3	W75_10	75th percentile			75th percentile of theta for white students for Score 10
F8.3	W50_10	50th percentile			50th percentile of theta for white students for Score 10
F8.3	W25_10	25th percentile			25th percentile of theta for white students for Score 10
F8.3	W05_10	5th percentile			5th percentile of theta for white students for Score 10
F8.3	B95_10	95th percentile	Box & whisker plot: Blacks		95th percentile of theta for black students for Score 10
F8.3	B75_10	75th percentile			75th percentile of theta for black students for Score 10
F8.3	B50_10	50th percentile			50th percentile of theta for black students for Score 10
F8.3	B25_10	25th percentile			25th percentile of theta for black students for Score 10
F8.3	B05_10	5th percentile			5th percentile of theta for black students for Score 10
F8.3	A95_11	95th percentile	Box & whisker plot: All	Score 11	95th percentile of theta for all students for Score 11
F8.3	A75_11	75th percentile			75th percentile of theta for all students for Score 11
F8.3	A50_11	50th percentile			50th percentile of theta for all students for Score 11
F8.3	A25_11	25th percentile			25th percentile of theta for all students for Score 11
F8.3	A05_11	5th percentile			5th percentile of theta for all students for Score 11
F8.3	M95_11	95th percentile	Box & whisker plot: Males		95th percentile of theta for male students for Score 11
F8.3	M75_11	75th percentile			75th percentile of theta for male students for Score 11
F8.3	M50_11	50th percentile			50th percentile of theta for male students for Score 11

F8.3	M25_11	25th percentile			25th percentile of theta for male students for Score 11
F8.3	M05_11	5th percentile			5th percentile of theta for male students for Score 11
F8.3	F95_11	95th percentile	Box & whisker plot: Females		95th percentile of theta for female students for Score 11
F8.3	F75_11	75th percentile			75th percentile of theta for female students for Score 11
F8.3	F50_11	50th percentile			50th percentile of theta for female students for Score 11
F8.3	F25_11	25th percentile			25th percentile of theta for female students for Score 11
F8.3	F05_11	5th percentile			5th percentile of theta for female students for Score 11
F8.3	W95_11	95th percentile	Box & whisker plot: Whites		95th percentile of theta for white students for Score 11
F8.3	W75_11	75th percentile			75th percentile of theta for white students for Score 11
F8.3	W50_11	50th percentile			50th percentile of theta for white students for Score 11
F8.3	W25_11	25th percentile			25th percentile of theta for white students for Score 11
F8.3	W05_11	5th percentile			5th percentile of theta for white students for Score 11
F8.3	B95_11	95th percentile	Box & whisker plot: Blacks		95th percentile of theta for black students for Score 11
F8.3	B75_11	75th percentile			75th percentile of theta for black students for Score 11
F8.3	B50_11	50th percentile			50th percentile of theta for black students for Score 11
F8.3	B25_11	25th percentile			25th percentile of theta for black students for Score 11
F8.3	B05_11	5th percentile			5th percentile of theta for black students for Score 11
F8.3	A95_12	95th percentile	Box & whisker plot: All	Score 12	95th percentile of theta for all students for Score 12
F8.3	A75_12	75th percentile			75th percentile of theta for all students for Score 12
F8.3	A50_12	50th percentile			50th percentile of theta for all students for Score 12
F8.3	A25_12	25th percentile			25th percentile of theta for all students for Score 12
F8.3	A05_12	5th percentile			5th percentile of theta for all students for Score 12
F8.3	M95_12	95th percentile	Box & whisker plot: Males		95th percentile of theta for male students for Score 12
F8.3	M75_12	75th percentile			75th percentile of theta for male students for Score 12
F8.3	M50_12	50th percentile			50th percentile of theta for male students for Score 12
F8.3	M25_12	25th percentile			25th percentile of theta for male students for Score 12

F8.3	M05_12	5th percentile			5th percentile of theta for male students for Score 12
F8.3	F95_12	95th percentile	Box & whisker plot: Females		95th percentile of theta for female students for Score 12
F8.3	F75_12	75th percentile			75th percentile of theta for female students for Score 12
F8.3	F50_12	50th percentile			50th percentile of theta for female students for Score 12
F8.3	F25_12	25th percentile			25th percentile of theta for female students for Score 12
F8.3	F05_12	5th percentile			5th percentile of theta for female students for Score 12
F8.3	W95_12	95th percentile	Box & whisker plot: Whites		95th percentile of theta for white students for Score 12
F8.3	W75_12	75th percentile			75th percentile of theta for white students for Score 12
F8.3	W50_12	50th percentile			50th percentile of theta for white students for Score 12
F8.3	W25_12	25th percentile			25th percentile of theta for white students for Score 12
F8.3	W05_12	5th percentile			5th percentile of theta for white students for Score 12
F8.3	B95_12	95th percentile	Box & whisker plot: Blacks		95th percentile of theta for black students for Score 12
F8.3	B75_12	75th percentile			75th percentile of theta for black students for Score 12
F8.3	B50_12	50th percentile			50th percentile of theta for black students for Score 12
F8.3	B25_12	25th percentile			25th percentile of theta for black students for Score 12
F8.3	B05_12	5th percentile			5th percentile of theta for black students for Score 12
F8.3	A95_OM	95th percentile	Box & whisker plot: All	Omits	95th percentile of theta for all students for omits
F8.3	A75_OM	75th percentile			75th percentile of theta for all students for omits
F8.3	A50_OM	50th percentile			50th percentile of theta for all students for omits
F8.3	A25_OM	25th percentile			25th percentile of theta for all students for omits
F8.3	A05_OM	5th percentile			5th percentile of theta for all students for omits
F8.3	M95_OM	95th percentile	Box & whisker plot: Males		95th percentile of theta for male students for omits
F8.3	M75_OM	75th percentile			75th percentile of theta for male students for omits
F8.3	M50_OM	50th percentile			50th percentile of theta for male students for omits
F8.3	M25_OM	25th percentile			25th percentile of theta for male students for omits
F8.3	M05_OM	5th percentile			5th percentile of theta for male students for omits
F8.3	F95_OM	95th percentile	Box & whisker plot: Females		95th percentile of theta for female students for omits
F8.3	F75_OM	75th percentile			75th percentile of theta for female students for omits

F8.3	F50_OM	50th percentile			50th percentile of theta for female students for omits
F8.3	F25_OM	25th percentile			25th percentile of theta for female students for omits
F8.3	F05_OM	5th percentile			5th percentile of theta for female students for omits
F8.3	W95_OM	95th percentile	Box & whisker plot: Whites		95th percentile of theta for white students for omits
F8.3	W75_OM	75th percentile			75th percentile of theta for white students for omits
F8.3	W50_OM	50th percentile			50th percentile of theta for white students for omits
F8.3	W25_OM	25th percentile			25th percentile of theta for white students for omits
F8.3	W05_OM	5th percentile			5th percentile of theta for white students for omits
F8.3	B95_OM	95th percentile	Box & whisker plot: Blacks		95th percentile of theta for black students for omits
F8.3	B75_OM	75th percentile			75th percentile of theta for black students for omits
F8.3	B50_OM	50th percentile			50th percentile of theta for black students for omits
F8.3	B25_OM	25th percentile			25th percentile of theta for black students for omits
F8.3	B05_OM	5th percentile			5th percentile of theta for black students for omits
F8.5	PARTIALLY PROFICIENT	Reserved 1 (Theta cut for Basic)	Reserved for future use (20 numeric and 5 alphanumeric)		
F8.5	MET	Reserved 2 (Theta cut for Met)			
F8.5	EXCEED	Reserved 3 (Theta cut for Exceed)			
F8.5	ICC1	Reserved 4 (ICC at cut for Basic)			
F8.5	ICC2	Reserved 5 (ICC at cut for Met)			
F8.5	ICC3	Reserved 6 (ICC at cut for Exceed)			
F8.3	RES7	Reserved 7			
F8.3	RES8	Reserved 8			
F8.3	RES9	Reserved 9			
F8.3	RES10	Reserved 10			
F8.3	RES11	Reserved 11			
F8.3	RES12	Reserved 12			
F8.3	RES13	Reserved 13			
F8.3	RES14	Reserved 14			
F8.3	RES15	Reserved 15			
F8.3	RES16	Reserved 16			
F8.3	RES17	Reserved 17			
F8.3	RES18	Reserved 18			
F8.3	RES19	Reserved 19			
F8.3	RES20	Reserved 20			

A5	RES21	Reserved 21			
A5	RES22	Reserved 22			
A5	RES23	Reserved 23			
A5	RES24	Reserved 24			
A5	RES25	Reserved 25			
A2	sx2fitflag	Fit Flag based on sx2 statistic	Replaces ZQ1 fit flag.		Equals NF (no fit) if p-value < .05, otherwise blank.
A2	sx2fitflag2	Fit Flag based on sx2 statistic for rater 2 if operational CR item	Replaces ZQ1 fit flag.		Equals NF (no fit) if p-value < .05, otherwise blank.
F8.3	sx2	IRT fit statistic for PARSCALE calibrated items.	Replaces ZQ1 fit statistc.		
F3	df_sx2	degress of freedom for sx2 statistic.			
F8.3	p_sx2	p-value for sx2 statistic			
F8.3	sx2r2	IRT fit statistic for PARSCALE calibrated CR item with second rater.	Replaces ZQ1 fit statistc.		
F3	df_sx2r2	degress of freedom for sx2r2 statistic.			
F8.3	p_sx2r2	p-value for sx2r2 statistic.			

Field Format	Field Name	Field Description	Notes	Notes 2
A6	TEST	Test Name	MATH03, READ05, SCIE08, SOCS06, WRIT05, etc.	
A2	SUBJ	Subject (RE, MA, SC, SS, WR)	REading, MAthematics, SCience, Social Studies, WRiting	
A2	GRADE	Grade	Grade in which an item administered	
A25	MME_ID	MME Item ID	Michigan item identifyer	
F12	CID	CID (currently 7 digits used)	Company ID number for an item (HAI or PEM)	
A2	TYPE	Item Type (MC, CR)	MC - multiple-choice, CR - constructed response	
A1	KEY	Item Answer Key (A, B, C, D)	For MC items	
F1	MAX	Item Maximal Score (3, 4, 6)	For CR items	
A3	STRAND	Item Strand		
A3	BNCHM	Item Benchmark		
A10	GLCE	Grade Level Expectation		
F1	DEPTKN	Depth of Knowledge		
A2	CYCLE	Year cycle (2 characters)		
A2	DOMAIN	Domain		
A2	LEVEL	Level		
A50	SCENARIO	Scenario		

A4	ADMYEAR	Administration Year	For each administration year a separate line will be provided	
A3	RELEASED	Released position or N/R		
A3	MATURITY	PP, PI, FT, OP, RL, EM	Pre-pilot, Pilot, Field-Test, Operational, Released, Emergency	
A3	FUNC	Item Function in Current Administration (CO, FC, EC, LI)	Core, Future core, Extended core, Linking	
A3	CHAR_COD	Character Code	See the spreadsheet 'Codes'	
F2	NFORMS	Number of Forms Item Appears On (1 - 5)	Indicates how many forms a matrix item appears on, ranges 1-5 (not supplied for core items).	
A60	FORMS	Form Numbers (string of 3x20 characters)	Indicates which forms a matrix item appears on, there will be as many form numbers as there are forms that item appears on (not supplied for core items).	
A60	POSITS	Test Positions (string of 3x20 characters)	Indicates positions in the test for each form that a matrix item appears on, there will be as many position numbers as there are forms that item appears on (shows only one number for core items).	
F6	NCOUNT	N-count	Number of calibration cases used to produce statistics	
F6	N_MAL	Ncount Males	N-counts for break-down groups	
F6	N_FEM	Ncount Females		
F6	N_WHI	Ncount White		
F6	N_BLA	Ncount Black		
F2	COM1	Percent for Comment Code 1		
F2	COM2	Percent for Comment Code 2		
F2	COM3	Percent for Comment Code 3		
F2	COM4	Percent for Comment Code 4		
F2	COM5	Percent for Comment Code 5		
F2	COM6	Percent for Comment Code 6		
F2	COM7	Percent for Comment Code 7		
F2	COM8	Percent for Comment Code 8		
F2	COM9	Percent for Comment Code 9		
F2	COM10	Percent for Comment Code 10 (not used yet)		
F2	COM11	Percent for Comment Code 11 (not used yet)		
F2	COM12	Percent for Comment Code 12 (not used yet)		
F2	COND_A	Percent for Condition Code A	Codition code distribution (for CR items only, see the spreadsheet "Codes")	
F2	COND_B	Percent for Condition Code B		
F2	COND_C	Percent for Condition Code C		

F2	COND_D	Percent for Condition Code D		
F2	COND_E	Percent for Condition Code E		
F2	COND_F	Percent for Condition Code F (not used yet)		
F2	COND_G	Percent for Condition Code G (not used yet)		
F2	COND_H	Percent for Condition Code H (not used yet)		
F2	A	Percent (option A or scorepoint 0)	Percent of ALL calibration cases	
F2	B	Percent (option B or scorepoint 1)		
F2	C	Percent (option C or scorepoint 2)		
F2	D	Percent (option D or scorepoint 3)		
F2	M	Percent (mult. marks or scorepoint 4)		
F2	S5	Percent (scorepoint 5)		
F2	S6	Percent (scorepoint 6)		
F2	O	Percent (Omits)		
F2	MAA	Male Percent (A or 0)	Percent for MALE calibration cases	
F2	MAB	Male Percent (B or 1)		
F2	MAC	Male Percent (C or 2)		
F2	MAD	Male Percent (D or 3)		
F2	MAM	Male Percent (MM or 4)		
F2	MAS5	Male Percent (scorepoint 5)		
F2	MAS6	Male Percent (scorepoint 6)		
F2	MAO	Male Percent (Omits)		
F2	FEA	Female Percent (A or 0)	Percent for FEMALE calibration cases	
F2	FEB	Female Percent (B or 1)		
F2	FEC	Female Percent (C or 2)		
F2	FED	Female Percent (D or 3)		
F2	FEM	Female Percent (MM or 4)		
F2	FES5	Female Percent (scorepoint 5)		
F2	FES6	Female Percent (scorepoint 6)		
F2	FEO	Female Percent (Omits)		
F2	WHA	White Percent (A or 0)	Percent for WHITE calibration cases	
F2	WHB	White Percent (B or 1)		
F2	WHC	White Percent (C or 2)		
F2	WHD	White Percent (D or 3)		
F2	WHM	White Percent (MM or 4)		
F2	WHS5	White Percent (scorepoint 5)		

F2	WHS6	White Percent (scorepoint 6)		
F2	WHO	White Percent (Omits)		
F2	BLA	Black Percent (A or 0)	Percent for BLACK calibration cases	
F2	BLB	Black Percent (B or 1)		
F2	BLC	Black Percent (C or 2)		
F2	BLD	Black Percent (D or 3)		
F2	BLM	Black Percent (MM or 4)		
F2	BLS5	Black Percent (scorepoint 5)		
F2	BLS6	Black Percent (scorepoint 6)		
F2	BLO	Black Percent (Omits)		
F8.4	PVAL	P-value or Item Mean	P-value or arithmetic mean of item scores (all cases)	
F8.4	MPVAL	P-value or Item Mean for Male	Impact analysis: item means for break-down groups	
F8.4	FPVAL	P-value or Item Mean for Female		
F8.4	WPVAL	P-value or Item Mean for White		
F8.4	BPVAL	P-value or Item Mean for Black		
F8.4	ADJPVAL	Adjusted P-value	Adjusted P-value = (Arithmetic mean - MIN item score) / (MAX item score - MIN item score)	
A5	DIFFICFL	Difficulty flag	Based on Test Construction Specifications	
F8.4	SDEV	Item Standard Deviation	Standard deviation of item scores	
F8.4	ITOT	Item-Total Correlation	Pearson product-moment correlation (Point-Biserial correlation for dihotomous items)	
F8.4	ITOTBIS	Biserial / Polyserial Correlation	For MC: biserial, for CR:poliserial (optional)	
F8.4	ITOTC	Point-Biserial Correlation (corrected)	For MC items (corrected for maximal possible value)	
A2	ITOTFL	Item-Total correlation flag	Based on Test Construction Specifications	
F8.4	APB	P-b correlation for option A	Options point-biserial correlations (for CR items only Omits Rpb is supplied)	
F8.4	BPB	P-b correlation for option B		
F8.4	CPB	P-b correlation for option C		
F8.4	DPB	P-b correlation for option D		
F8.4	OPB	P-b correlation for Omits		
A7	MISKFL	Flag for potential miskeying	Based on Test Construction Specifications	
F8.4	MCHI_MF	Mantel CHSQ Male-Female	DIF analyses: Mantel chi-square (for both dichotomous and polytomous items), Mantel-Haenszel Delta and corresponding standard error for dichotomous items (not supplied for polytomous items)	
F8.4	MHD_MF	Mantel-Haenszel Delta Male-Female		
F8.4	MHDSE_MF	Mantel-Haenszel Delta St. Error Male-Female		
F8.4	MCHI_WB	Mantel CHSQ White-Black		
F8.4	MHD_WB	Mantel-Haenszel Delta White-Black		

F8.4	MHDSE_WB	Mantel-Haenszel Delta St. Error White-Black		
F8.4	SMDS_MF	SMD signed M-F	DIF analyses: Standardized Mean Difference (signed: mean of algebraic differences; unsigned: mean of absolute differences); Effect size of signed SMD	
F8.4	SMDES_MF	SMD signed Effect Size for M-F		
F8.4	SMDU_MF	SMD unsigned M-F		
F8.4	SMDS_WB	SMD signed W-B		
F8.4	SMDES_WB	SMD signed Effect Size for W-B		
F8.4	SMDU_WB	SMD unsigned W-B		
A2	DIF_MF	DIF category for M-F (A, B, C)	DIF level categorization: A - no or negligible, B - moderate, C - substantial.	
A2	DIF_WB	DIF category for W-B (A, B, C)		
A6	FG_MF	Favored group for M-F (Male, Female)	Favored group if DIF level equal to B or C	
A6	FG_WB	Favored group for W-B (White, Black)		
F8.5	BPAR	B parameter (scaled)	Scaled (equated) IRT parameters: for MC items: A, B, and C; for CR items: A, B, D1, D2, D3, D4, D5, D6, and corresponding standard errors. For Rasch model A and C will not be supplied.	
F8.5	BSE	SE for B parameter (scaled)		
F8.5	D1	D1 category parameter (scaled)		
F8.5	D1SE	SE for D1 category parameter (scaled)		
F8.5	D2	D2 category parameter (scaled)		
F8.5	D2SE	SE for D2 category parameter (scaled)		
F8.5	D3	D3 category parameter (scaled)		
F8.5	D3SE	SE for D3 category parameter (scaled)		
F8.5	D4	D4 category parameter (scaled)		
F8.5	D4SE	SE for D4 category parameter (scaled)		
F8.5	D5	D5 category parameter (scaled)		
F8.5	D5SE	SE for D5 category parameter (scaled)		
F8.5	D6	D6 category parameter (scaled)		
F8.5	D6SE	SE for D6 category parameter (scaled)		
F8.4	MSQIN	Mean-square infit	Rasch fit index and flag: blank (0.5 < 1.5), MM (misfit moderate: 1.5 < 2.0), MH (misfit high: 2.0 <), TP (too predicTable: < 0.5). Not supplied for 3PL and 2PPC models.	
F8.4	MSQOUT	Mean-square outfit		
A2	MSQFITFL	Mean-square fit flag (blank, MM, MH, TP)		
F1	FITLEV	Misfit level (0, 1, 2)		
F8.5	INFO1	Item information at cut point 1	Item information at performance level cut-points.	
F8.5	INFO2	Item information at cut point 2		
F8.5	INFO3	Item information at cut point 3		
F8.3	TH01	Theta point 1	Theta points for plotting conditional item means.	
F8.3	TH02	Theta point 2		

F8.3	TH03	Theta point 3		
F8.3	TH04	Theta point 4		
F8.3	TH05	Theta point 5		
F8.3	TH06	Theta point 6		
F8.3	TH07	Theta point 7		
F8.3	TH08	Theta point 8		
F8.3	TH09	Theta point 9		
F8.3	TH10	Theta point 10		
F8.3	AD01	Conditional Item Mean for Decile 1	Conditional item means plot: All	
F8.3	AD02	Conditional Item Mean for Decile 2		
F8.3	AD03	Conditional Item Mean for Decile 3		
F8.3	AD04	Conditional Item Mean for Decile 4		
F8.3	AD05	Conditional Item Mean for Decile 5		
F8.3	AD06	Conditional Item Mean for Decile 6		
F8.3	AD07	Conditional Item Mean for Decile 7		
F8.3	AD08	Conditional Item Mean for Decile 8		
F8.3	AD09	Conditional Item Mean for Decile 9		
F8.3	AD10	Conditional Item Mean for Decile 10		
F8.3	MD01	Conditional Item Mean for Decile 1	Conditional item means plot: Males	
F8.3	MD02	Conditional Item Mean for Decile 2		
F8.3	MD03	Conditional Item Mean for Decile 3		
F8.3	MD04	Conditional Item Mean for Decile 4		
F8.3	MD05	Conditional Item Mean for Decile 5		
F8.3	MD06	Conditional Item Mean for Decile 6		
F8.3	MD07	Conditional Item Mean for Decile 7		
F8.3	MD08	Conditional Item Mean for Decile 8		
F8.3	MD09	Conditional Item Mean for Decile 9		
F8.3	MD10	Conditional Item Mean for Decile 10		
F8.3	FD01	Conditional Item Mean for Decile 1	Conditional item means plot: Females	
F8.3	FD02	Conditional Item Mean for Decile 2		
F8.3	FD03	Conditional Item Mean for Decile 3		
F8.3	FD04	Conditional Item Mean for Decile 4		
F8.3	FD05	Conditional Item Mean for Decile 5		
F8.3	FD06	Conditional Item Mean for Decile 6		

F8.3	FD07	Conditional Item Mean for Decile 7		
F8.3	FD08	Conditional Item Mean for Decile 8		
F8.3	FD09	Conditional Item Mean for Decile 9		
F8.3	FD10	Conditional Item Mean for Decile 10		
F8.3	WD01	Conditional Item Mean for Decile 1	Conditional item means plot: Whites	
F8.3	WD02	Conditional Item Mean for Decile 2		
F8.3	WD03	Conditional Item Mean for Decile 3		
F8.3	WD04	Conditional Item Mean for Decile 4		
F8.3	WD05	Conditional Item Mean for Decile 5		
F8.3	WD06	Conditional Item Mean for Decile 6		
F8.3	WD07	Conditional Item Mean for Decile 7		
F8.3	WD08	Conditional Item Mean for Decile 8		
F8.3	WD09	Conditional Item Mean for Decile 9		
F8.3	WD10	Conditional Item Mean for Decile 10		
F8.3	BD01	Conditional Item Mean for Decile 1	Conditional item means plot: Blacks	
F8.3	BD02	Conditional Item Mean for Decile 2		
F8.3	BD03	Conditional Item Mean for Decile 3		
F8.3	BD04	Conditional Item Mean for Decile 4		
F8.3	BD05	Conditional Item Mean for Decile 5		
F8.3	BD06	Conditional Item Mean for Decile 6		
F8.3	BD07	Conditional Item Mean for Decile 7		
F8.3	BD08	Conditional Item Mean for Decile 8		
F8.3	BD09	Conditional Item Mean for Decile 9		
F8.3	BD10	Conditional Item Mean for Decile 10		
F8.3	A95_A0	95th percentile	Box & whisker plot: All	Option A /Score 0
F8.3	A75_A0	75th percentile		
F8.3	A50_A0	50th percentile		
F8.3	A25_A0	25th percentile		
F8.3	A05_A0	5th percentile		
F8.3	M95_A0	95th percentile	Box & whisker plot: Males	
F8.3	M75_A0	75th percentile		
F8.3	M50_A0	50th percentile		
F8.3	M25_A0	25th percentile		

F8.3	M05_A0	5th percentile		
F8.3	F95_A0	95th percentile	Box & whisker plot: Females	
F8.3	F75_A0	75th percentile		
F8.3	F50_A0	50th percentile		
F8.3	F25_A0	25th percentile		
F8.3	F05_A0	5th percentile		
F8.3	W95_A0	95th percentile	Box & whisker plot: Whites	
F8.3	W75_A0	75th percentile		
F8.3	W50_A0	50th percentile		
F8.3	W25_A0	25th percentile		
F8.3	W05_A0	5th percentile		
F8.3	B95_A0	95th percentile	Box & whisker plot: Blacks	
F8.3	B75_A0	75th percentile		
F8.3	B50_A0	50th percentile		
F8.3	B25_A0	25th percentile		
F8.3	B05_A0	5th percentile		
F8.3	A95_B1	95th percentile	Box & whisker plot: All	Option B /Score 1
F8.3	A75_B1	75th percentile		
F8.3	A50_B1	50th percentile		
F8.3	A25_B1	25th percentile		
F8.3	A05_B1	5th percentile		
F8.3	M95_B1	95th percentile	Box & whisker plot: Males	
F8.3	M75_B1	75th percentile		
F8.3	M50_B1	50th percentile		
F8.3	M25_B1	25th percentile		
F8.3	M05_B1	5th percentile		
F8.3	F95_B1	95th percentile	Box & whisker plot: Females	
F8.3	F75_B1	75th percentile		
F8.3	F50_B1	50th percentile		
F8.3	F25_B1	25th percentile		
F8.3	F05_B1	5th percentile		
F8.3	W95_B1	95th percentile	Box & whisker plot: Whites	
F8.3	W75_B1	75th percentile		
F8.3	W50_B1	50th percentile		
F8.3	W25_B1	25th percentile		

F8.3	W05_B1	5th percentile		
F8.3	B95_B1	95th percentile	Box & whisker plot: Blacks	
F8.3	B75_B1	75th percentile		
F8.3	B50_B1	50th percentile		
F8.3	B25_B1	25th percentile		
F8.3	B05_B1	5th percentile		
F8.3	A95_C2	95th percentile	Box & whisker plot: All	Option C /Score 2
F8.3	A75_C2	75th percentile		
F8.3	A50_C2	50th percentile		
F8.3	A25_C2	25th percentile		
F8.3	A05_C2	5th percentile		
F8.3	M95_C2	95th percentile	Box & whisker plot: Males	
F8.3	M75_C2	75th percentile		
F8.3	M50_C2	50th percentile		
F8.3	M25_C2	25th percentile		
F8.3	M05_C2	5th percentile		
F8.3	F95_C2	95th percentile	Box & whisker plot: Females	
F8.3	F75_C2	75th percentile		
F8.3	F50_C2	50th percentile		
F8.3	F25_C2	25th percentile		
F8.3	F05_C2	5th percentile		
F8.3	W95_C2	95th percentile	Box & whisker plot: Whites	
F8.3	W75_C2	75th percentile		
F8.3	W50_C2	50th percentile		
F8.3	W25_C2	25th percentile		
F8.3	W05_C2	5th percentile		
F8.3	B95_C2	95th percentile	Box & whisker plot: Blacks	
F8.3	B75_C2	75th percentile		
F8.3	B50_C2	50th percentile		
F8.3	B25_C2	25th percentile		
F8.3	B05_C2	5th percentile		
F8.3	A95_D3	95th percentile	Box & whisker plot: All	Option D /Score 3
F8.3	A75_D3	75th percentile		
F8.3	A50_D3	50th percentile		
F8.3	A25_D3	25th percentile		

F8.3	A05_D3	5th percentile		
F8.3	M95_D3	95th percentile	Box & whisker plot: Males	
F8.3	M75_D3	75th percentile		
F8.3	M50_D3	50th percentile		
F8.3	M25_D3	25th percentile		
F8.3	M05_D3	5th percentile		
F8.3	F95_D3	95th percentile	Box & whisker plot: Females	
F8.3	F75_D3	75th percentile		
F8.3	F50_D3	50th percentile		
F8.3	F25_D3	25th percentile		
F8.3	F05_D3	5th percentile		
F8.3	W95_D3	95th percentile	Box & whisker plot: Whites	
F8.3	W75_D3	75th percentile		
F8.3	W50_D3	50th percentile		
F8.3	W25_D3	25th percentile		
F8.3	W05_D3	5th percentile		
F8.3	B95_D3	95th percentile	Box & whisker plot: Blacks	
F8.3	B75_D3	75th percentile		
F8.3	B50_D3	50th percentile		
F8.3	B25_D3	25th percentile		
F8.3	B05_D3	5th percentile		
F8.3	A95_4	95th percentile	Box & whisker plot: All	Score 4
F8.3	A75_4	75th percentile		
F8.3	A50_4	50th percentile		
F8.3	A25_4	25th percentile		
F8.3	A05_4	5th percentile		
F8.3	M95_4	95th percentile	Box & whisker plot: Males	
F8.3	M75_4	75th percentile		
F8.3	M50_4	50th percentile		
F8.3	M25_4	25th percentile		
F8.3	M05_4	5th percentile		
F8.3	F95_4	95th percentile	Box & whisker plot: Females	
F8.3	F75_4	75th percentile		
F8.3	F50_4	50th percentile		
F8.3	F25_4	25th percentile		

F8.3	F05_4	5th percentile		
F8.3	W95_4	95th percentile	Box & whisker plot: Whites	
F8.3	W75_4	75th percentile		
F8.3	W50_4	50th percentile		
F8.3	W25_4	25th percentile		
F8.3	W05_4	5th percentile		
F8.3	B95_4	95th percentile	Box & whisker plot: Blacks	
F8.3	B75_4	75th percentile		
F8.3	B50_4	50th percentile		
F8.3	B25_4	25th percentile		
F8.3	B05_4	5th percentile		
F8.3	A95_5	95th percentile	Box & whisker plot: All	Score 5
F8.3	A75_5	75th percentile		
F8.3	A50_5	50th percentile		
F8.3	A25_5	25th percentile		
F8.3	A05_5	5th percentile		
F8.3	M95_5	95th percentile	Box & whisker plot: Males	
F8.3	M75_5	75th percentile		
F8.3	M50_5	50th percentile		
F8.3	M25_5	25th percentile		
F8.3	M05_5	5th percentile		
F8.3	F95_5	95th percentile	Box & whisker plot: Females	
F8.3	F75_5	75th percentile		
F8.3	F50_5	50th percentile		
F8.3	F25_5	25th percentile		
F8.3	F05_5	5th percentile		
F8.3	W95_5	95th percentile	Box & whisker plot: Whites	
F8.3	W75_5	75th percentile		
F8.3	W50_5	50th percentile		
F8.3	W25_5	25th percentile		
F8.3	W05_5	5th percentile		
F8.3	B95_5	95th percentile	Box & whisker plot: Blacks	
F8.3	B75_5	75th percentile		
F8.3	B50_5	50th percentile		
F8.3	B25_5	25th percentile		

F8.3	B05_5	5th percentile		
F8.3	A95_6	95th percentile	Box & whisker plot: All	Score 6
F8.3	A75_6	75th percentile		
F8.3	A50_6	50th percentile		
F8.3	A25_6	25th percentile		
F8.3	A05_6	5th percentile		
F8.3	M95_6	95th percentile	Box & whisker plot: Males	
F8.3	M75_6	75th percentile		
F8.3	M50_6	50th percentile		
F8.3	M25_6	25th percentile		
F8.3	M05_6	5th percentile		
F8.3	F95_6	95th percentile	Box & whisker plot: Females	
F8.3	F75_6	75th percentile		
F8.3	F50_6	50th percentile		
F8.3	F25_6	25th percentile		
F8.3	F05_6	5th percentile		
F8.3	W95_6	95th percentile	Box & whisker plot: Whites	
F8.3	W75_6	75th percentile		
F8.3	W50_6	50th percentile		
F8.3	W25_6	25th percentile		
F8.3	W05_6	5th percentile		
F8.3	B95_6	95th percentile	Box & whisker plot: Blacks	
F8.3	B75_6	75th percentile		
F8.3	B50_6	50th percentile		
F8.3	B25_6	25th percentile		
F8.3	B05_6	5th percentile		
F8.3	A95_OM	95th percentile	Box & whisker plot: All	Omits
F8.3	A75_OM	75th percentile		
F8.3	A50_OM	50th percentile		
F8.3	A25_OM	25th percentile		
F8.3	A05_OM	5th percentile		
F8.3	M95_OM	95th percentile	Box & whisker plot: Males	
F8.3	M75_OM	75th percentile		
F8.3	M50_OM	50th percentile		
F8.3	M25_OM	25th percentile		

F8.3	M05_OM	5th percentile		
F8.3	F95_OM	95th percentile	Box & whisker plot: Females	
F8.3	F75_OM	75th percentile		
F8.3	F50_OM	50th percentile		
F8.3	F25_OM	25th percentile		
F8.3	F05_OM	5th percentile		
F8.3	W95_OM	95th percentile	Box & whisker plot: Whites	
F8.3	W75_OM	75th percentile		
F8.3	W50_OM	50th percentile		
F8.3	W25_OM	25th percentile		
F8.3	W05_OM	5th percentile		
F8.3	B95_OM	95th percentile	Box & whisker plot: Blacks	
F8.3	B75_OM	75th percentile		
F8.3	B50_OM	50th percentile		
F8.3	B25_OM	25th percentile		
F8.3	B05_OM	5th percentile		

Appendix C. Statistics Used on Item Labels for Item Review Committees

CID Company identification number for the item.

Maturity Function of the reviewed item.

Form Form numbers that contain the reviewed item.

Position Position numbers in the test for the reviewed item (given for each form that the item appears on).

Type Item type: MC – multiple-choice item, CR – constructed-response item, WR – writing.

Key The correct answer for an MC item.

Max The maximum score point for a CR or a writing item.

P-value The percent of students who answered the item correctly. Its theoretical range is 0-1. It indicates item difficulty. Items with high p-values, such as .90, are relatively easy items. Those with p-values below .50 are relatively difficult items. P-values depend on the group of examinees who take the test.

Adj. P value Computed by dividing the item mean by the score range. It is equivalent to the p-value for the MC items when the score point is awarded either 1 or 0.

N-count The number of tested students who were administered the item.

Rasch Difficulty The usual range of Rasch difficulties is from -3 to +3 with mean of 0 and standard deviation of 1. 0 means medium difficulty. Positive values mean difficult items. Negative values mean easy items.

PB Correlation Point-biserial correlation shows the relationship between a student's performance on the item and performance on the test as a whole. A high point-biserial correlation (e.g., above .50) indicates that students who answered the item correctly on the item achieved higher total scores on the test than those who answered the item incorrectly on the item. Values less than .25 may indicate a weaker than desired relationship. Note that extremely difficult or extremely easy items may have point-biserial correlation artificially reduced.

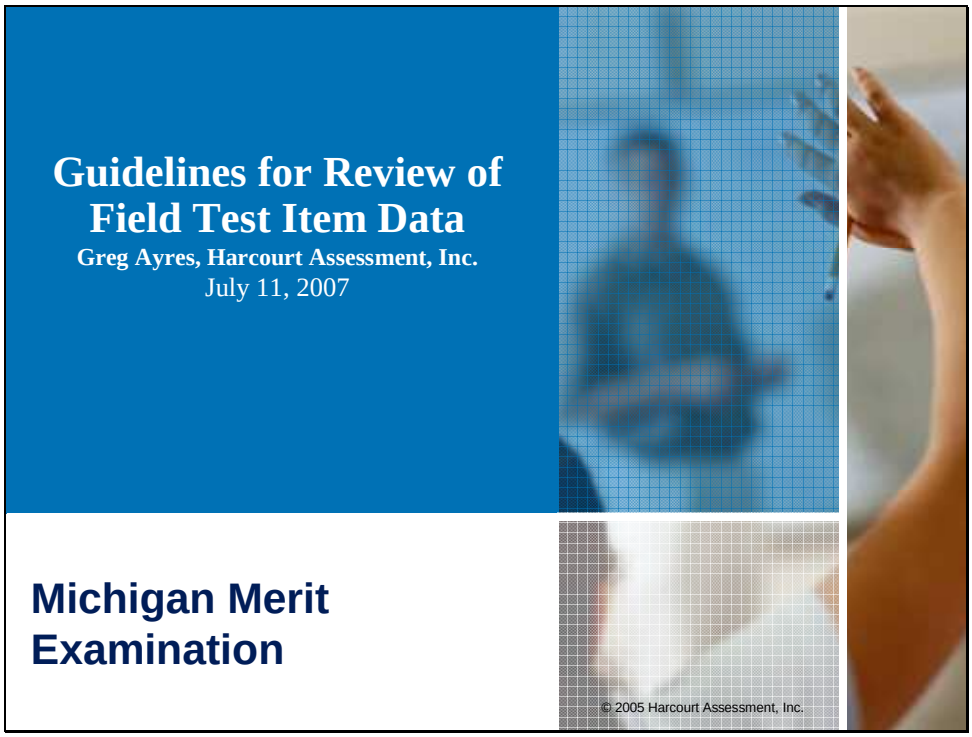
Item-Total Corr. Item-total correlation shows the relationship between a student's performance on the item and performance on the test as a whole. A high item-total correlation (e.g., above .50) indicates that students who earned more points on the item achieved higher total scores on the test than those who earned fewer points on

	the item. Values less than .25 may indicate a weaker than desired relationship. Note that extremely difficult or extremely easy items may have item-total correlation artificially reduced.
FIT Flag	This flag indicates that two fit indices are out of the desired range. It means the Item may have not misfit or overfit the measurement model specified for the test analysis.
Difficulty Flag	This flag indicates that P-value, or adjusted p-value, or Rasch difficulty is out of the desired range.
PB Correlation Flag	This flag indicates that a MC item point-biserial correlation is smaller than the desired range of larger than 0.25.
Item-Total Corr. Flag	This flag indicates that a CR or a Writing item point-biserial correlation is smaller than the desired range of larger than 0.25.
Option Quality Flag	This flag indicates that a MC item may have a key problem. It could be that the key is not correct or it was miskeyed in scoring.
Score Point Dist. Flag	This flag indicates that a CR or a Writing item may have a scoring rubric problem. It could be the sample answer for each score point was not correctly identified.
Option Analysis	Percent of students who selected options A, B, C, and D, or did not choose any option (Omit) for all students and for subgroups by gender and ethnicity.
Score Point Distribution	Percent of students who earned each valid score point and who did not answer the CR or writing item for all students and for subgroups by gender and ethnicity.
Option PB Correlation	Point-biserial correlation for each of a MC item options. The key option point-biserial correlation should be positive and high. The non-keyed option point-biserial should be negative and low.
Omit PB Correlation	Point-biserial correlation for omit of a CR or Writing item. The omit point-biserial correlation should be negative.
Invalid Codes	The codes for invalid responses for a CR or a writing item.
DIF	Differential Item Functioning index. It indicates whether the reviewed item favors a particular subgroup of the student population; thus that group of students may have a higher chance of answering the item correctly or earn

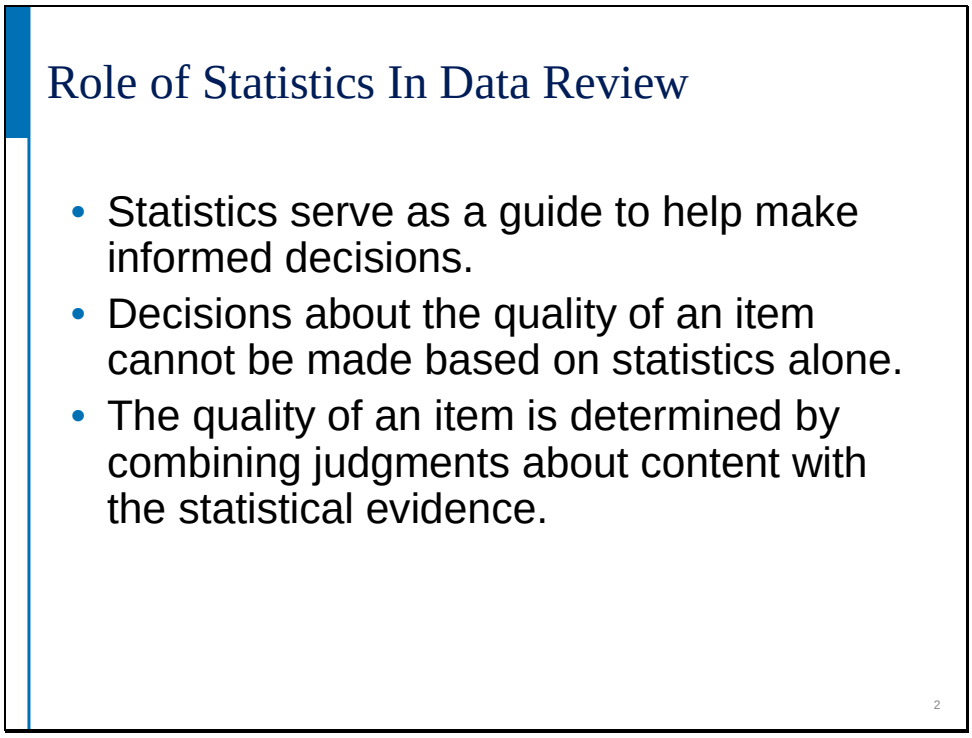
higher score point than the contrasted group. The focused group is often the minority group such as female in the gender group comparison, and black in the ethnic group comparison. The reference group is often the majority group which is male in the gender group comparison, and white in the ethnic group comparison.

Appendix D Guidelines for Bias Review of Field Test Item Data

Slide 1



Slide 2



Slide 3

Statistical Evidence

- Psychometricians collect evidence about item and test characteristics.
- Statistical evidence needs to be weighed to determine whether the item is a good candidate for an operational form.

3

Slide 4

Item Statistics

MME **Grade: 11** **Subject: Math** **Admin: Spring 2007**

ID: 3423345
Form: 8
Position: 13
Scenario: NA

GLCE: F.2.h.06

- ☐ Accept as is
- ☐ Reject
- ☐ Accept with revision

Table 1. Item Information

Type:	MC	**P-value:**	.62	**B parameter:**		**Difficulty Flag:**	
Key:	B	**N-count:**	3695	**PB Correlation:**	0.50	**PB Correlation Flag:**	
		Maturity:	FT	**Fit Flag:**		**Option Quality Flag:**	

4

Slide 5

Table 2. Breakout Group Descriptives and Option Analysis

		N-count	Percent of Students Selected Option				
			A	B *	C	D	Omit
Group	All	3695	7	62	15	16	0
	Male	1797	8	59	16	17	0
	Female	1898	7	64	13	15	0
	White	2913	7	65	13	14	0
	Black	519	7	44	22	26	0
Option PB Correlations			-0.15	0.50	-0.27	-0.29	-0.03

Table 3. Differential Item Functioning

Reference/ Focal Group	Male/ Female	White/ Black
Flag	B	
Favored Group	female	

5

Slide 6

Table 2. Breakout Group Descriptives and Score Point Distributions

		N-count	Item Mean	Percent of Students at Each Score Point							
				0	1	2	3	4	5	6	Omit
Group	All	1977	1.94	5	38	27	20	8	2		
	Male	998	1.75	7	43	26	17	6	1		
	Female	979	2.13	3	33	29	23	10	3		
	White	1572	2.03	4	35	28	22	9	2		
	Black	277	1.43	10	52	24	12	2			
Omit PB Correlation											

Table 3. Condition Code Distributions

Frequency of Students at Each Condition Code		
A	B	C
.400		1.21

Table 4. Differential Item Functioning

Reference/ Focal Group	Male/ Female	White/ Black
Flag	C	
Favored Group	female	

6

Slide 7

Classical Item Difficulty: P-value

- MC items: **P-value** is the percentage of students who answered the item correctly.
- CR items: Adjusted **P-value** is the item mean divided by its range (max score – min score).
- Theoretical range from 0 to 1, with values over 0.9 indicating items that may be too easy, and values below 0.3 indicating items that may be too difficult
- Group dependent (not comparable across administration years)

7

Slide 8

Item Discrimination: Item-Total Correlation

- Item-total correlation indicates agreement between **item** scores and **total** test scores.
  - Point-biserial correlation is a specific type of item-total correlation used for dichotomous items (e.g., MC items).
- Theoretical range from -1 to 1
- High item-total correlation indicates that students who answered an item correctly, or who received a higher score-point on an item, also have higher total test scores (and vice versa).
- Item-total correlation greater than 0.25 are acceptable; those below 0.25 should be scrutinized.

8

Slide 9

Option Analysis / Score Point Distribution

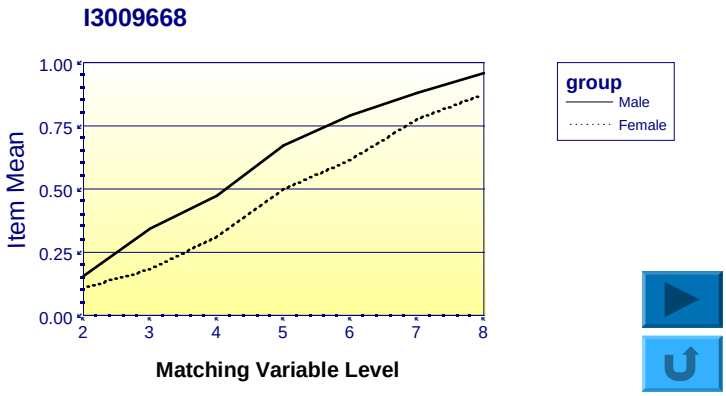
- Shows the percentage of students choosing each option on MC items, or earning a score point on CR items
- This percentage is given for all students and students grouped by ethnicity and gender.
- Option point-biserial correlation indicates the agreement between choosing each option (or earning a score point) and the total score on the test.

9

Slide 10

Differential Item Functioning (DIF) Analysis

- DIF refers to the **unexpected** differences in performance on a studied item between a reference and a focal group **after they have been matched** with respect to the total score on the test.



10

Slide 11

DIF and Item Bias

- An item is biased if it measures attributes irrelevant to the intended construct or is somehow a less acceptable measure of the construct for one subgroup.
- DIF does not necessarily mean that an item is biased. DIF only indicates that the examinees of equal proficiency from different subgroups have an unequal probability of responding correctly to an item.
- The results of DIF analyses provide a convenient starting point for the study of item bias.

11

Slide 12

DIF Levels

- Items are classified into one of the three DIF categories.
  - Category A: Negligible DIF, no group favored
  - Category B: Moderate DIF, one group is slightly favored by the studied item
  - Category C: Large DIF, one group is strongly favored by the studied item
- Items in category B and C are flagged and should be carefully examined for potential bias against a particular group.

12

Slide 13

DIF Table

- DIF flag: An indication of moderate DIF (flag B) or large DIF (flag C)
- Fav group: The flag for indicating which group is favored by the studied item

13

Slide 14

Summary

- Make informed decisions based on the data.
- Information on content and statistics determines the quality of an item.
- Weigh the statistical evidence and content, and then determine whether the items are good candidates for a live form.

14

Slide 15

Next

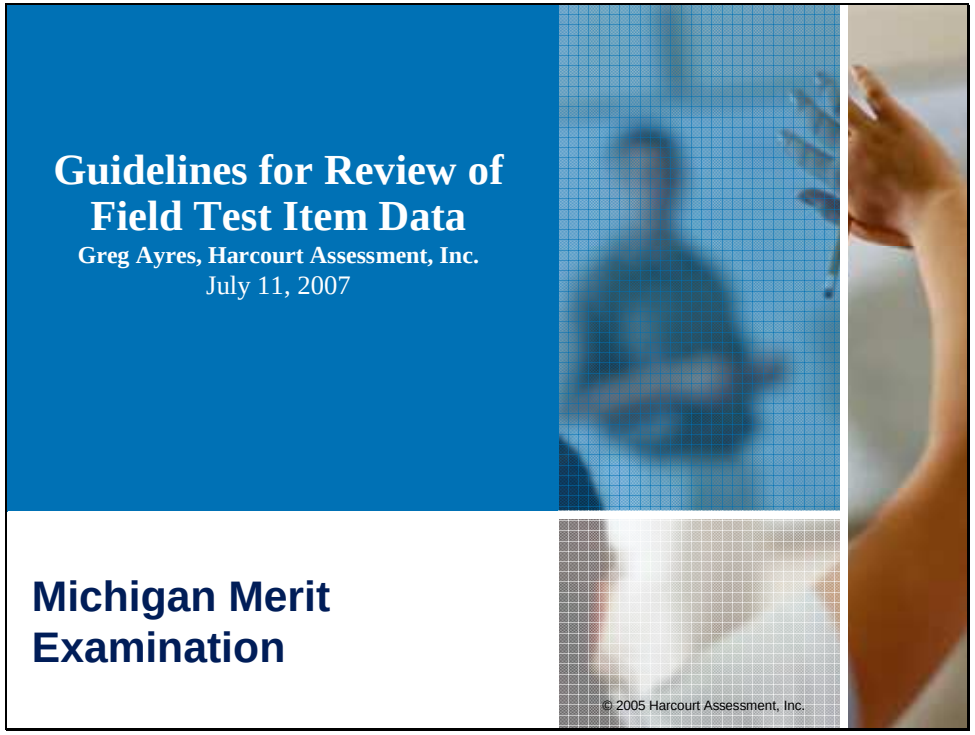
- Ask any questions that you may have
- Work in your respective subject area groups
- Enjoy the process

Thank you!

15

Appendix E Guidelines for Content Review of Field Test Item Data

Slide 1



Slide 2

Role of Statistics In Data Review

- Statistics serve as a guide to help make informed decisions.
- Decisions about the quality of an item cannot be made based on statistics alone.
- The quality of an item is determined by combining judgments about content with the statistical evidence.

2

Slide 3

Statistical Evidence

- Psychometricians collect evidence about item and test characteristics.
- Statistical evidence needs to be weighed to determine whether the item is a good candidate for an operational form.

3

Slide 4

Item Statistics

MME Grade: 11 Subject: Math Admin: Spring 2007

ID: 3423170
Form: 2
Position: 14
Scenario: NA

GLCE: G.1.h.05

- □ Accept as is
- □ Reject
- □ Accept with revision

Table 1. Item Information

Type:	MC	**P-value:**	.32	**B parameter:**		**Difficulty Flag:**	
Key:	C	**N-count:**	3718	**PB Correlation:**	0.24	**PB Correlation Flag:**	CL
		Maturity:	FT	**Fit Flag:**		**Option Quality Flag:**	P

4

Slide 5

Table 2. Breakout Group Descriptives and Option Analysis

		N-count	Percent of Students Selected Option				
			A	B	C *	D	Omit
Group	All	3718	14	31	32	22	0
	Male	1810	14	29	37	21	0
	Female	1908	14	34	29	23	0
	White	2898	13	31	33	22	0
	Black	539	17	34	28	20	0
Option PB Correlations			-0.24	-0.16	0.24	0.12	-0.04

Table 3. Differential Item Functioning

Reference/ Focal Group	Male/ Female	White/ Black
Flag		
Favored Group		

5

Slide 6

Table 2. Breakout Group Descriptives and Score Point Distributions

		N-count	Item Mean	Percent of Students at Each Score Point							
				0	1	2	3	4	5	6	Omit
Group	All	1977	1.94	5	38	27	20	8	2		
	Male	998	1.75	7	43	26	17	6	1		
	Female	979	2.13	3	33	29	23	10	3		
	White	1572	2.03	4	35	28	22	9	2		
	Black	277	1.43	10	52	24	12	2			
Omit PB Correlation											

Table 3. Condition Code Distributions

Frequency of Students at Each Condition Code		
A	B	C
.400		1.21

Table 4. Differential Item Functioning

Reference/ Focal Group	Male/ Female	White/ Black
Flag	C	
Favored Group	female	

6

Slide 7

Classical Item Difficulty: P-value

- MC items: **P-value** is the percentage of students who answered the item correctly.
- CR items: Adjusted **P-value** is the item mean divided by its range (max score – min score).
- Theoretical range from 0 to 1, with values over 0.9 indicating items that may be too easy, and values below 0.3 indicating items that may be too difficult
- Group dependent (not comparable across administration years)

7

Slide 8

Item Discrimination: Item-Total Correlation

- Item-total correlation indicates agreement between **item** scores and **total** test scores.
  - Point-biserial correlation is a specific type of item-total correlation used for dichotomous items (e.g., MC items).
- Theoretical range from -1 to 1
- High item-total correlation indicates that students who answered an item correctly, or who received a higher score-point on an item, also have higher total test scores (and vice versa).
- Item-total correlation greater than 0.25 are acceptable; those below 0.25 should be scrutinized.

8

Slide 9

Option Analysis / Score Point Distribution

- Shows the percentage of students choosing each option on MC items, or earning a score point on CR items
- This percentage is given for all students and students grouped by ethnicity and gender.
- Option point-biserial correlation indicates the agreement between choosing each option (or earning a score point) and the total score on the test.

9

Slide 10

Summary

- Make informed decisions based on the data.
- Information on content and statistics determines the quality of an item.
- Weigh the statistical evidence and content, and then determine whether the items are good candidates for a live form.

10

Slide 11

Next

- Ask any questions that you may have
- Work in your respective subject area groups
- Enjoy the process

Thank you!

11