MICHIGAN MERIT EXAMINATION

Technical Manual

2011 Testing Cycle

December 22, 2011

Preface

This purpose of this manual is to document the technical characteristics of the 2011 Michigan Merit Examination (MME) based on the results of the 2011 operational administration. Analytic results are provided by Michigan's Bureau of Assessment and Accountability (BAA), formerly, Office of Educational Assessment and Accountability (OEAA), ACT, Inc., Measurement, Inc., and Pearson Educational Measurement (PEM). This manual includes information regarding: (1) changes implemented in the 2011 MME administration, (2) background to the test, (3) test development analyses, (4) erasure analyses, (5) ACT writing scoring analyses, (6) model fit analyses, (7) scaling and equating information related to linking across MME forms, (8) reliability and validity information, (9) item analysis information, (10) standard setting information, and (11) information related to Adequate Yearly Progress and Education YES.

The Michigan Merit Examination (MME) is used to assess Grade 11 and eligible Grade 12 students on Michigan's English language arts (ELA), mathematics, science, and social studies high school content standards and expectations. It is designed differently than other statewide assessments in that the MME has three distinct components: (1) the ACT Plus Writing college entrance examination, (2) *WorkKeys* job skill assessments in *Reading for Information, Applied Mathematics,* and *Locating Information*; and (3) Michigan-specific assessments in mathematics, science, and social studies. Each component is administered on a different day. The ACT Plus Writing component is administered on Day 1, the *WorkKeys* component is administered on Day 2, and the Michigan component is administered on Day 3.

We encourage individuals who are interested in receiving more detailed information on topics discussed in this manual, or on related topics, to contact the Bureau of Assessment and Accountability (BAA) at the Michigan Department of Education (http://www.michigan.gov/mde).

Bureau of Assessment and Accountability
Michigan Department of Education
608 W. Allegan Street
P.O. Box 30008
Lansing, MI 48909

Table of Contents

PREFACE I

CHAPTER 1: CHANGES FOR THE MME 2011 ADMINISTRATION 1

CHAPTER 2: BACKGROUND TO THE MICHIGAN MERIT EXAMINATION 2
School Structure and MME Administration 2
Students to Be Tested 2
MME Assessment Components and Schedule 3
Appropriate Uses for Scores and Reports 4
School, District, Intermediate School District, and State Reports 6
Organizations Involved in MME Testing 7
Michigan Department of Education (MDE) 7
Bureau of Assessment and Accountability (BAA) 7
Center for Educational Performance and Information (CEPI) 8
Department of Information Technology (DIT) 8
Department of Educational Technology 8
Contractors and Subcontractors 9
Educators 9
Technical Advisory Committee 11
Michigan State Board of Education 11

CHAPTER 3: TEST DEVELOPMENT (ACT, *WORKKEYS*, AND MICHIGAN COMPONENTS) 13
Test Development for Day 1 and Day 2 13
The ACT Plus Writing Test 13
Philosophical Basis for the ACT 13
Description of the ACT Plus Writing 14
The English Test 15
The Mathematics Test 15
The Reading Test 16
The Science Test 16
The Writing Test 17
Test Development Procedures for the ACT Multiple-Choice Tests 17
Reviewing Test Specifications 17
Content specifications 17
Statistical specifications 18
Selection of Item Writers 22
Item Construction 22
Review of Items 22
Item Tryouts 23
Item Analysis of Tryout Units 23
Assembly of New Forms 23
Content and Fairness Review of Test Forms 24
Review Following Operational Administration 25
Test Development Procedures for the ACT Writing Test 26
Selection and Training of Prompt Writers 26
Prompt Construction 26
Content and Fairness Review of Prompts 26
Field Testing of Prompts 26
Review of Field Tests and Operational Administration 26
ACT Scoring Procedures 26
Technical Characteristics of the ACT Tests 28
The *WorkKeys* Assessments Components: Reading for Information, Applied Mathematics, and Locating Information 28
The WorkKeys Assessment Development Process 29
Skill Definition 29
Test Specifications 30
Prototyping 30

Pretesting 31
Operational Forms 31
WorkKeys Assessment Descriptions 32
Applied Mathematics 32
Reading for Information 34
Locating Information 37
Technical Characteristics of the WorkKeys Tests 38
Test Development for Day 3 38
Test Specifications 38
Step 1: Specification Development 38
Step 2: Item Writer Training 39
Step 3: Item Development 39
Step 4: Item Review 40
Step 5: Contractor Review 41
Step 6: Field Testing 41
Step 7: Field Test Item Review 41
Field Testing Procedures: Item Development, Review, Field Test Design, and Statistics 41
Field Testing Design 41
Field Test Sampling 42
Item Specifications 42
Item Statistics 42
For Multiple-Choice (MC) Items 42
Differential Item Functioning 43
Field Testing Embedding 43
Post-Field-Test Item Review 43
Field Test Item Statistics and Data 43
Statistics Prepared for Review Committees 44
Item Reviews 47
Bias/Sensitivity and Content Committee Review 47
Item Revision Procedures 47
Item Banking 48
Procedures 48
Existing Architecture 48
Populate the Item Bank 48
Data Verification Process: 48
Quality Control Procedures for Item Bank 49
Quality Control of Items 49
Quality Control of Metadata 49
Quality Control of Item Statistics 50
Quality Control of Functionality 50
Data Included in Item Bank 50
Construction of Operational Test Forms 50
Assessment Blueprints 50
Mathematics 51
Science 51
Social Studies 52
Accommodated Formats 52
Item Selection 52
Select Assessment Items to Meet the Assessment Blueprints 52
Assess the Statistical Characteristics of the Selected Assessment Items 52
For Multiple-Choice (MC) Items 53
Review and Approve Test Forms 53
Accommodated Test Forms 54
Accommodated Format Production: Day 1 ACT Plus Writing 54
Large Print 55
Oral Presentation 55
Translated and Video Formats: State-Allowed Administrations. 55
Accommodated Format Production: Day 2 *WorkKeys* 56
Translations 56

Braille 56
Other media 57
Accommodated Format Production: Day 3 Michigan Components 57
Braille 57
Large Print 57
Oral Administration 57
Bilingual Tests 58

CHAPTER 4: ADMINISTRATION 59
Preparation for Test Administration 60
MME Day 1 and Day 2: Materials Processing 63
Day 1 Processing 64
Day 2 Processing 64
Test Security—Days 1 and 2 65
Materials Return—Day 1 and Day 2 65
Materials Discrepancy Checking Process—Day 1 and Day 2 66
Processing Assessment Materials Returned by Schools—Day 1 and Day 2 66
MME Day 3 Michigan Components: Materials Processing 67
Accommodations for Students with Disabilities (SWD) and English Language Learners (ELL) 76
Spring 2011 Test Organization 77
Spring 2011 Michigan Merit Examination (MME) Accommodations Summary Table 84
Accommodation 84
MME Day 1 (The ACT Plus Writing) 84
MME Day 2 84
MME Day 3 84
Standard/Nonstandard 84
State Use Questions 111

CHAPTER 5: TEST DEVELOPMENT ANALYSES 122
MME Components 122
Test Specifications and Alignment Between Contributing Components 122
Alignment of the 2011 MME with HSCEs: Item Selection for Day 1 and Day 2 Scoring 123
Test Development for Michigan Components 125
Historical Alignment Analyses Prior to 2011 Administration 127
Post-Hoc Alignment Studies of the Pilot Michigan Merit Examination 127
Post-Hoc Alignment Studies of the 2010 Michigan Merit Examination 128

CHAPTER 6: ERASURE ANALYSES 131
Description and Purpose 131
Data and Methods 131
Day 1 and 2 Analysis 132
Overview 132
Input 133
Output 133
Day 3 Erasure Analysis 135

CHAPTER 7: ACT WRITING TRAINING AND SCORING 140
Results of Constructed Response Scoring Procedures 140
Rangefinding 140
Rater Training 140
Scoring 141
Comment Codes 141
Rater Monitoring 141
Rater Validity Checks 141
Inter-Rater Reliability 142

CHAPTER 8: MODEL FIT 143
Item-Model Fit 143

Item-Model Fit Indexes 143
Person-Model Fit 160
The *Lz* Person Fit Index 160

CHAPTER 9: SCALING AND EQUATING 168
Quality Control Protocols for MME Calibrations 168
Equating for ACT 169
Equating for *WorkKeys* 170
Equating for MME Social Studies 170
Equating for MME Writing, Reading, Mathematics, Science 171
Calibration Summary Reports 175
IRT Model Fit and Plots 175
Item Analysis 175
Theta Generation 176
IRT Models 176
Algorithms for the Scoring Programs 176
Results of Test Runs 177
Scoring Procedures for the Spring 2011 MME Administration 177

CHAPTER 10: SCORE PRECISION 178
Internal Consistency Reliability 178
Further Evidence of Reliability on ACT Writing 178
Empirical IRT Reliability 179
MME Scale Score Reliability 179
SEM/Information Curves with Cut Scores (Imposed) 180
Classification Consistency and Classification Accuracy 180

CHAPTER 11: VALIDITY 184
Construct Validity Evidence from Content and Curricular Validity 184
Relation to Statewide Content Standards 184
MME Alignment Studies 185
Educator Input 185
Construct-related Validity Evidence from Criterion Validity Analyses 186
Criterion-related Validity Evidence for MME Science 188
DIF Analyses of the Spring 2011 MME 189
Matching Criterion 190
Validity Evidence for the Day 1 Stand Alone Component: ACT Assessment 195
Measuring Educational Achievement 195
Making Admissions Decisions 195
Course Placement Decisions 195
Indicators of Educational Effectiveness 196
Evaluating Probable College Success 196
Validity Evidence for the Day 2 Stand Alone Component: the *WorkKeys* Assessments 196
Content-Related Evidence 196
Criterion-Related Evidence 196
Construct-Related Evidence 197
Gender and Race/Ethnicity Analyses 197
Fairness Review 197
Fairness Review for the ACT 197
Item Writing, Review, and Pretesting 197
Operational Forms Construction 198
Differential Item Functioning (DIF) 198
Fairness review for *WorkKeys* 199
Conclusion 199

CHAPTER 12: ITEM ANALYSIS 200
Post-Field-Test Item Review 200
Data 200

Statistics and Graphs Prepared for Review Committees **201**

CHAPTER 13: STANDARD SETTING **206**

CHAPTER 14: ADEQUATE YEARLY PROGRESS AND EDUCATION YES **207**
Legislative Grounding **207**
Procedures for Using Assessment Data for Accountability **207**
Achievement Status **209**
Achievement Change **210**

CHAPTER 15: STATE SUMMARY DATA **217**

CHAPTER 16: MME SCALE SCORE HISTORY **240**

REFERENCES **242**

APPENDICES **245**
Appendix A: Plots of MME Test Information Functions **246**
Spring 2011 Writing Initial Form 246
Spring 2011 Reading Initial Form 247
Spring 2011 Mathematics Initial Form 1101 248
Spring 2011 Science Initial Form 1101 249
Spring 2011 Social Studies Initial Form 250
Appendix B: Data Created for Field-Test Items **251**
Appendix C: Statistics and Terms Used on Item Labels for Item Review Committees **302**
Appendix D: Guidelines for Bias Review of Field Test Item Data **304**
Appendix E: Guidelines for Content Review of Field Test Item Data **312**

Chapter 1: Changes for the MME 2011 Administration

Statewide assessments undergo periodic modifications or changes to their design, administration, and reporting. In the Spring 2011 MME test cycle, there were several such modifications or changes made to the MME. These modifications or changes are detailed below.

Modifications to the Michigan-Developed Test Design

For the Spring 2011 test cycle, the total number of items for the Day 3 Michigan-developed Social Studies test increased from 28 (on the Spring 2010 test) to 32 operational items plus 16 field test items. This change was due to the fact that Spring 2011 was the first test cycle that Social Studies was completely aligned with and assessed on the Michigan High School Content Expectations (HSCE's).

Changes to Administration Time

For initial (standard) and makeup test forms, the administration time for Day 3 Science was reduced by five minutes (from 40 to 35) and the administration time for Social Studies increased by five minutes (from 30 to 35) as compared with the time allowed for Spring 2009.

Chapter 2: Background to the Michigan Merit Examination

School Structure and MME Administration

Michigan's 57 intermediate school districts provide leadership, services, resources, and programs to Michigan districts and schools. The intermediate school districts include more than 550 public school districts, which consist of approximately 4,000 school buildings and approximately 125,000 students per grade. There were approximately 1,000 high schools in operation during the Spring 2011 test cycle. Public school academies (charter schools) are also required to administer the MME assessments. There are approximately 190 public school academies in the state. The MME assessments are administered to all Grade 11 and eligible Grade 12 students, including those with exceptional needs and English language learners.

There were 1,014 home schooled students reported in the Michigan Student Data System (MSDS) during the 2010–2011 school year in the state of Michigan. However, the actual number could be greater than reported here because homeschooled students are not required to be reported in MSDS. Homeschooled students are given the opportunity to be assessed at their local public school district. According to the Michigan Department of Education's Office of Nonpublic & Home Schools, for the 2010–2011 school year, there were 694 open-active nonpublic schools, with approximately 128,407 students in total. The total number of nonpublic school students per grade varied from over 9,000 for the lower grades and to about 8,000 students per grade for the upper grades. Michigan law requires the state to provide MME assessment opportunities to high school students who attend nonpublic schools. This is accomplished by giving nonpublic schools the opportunity to establish as ACT test centers or to send students to one of several auxiliary test centers located in several regions throughout the state. Participation of nonpublic schools and students is voluntary.

Students to Be Tested

Schools administered all three components of the MME to all public school and charter school students enrolled in Grade 11 during the Spring 2011 testing window. There are two exceptions:

1. A Grade 11 student is NOT to be tested on the MME if the student's IEP indicates that the student should take MI-Access, Michigan's alternate assessment. A student who takes MI-Access in Spring 2011 wasn't allowed to take any portion of the MME in Spring 2011.
2. A Grade 11 student (retained or reclassified as Grade 11) is NOT to be tested on the MME if the student has taken the complete MME in a previous year and has achieved a performance level of either 1,2,3, or 4 in **each** MME subject area, including reading, writing, mathematics, science, and social studies. A student who has a reported performance level of "N/A" or a blank performance level, in any MME subject area, is considered to have not yet taken the complete MME. They are eligible to retake the test, but would not be required to. In case they retake the MME, these students **must** take the complete (all three day) MME.

Michigan law now requires that the complete MME be administered to a student once and only once. A Grade 12 student is only eligible, but not required, to take the MME only if either of the following is true:

- The student is a first-time tester who has not previously taken the MME.
- The student has taken the MME previously but received an invalid MME score (blank or "N/A" performance level) in any of the MME subjects tested, including reading, writing, mathematics, science, or social studies.

A 12th grader not previously tested due to grade reclassification (e.g., a tenth grader the year before and now categorized as a twelfth grader without ever having been an eleventh grader) must be tested or a school is likely to face accountability consequences.

MME Assessment Components and Schedule

The MME is composed of three distinct components: (1) the ACT Plus Writing college entrance examination, (2) WorkKeys job skills assessments in *Reading for Information, Applied Mathematics,* and *Locating Information,* and (3) the Michigan-developed assessments in mathematics, science, and social studies. Table 2.1 presents the MME assessment components for the 2010–2011 school year.

Table 2.1: MME Components and Sections

MME Day	MME Component	Sections	Reading	Writing	Mathematics	Science	Social Studies
Day 1	ACT Plus Writing	English		S			
		Mathematics			S		
		Reading	S				
		Science				S	
		Writing		A			
Day 2	WorkKeys	Reading for Information	S				
		Applied Mathematics			S		
		Locating Information			S		S
Day 3	Michigan Component	Mathematics			A		
		Science				A	
		Social Studies					A

Note: The shaded area shows the sections in each component that contribute to a student's MME score in each subject area. An "A" means all operational items in that section contribute to the student's MME score, and an "S" means select items in that section contribute to the MME score.

Appropriate Uses for Scores and Reports

Following administration of the MME assessment, reports are provided to help educators understand and use the MME assessment results. Under the No Child Left Behind act, all schools are required to demonstrate that their students are making progress and that student achievement is increasing. Information from the MME helps to make that determination at the high school level. The reports provide educators, parents, the public as well as students with an understanding of the educational progress of Michigan students.

Properly used, MME assessment results can:

- measure the status of high school academic achievement as compared with high school content expectations, and the extent to which academic achievement is improving over time;
- evaluate instruction, curriculum and policies designed to improve academic achievement; and
- target academic remediation and support for students where it is needed.

Data Reporting Guidelines and Restrictions

In September 2009, the BAA (then OEAA) published an updated ethics document, the *Assessment Integrity Guide*, which replaced *Professional Assessment and Accountability Practices for Educators*, published in August 2005. Section 6 of this report provides the following specific instructions for appropriate and ethical data reporting that must be followed by school personnel when using data generated from the MME assessment:

<u>School personnel will:</u>

1. Understand and comply with Michigan and United States laws that apply to the handling of family privacy and student data including but not limited to the Family Educational Rights and Privacy Act (1997) and the Michigan Freedom of Information Act (1996).
2. Focus on student achievement to improve individual and program performance.
3. Maintain student confidentiality at all times.
4. Ensure that the information is reported to parents and teachers as soon as possible to determine individual strengths and weaknesses.
5. Ensure that student information is accurate before placing it in the student's permanent records.
6. Analyze student attainment and scores in conjunction with MDE Grade Level Content Expectations, or High School Level Content expectations, and Benchmarks.
7. Analyze results in the context of the school program as a whole, not in isolation.
8. Remind the community that various factors affect test performance and factors such as the following need be taken into consideration when analyzing test results: cultural backgrounds, health conditions, economic status, and former educational experiences.

<u>School personnel will not:</u>

1. Expose any personally identifiable information to anyone other than the student or parents/legal guardian or designated school personnel. (Public law requires the protection of student information).
2. Report on sub-groups of students that would lead to inadvertent identification of students. State results are reported for sub-group sizes of ten students per group or more. Smaller group sizes may inadvertently expose student identities.
3. Use names, student ID numbers, birthdates, gender, race or student ID numbers which may appear on reports on any public information. Names may be used on recognized achievement awards.

4. Falsify student records to alter the accuracy of reported results.
5. Misuse or misrepresent the meaning and interpretation of any student scores. (pp.25-26)

Brief descriptions of MME score reports are provided below. More extensive descriptions with samples are included in *Spring 2011 Guide to Reports,* MME. The guide also outlines information about scale score, performance level, machine-scoring process, and hand-scoring process, and includes notes to help interpret score report data. *The Guide to Reports* are available at the BAA website: http://www.michigan.gov/documents/mde/2011_MME_GTR_Final_355746_7.pdf.

Reports for Individual Students

Individual Student Reports

The Individual Student Report (ISR) provides a detailed description of each student's performance in the content areas assessed on the MME. The ISR is designed to help educators identify their students' academic strengths and areas that may need improvement. Schools may include these reports in student record files.

At the top of the ISR, a student performance summary appears to the right of the student's name and demographic information. The student performance summary contains the student's scale score, and the performance level attained for each subject.

The main section of the ISR presents detailed information on the individual student's performance for each high school content standards. Selected ACT items, selected *WorkKeys* items and all Michigan-developed operational items are included. The number of points earned out of the total number of possible points is reported for each content standard assessed. ACT and *WorkKeys* scores are also included on the ISR.

Parent Reports

The Parent Report presents a summary description of the student's performance by high school content standard for each subject area assessed on the MME, as well as scale scores and performance level information. ACT and *WorkKeys* scores are also included on the Parent Report. One copy of the Parent Report is produced for each student and schools are required to distribute these immediately upon receipt to the respective parent/guardian.

Student Labels

Student Record Labels present summaries of individual scale scores and performance levels in all content areas in label format. The labels are distributed to the schools for placement in the student record files (CA-60). The Record Labels are printed for each student who took the MME.

Student Rosters

Student Rosters are distributed to schools. A Student Roster presents a listing of students by grade. Each student's name, UIC and Date of Birth, subject area scale score, performance level and number of points attained per standard are presented. The top row above the listing of student names shows the possible number of points per standard by subject area.

The top section of the School Roster identifies the grade level reported, the assessment cycle, and the subject area. The school name and code, and the district name and code are also provided.

The MME doesn't use Class/Group ID sheets to group students by teachers or counselors. Schools can choose to provide codes for grouping students and then the Student Roster would list students grouped according to the codes.

School, District, Intermediate School District, and State Reports

There are four different types of aggregate reports generated for schools, districts, ISDs and the state, respectively: (1) summary reports at the school, district, and state level; (2) comprehensive reports at the district and ISD levels; (3) Demographic Reports at the school, district, and state level, and (4) Student Rosters at the school level.

Aggregate reports provide a comparative set of mean scale score information summarized by school, district, ISD and state. The aggregate reports are generated for three student populations:

- all students;
- students with disabilities (SWD); and
- all except students with disabilities (AESWD).

Summary Reports summarize student performance at the school, district and state levels. The top section of each Summary Report identifies the title of the report, the level of aggregation (school, district, state), the student population included in the report, the grade level and the assessment cycle, School and district names and codes are given as appropriate to the report level.

The first page of the Summary Report provides summary data for each content area for the current year and the four previous years. Also presented for each year are 1.) the number of students assessed, 2.) the mean scale score, 3.) the scale score margin of error, 4.) the percentage of students at each performance level, and 5.) the percentage of students at proficient and advanced levels combined.

The second page of the MME Summary Report presents subscore information for each content area and strands within content area for the current test cycle. The subscore data include the number of students assessed, the mean points earned by standard, the total number of possible points for each standard, and the percentage of students earning each raw score point or point range.

Comprehensive Reports are produced for the district and ISD and provide the mean scale score information and percentage of students at each performance level by subject. The district report shows district-level results and results for each school within the district. The ISD report provides ISD-level results and results for each district, and each Public School Academy (PSA) located in the ISD.

Demographic Reports provide a summary of scores by demographic subgroup for each subject area assessed. The data reported includes the number of students assessed in each subgroup, the mean scale score for that subgroup, the percentage of students at each performance level, and the percentage of students who are advanced or proficient combined.

The top section of each Demographic Report identifies the title of the report, the level of aggregation (school, district, state), the student population included in the report, the grade level, and the assessment cycle. School and district names and codes are included as applicable.

The main section of the Demographic Report provides MME results for each demographic subgroup and the total student population. Ethnicity subgroups are defined by federal requirements. Under "Additional Reporting Groups" the Economically Disadvantaged and English Language Learners reporting subgroups

are broken into two categories each: 1.) students who are in the category (Yes) and all students who are who are not (No). No summary scores are provided for subgroups with fewer than ten students. The demographic subgroups reported are:

- gender;
- ethnicity;
- economically disadvantaged (ED);
- English language learners (ELL);
- formerly limited English proficient (FLEP);
- migrant;
- homeless;
- students with accommodations. MME reports accommodation conditions for the following:
  - standard (all);
  - nonstandard (all);
  - standard (ELL only); and
  - nonstandard (ELL only).

More detailed information about reporting is provided in Chapter 4 Administration.

Organizations Involved in MME Testing

Michigan Department of Education (MDE)

Bureau of Assessment and Accountability (BAA)

A primary function of the Bureau of Assessment and Accountability (BAA), formerly, Office of Educational Assessment and Accountability (OEAA), located within the Michigan Department of Education (MDE), is to establish, develop, and conduct a state assessment system that fairly and accurately reflects the state's content standards. These assessments include Michigan Educational Assessment Program (MEAP), MI-Access, MEAP-Access, English Language Proficiency Assessment (ELPA), and the Michigan Merit Examination (MME).

The BAA staff directs and manages the implementation of the statewide assessment programs. In addition to planning, scheduling, and directing all assessment activities, the staff is extensively involved in item reviews, security, and quality control procedures.

The BAA is also responsible for assessment and accountability reporting, including

- the State of Michigan's *Education Yes!*;
- the federal No Child Left Behind Act (NCLB);
- the National Assessment of Educational Progress (NAEP);
- special reports for legislators, educators, and other stakeholders;
- data for MDE programs and other state agencies;
- external research requests;
- the federal Individuals with Disabilities Education Act (IDEA).

Center for Educational Performance and Information (CEPI)

The Center for Educational Performance and Information (CEPI) collects and reports data about Michigan K–12 public schools. CEPI initiatives in data collection and reporting facilitate school districts' compliance with NCLB and the MDE's accreditation plan, *Education Yes!* CEPI is located in the Department of Management and Budget.

State and federal laws require Michigan's K–12 public schools to collect and submit data about students, educational personnel, and individual schools. Districts report their data to CEPI via the Michigan Student Data System (MSDS)—formerly known as Single Record Student Database (SRSD), the School Infrastructure Database (SID) and the Registry of Educational Personnel (REP). These data are used for several purposes, including:

- determining state aid payments;
- calculating adequate yearly progress;
- determining school accreditation;
- generating graduation/dropout rates;
- documenting teacher qualifications;
- measuring what constitutes a "safe" school.

CEPI maintains the State of Michigan's database of school directory information, the Educational Entity Master (EEM). The Report Card data that comes from the EEM includes:

- names of superintendents and principals;
- school and district addresses and telephone numbers;
- e-mail and Web site addresses.

Department of Information Technology (DIT)

Formed in October 2001 by executive order, the Department of Information Technology (DIT) was created to centralize and consolidate the technology resources in Michigan government. DIT's strategic plan outlines five major goals:

- expand Michigan's services to reach anyone at anytime from anywhere;
- transform Michigan services through sharing and collaboration;
- manage technology to provide better service and faster delivery;
- make Michigan a "Great Workplace" and the employer of choice for technology professionals;
- create a statewide community of partnerships.

Staff members from the DIT assist with preparing the School Report Card, including providing a process for reviewing and resolving appeals from elementary, middle, and high schools.

Department of Educational Technology

In March 2006, the MDE published the report *Leading Educational Transformation for Today's Global Society*, which outlines Michigan's educational technology plan. The report calls for leadership at all levels to meet a single goal: preparing Michigan students to become productive citizens in a global society. The report specifies eight objectives with strategies, performance indicators, and action steps to focus current efforts and to utilize available state- level resources.

Objective 4 states: "Every Michigan educator will use data effectively for classroom decision making and school improvement planning through an integrated local and statewide decision support system."

The Center for Educational Performance and Information will leads a collaboration of State of Michigan agencies to plan and implement comprehensive educational data management to meet federal and state reporting requirements and time lines.

Contractors and Subcontractors

The **BAA** has several contractors for the MME test program. The contractors for the MME Spring 2011 test cycle include ACT Inc., Measurement Inc., Pearson Educational Measurement (PEM) and Cheeney Media Concepts Corporation (Cheeney Media).

ACT is responsible for Day 1 (ACT Plus Writing—college readiness) and Day 2 (*WorkKeys*—work skills assessment) materials, administration, and calibration, scaling, and equating, as well as the derivation of MME subject scores using scores from all three assessment components. Measurement Inc. is in charge of MME Day 3 (Michigan components), including administration, Day 3 scoring, and MME Day 3 reporting. PEM, under the direction of the **BAA**, is responsible for test development for Day 3 Michigan components.

Cheeney Media produces the accommodated formats for Day 3 Michigan components, including large print, Braille, reader scripts, audio accommodations (audio DVDs, video DVDS, audio cassettes) and translated formats in Spanish and Arabic. Subcontracted under Cheeney Media, American Printing House for the Blind, Inc. (APH) creates the Braille and enlarged print versions for the MME assessments. APH assists test developers, including state and federal departments of education, with best practices and appropriate accommodations for assessing blind and visually impaired students.

Educators

The purpose of the Michigan Merit Examination is to accurately measure and report student achievement as measured by knowledge of the Michigan high school content standards. Educators who assist in developing and administering the assessments play crucial roles in helping to achieve fair and accurate student results.

The development of the Michigan Merit Examination is a meticulous process involving thousands of Michigan administrators, teachers, and curriculum experts. The Michigan Revised School Code and the State School Aid Act require the establishment of educational standards and the assessment of students' academic achievement. Accordingly, the State Board of Education, with the input of educators throughout Michigan, approved a system of academic standards and a framework within which local school districts could develop and implement curricula.

The MME assessment is based on the state High School Content Expectations (HSCEs). In 2004, the State Board of Education and the Michigan Department of Education embraced the challenge to initiate a "high school redesign" project. Since then, the national call to create more rigorous learning for high school students has become a major priority for state leaders across the country. The Cherry Commission Report (2005) highlighted several goals for Michigan including the development of HSCEs that reflect both rigorous and a relevant curricular focus. Dovetailing with this call to "curricular action" is Michigan's legislative change in high school assessment. The Michigan Merit Examination, based on rigorous high school learning standards, was implemented in 2007 and will be fully aligned with these standards by 2011.

The Michigan Department of Education's Office of Educational Improvement and Innovation (OEII, formerly Office of School Improvement) led the development of grade level content expectations (for grades K-8) and high school content expectations (for grades 9-12). Content area work groups of academicians chaired by nationally known scholars in the respective field, were commissioned to conduct a scholarly review and identify content standards and expectations. These content standards and expectations went through an extensive field and national review and reflect best practices and current research in the teaching and learning of respective field. They not only build from the *Michigan Curriculum Framework Standards and Benchmarks* (1996), the *Career and Employability Skills Content Standards and Benchmarks* (2001), and include the *Michigan State Board of Education's Policy on Learning Expectations for Michigan Students* (2002), but are also closely aligned with national standards and frameworks in the respective subjects. The Michigan State Board of Education approved the English Language Arts and Mathematics High School Content Expectations in April, 2006, the Science High School Content Expectations in December, 2006, and the Social Studies High School Content Expectations in October, 2007. More information can be found at www.michigan.gov/osi.

Please note the 2010–2011 school year was the first year for MME Social Studies assessment to apply the HSCEs. Previously it was based on the Michigan Curriculum Framework. The state HSCEs were first applied to the MME English Language Arts assessment and Mathematics assessment in Spring 2007 and to the MME Science assessment in Spring 2009.

Advisory and Review Committees

The BAA actively seeks the input and feedback via advisory and review committees in the development and implementation of assessment and accountability systems to further the educational goal of improving what students know and can do in relation to the state content standards. Programs that utilize these committees include:

- the MME and MEAP, the assessments for most students in K–12 education programs throughout the State of Michigan;
- the alternate assessments for students with disabilities (MI-Access and MEAP-Access);
- the English Language Proficiency Assessment (ELPA), a screening assessment for students new to this country who have limited English proficiency.

This ensures that Michigan assessments are high quality and gives Michigan educators a valuable professional development opportunity that increases their familiarity with the high school content standards and thereby enhances their teaching effectiveness.

All committees use structured, nationally recognized processes for their work to ensure that the advice provided is clear, well documented, and efficiently obtained. The time and expertise of the committee members are valued. Below are brief descriptions of some of the key committees, their purposes, and characteristics of members.

- **Content Advisory Committees (CACs)** review assessment items and key content decisions to ensure that the content of items and tests measure important elements of the state content standards in each subject and that each item is clearly worded for the students, has one clear "best" answer, and is factually consistent with the most current knowledge in the field. Based on the advice of these committees and other key information, the BAA will accept test items for use, drop items from further consideration, or edit items. Separate Content Advisory Committees are required for each subject tested. Committee members are very familiar with the subject, the related state content standards, benchmarks, and expectations, and hold detailed knowledge of the students being assessed. Some committee members have in-depth content expertise such as mathematicians who serve on the mathematics committee. Child development experts serve on several committees. The

majority of the committee members must be current teacher experts at the high school level and in the subject tested.

- **Bias and Sensitivity Review Committees (BSCs)** review each text selection, item, and writing prompt for fairness, to assure that no group is unfairly advantaged or disadvantaged compared with any other group by any MME content. The committee rejects items it considers inappropriate, suggests revisions to some, and passes on the majority of the items to the next review committee.
- **Standard Setting Committees** are charged with establishing the performance levels used to report student results on the state assessments. For the MME, these committees develop the performance level descriptors (PLDs) and recommend cut scores for the four levels. The standard setting committee is an advisory committee that looks at actual student assessments, item by item. The committee decides the performance level these assessment results represent, using clearly defined performance level descriptions, and discusses the rationale for the decision with the other committee members. The majority of standard setting committee members must be current teacher experts at the grade and subject tested. Other committee members include administrators, curriculum specialists, counselors, parents, and business leaders. Committees represent the geographic and ethnic diversity of the state.
- **Professional Practices Committee** assisted with developing the document, *Professional Assessment and Accountability Practices for Educators*. This document, published in August 2005, presents the expected ethical conduct of educators who administer the assessments. For assessments to yield fair and accurate results, they must be given under the same standardized conditions to all students. *Professional Assessment and Accountability Practices for Educators* is intended to be used by districts and schools in the fair, appropriate, and ethical administration of the assessments.

Technical Advisory Committee

The Technical Advisory Committee (TAC) independently monitors all assessment development and implementation processes, including information gathered in field tests and review of item development. The TAC may make recommendations for revisions in design, administration, scoring, processing, or use in the examination.

The TAC was first established in 1993 to assist the MDE in developing a high school proficiency assessment as a requirement for high school graduation, required by PA 118 of 1991. At that time, the purpose of the TAC was to assist the MDE in implementing provisions of the law. The TAC continues to advise and assist the BAA (formerly OEAA) to ensure the MME assessments are developed in keeping with technical guidelines that meet national standards. The TAC is composed of individuals from Michigan and across the nation who are recognized experts in developing or reviewing high stakes assessment programs.

Michigan State Board of Education

The State Board of Education provides leadership and general supervision over all public education, including adult education and instructional programs in state institutions, with the exception of higher education institutions granting baccalaureate degrees. The State Board of Education serves as the general planning and coordinating body for all public education, including higher education, and advises the legislature concerning the financial requirements of public education.

The State Board of Education established the standards at key checkpoints during Michigan students' academic careers. With the input of educators throughout Michigan, the State Board of Education approved a system of academic standards and a framework within which local school districts could develop and implement curricula. MME assessment results show how Michigan students and schools perform compared

to standards established by the State Board of Education. MME assessments are criterion-referenced assessments, meaning that student performance is measured against a set standard—in this case, the High School Content Expectations—and results are reported relative to that standard. The standards are developed by Michigan educators and approved by the State Board of Education.

Chapter 3: Test Development (ACT, *WorkKeys*, and Michigan Components)

The MME tests consist of three components (i.e., Day 1 ACT Plus Writing, Day 2 *WorkKeys* and Day 3 Michigan Component). The following section deals with the test development process for Day 1 and Day 2 affiliated with ACT. The second half of the chapter documents the test development process for the Day 3 Michigan component, which includes a wide range of related topics (e.g., item development, item review, field testing, post-test field test item review, operational test construction, item bank, and test form construction). Test alignment information is presented in Chapter 5 ("Test Development Analyses").

Test Development for Day 1 and Day 2

The ACT Plus Writing Test

The ACT Test Program is a comprehensive system of data collection, processing, and reporting designed to help high school students develop postsecondary educational plans and to help postsecondary educational institutions meet the needs of their students. One component of the ACT Test Program is the ACT Plus Writing Test, a battery of four multiple-choice tests: English, Mathematics, Reading, and Science, and a Writing Test. The ACT Test Program also includes an interest inventory, and it collects information about students' high school courses and grades, educational and career aspirations, extracurricular activities, and special educational needs. The ACT Plus Writing is taken under standardized conditions.

ACT Test data are used for many purposes. High schools use ACT data in academic advising and counseling, evaluation studies, accreditation documentation, and public relations. Colleges use ACT results for admissions and course placement. States use the ACT Test as part of their statewide assessment systems. Many of the agencies that provide scholarships, loans, and other types of financial assistance to students tie such assistance to students' academic qualifications. Many state and national agencies also use ACT data to identify talented students and award scholarships.

Philosophical Basis for the ACT

Underlying the ACT tests of educational achievement is the belief that students' preparation for college is best assessed by measuring, as directly as possible, the academic skills that they will need to perform college-level academic work. The required academic skills can be assessed most directly by reproducing as faithfully as possible the complexity of college-level work. Therefore, the tests of educational achievement are designed to determine how skillfully students solve problems, grasp implied meanings, draw inferences, evaluate ideas, and make judgments in content areas important to success in college.

Accordingly, the ACT tests of educational achievement are oriented toward the general content areas of college and high school instructional programs. The test questions require students to integrate the knowledge and skills they possess in major curriculum areas with the information provided by the test. Thus, scores on the tests have a direct and obvious relationship to the students' educational achievement in curriculum-related areas and possess a meaning that is readily grasped by students, parents, and educators. Tests of general educational achievement are used in the ACT because, in contrast to other types of tests, they best satisfy the diverse requirements of tests used to facilitate the transition from secondary to postsecondary education. By comparison, measures of examinee knowledge of specific course content (as opposed to curriculum areas) do not readily provide a common baseline for comparing students for the

purposes of admission, placement, or awarding scholarships because high school courses vary extensively. In addition, such tests might not measure students' skills in problem solving and in the integration of knowledge from a variety of courses.

Tests of educational achievement can also be contrasted with tests of academic aptitude. The stimuli and test questions for aptitude tests are often chosen precisely for their dissimilarity to instructional materials, and each test within a battery of aptitude tests is designed to be homogeneous in psychological structure. With such an approach, these tests may not reflect the complexity of college-level work or the interactions among the skills measured. Moreover, because aptitude tests are not directly related to instruction, they may not be as useful as tests of educational achievement for making placement decisions in college.

The advantage of tests of educational achievement over other types of tests for use in the transition from high school to college becomes evident when their use is considered in the context of the educational system. Because tests of educational achievement measure many of the same skills that are taught in high school, the best preparation for tests of educational achievement is high school course work. Long-term learning in school, rather than short-term cramming and coaching, becomes the best form of test preparation. Thus, tests of educational achievement tend to serve as motivators by sending students a clear message that high test scores are not simply a matter of innate ability but reflect a level of achievement that has been earned as a result of hard work.

Because the ACT stresses such general concerns as the complexity of college-level work and the integration of knowledge from a variety of sources, students may be influenced to acquire skills necessary to handle these concerns. In this way, the ACT may serve to aid high schools in developing in their students the higher-order thinking skills that are important for success in college and later life.

The tests of the ACT therefore are designed not only to accurately reflect educational goals that are widely accepted and judged by educators to be important, but also to give educational considerations, rather than statistical and empirical techniques, paramount importance.

Description of the ACT Plus Writing

The ACT Plus Writing contains four multiple-choice tests—English, Mathematics, Reading, and Science—and a Writing Test. These tests are designed to measure skills that are most important for success in postsecondary education and that are acquired in secondary education.

The content specifications describing the knowledge and skills to be measured by the ACT were determined through a detailed analysis of relevant information: First, the curriculum frameworks for grades seven through twelve were obtained for all states in the United States that had published such frameworks. Second, textbooks on state-approved lists for courses in grades seven through twelve were reviewed. Third, educators at the secondary and postsecondary levels were consulted on the importance of the knowledge and skills included in the reviewed frameworks and textbooks.

Because one of the primary purposes of the ACT is to assist in college admission decisions, in addition to taking the steps described above, ACT conducted a detailed survey to ensure the appropriateness of the content of the ACT tests for this particular use. College faculty members across the nation who were familiar with the academic skills required for successful college performance in language arts, mathematics, and science were surveyed. They were asked to rate numerous knowledge and skill areas on the basis of their importance to success in entry-level college courses and to indicate which of these areas students should be expected to master before entering the most common entry-level courses. They were also asked to

identify the knowledge and skills whose mastery would qualify a student for advanced placement. A series of consultant panels were convened, at which the experts reached consensus regarding the important knowledge and skills in English and reading, mathematics, and science, given current and expected curricular trends.

Curriculum study is ongoing at ACT. Curricula in each content area (English, reading, mathematics, science, and writing) in the ACT tests are reviewed on a periodic basis. ACT's analyses include reviews of tests, curriculum guides, and national standards; surveys of current instructional practice; and meetings with content experts (see ACT, *ACT National Curriculum Survey® 2009*, 2009).

The tests in the ACT are designed to be developmentally and conceptually linked to those of EXPLORE (Grades 8 and 9) and PLAN (Grade 10). To reflect that continuity, the names of the content area tests are the same across the three programs. Moreover, the programs are similar in their focus on thinking skills and in their common curriculum base. The test specifications for the ACT are consistent with, and should be seen as a logical extension of the content and skills measured in EXPLORE and PLAN.

The English Test

The ACT English Test is a 75-item, 45-minute test that measures understanding of the conventions of standard written English (punctuation, grammar and usage, and sentence structure) and of rhetorical skills (strategy, organization, and style). Spelling, vocabulary, and rote recall of rules of grammar are not tested. The test consists of five prose passages, each accompanied by a sequence of multiple-choice test items. Different passage types are employed to provide a variety of rhetorical situations. Passages are chosen not only for their appropriateness in assessing writing skills, but also to reflect examinees' interests and experiences. Most items refer to underlined portions of the passage and offer several alternatives to the portion underlined. These items include "NO CHANGE" to the underlined portion in the passage as one of the possible responses. Some items are identified by a number or numbers in a box. These items ask about a section of the passage, or about the passage as a whole. The examinee must decide which choice is most appropriate in the context of the passage, or which choice best answers the question posed.

Three scores are reported for the English Test: a total test score based on all 75 items, a subscore in Usage/Mechanics based on 40 items, and a subscore in Rhetorical Skills based on 35 items.

The Mathematics Test

The ACT Mathematics Test is a 60-item, 60-minute test that is designed to assess the mathematical reasoning skills that examinees across the United States have typically acquired in courses taken up to the beginning of Grade 12. The test presents multiple-choice items that require examinees to use their mathematical reasoning skills to solve practical problems in mathematics. Knowledge of basic formulas and computational skills are assumed as background for the problems, but memorization of complex formulas and extensive computation are not required. The material covered on the test emphasizes the major content areas that are prerequisite to successful performance in entry-level courses in college mathematics. Six content areas are included: pre-algebra, elementary algebra, intermediate algebra, coordinate geometry, plane geometry, and trigonometry.

The items included in the Mathematics Test cover four cognitive levels: knowledge and skills, direct application, understanding concepts, and integrating conceptual understanding. "Knowledge and skills" items require the examinee to use one or more facts, definitions, formulas, or procedures to solve problems that are presented in purely mathematical terms. "Direct application" items require the examinee to use one

or more facts, definitions, formulas, or procedures to solve straightforward problem sets in real-world situations. "Understanding concepts" items test the examinee's depth of understanding of major concepts by requiring reasoning from a concept to reach an inference or a conclusion. "Integrating conceptual understanding" items test the examinee's ability to achieve an integrated understanding of two or more major concepts so as to solve nonroutine problems.

Calculators, although not required, are permitted for use on the Mathematics Test. Almost any four-function, scientific, or graphing calculator may be used on the Mathematics Test. A few restrictions do apply to the calculator used. These restrictions can be found in the current year's *ACT User Handbook* or on ACT's website at **www.act.org**.

Four scores are reported for the Mathematics Test: a total test score based on all 60 items, a subscore in Pre-Algebra/Elementary Algebra based on 24 items, a subscore in Intermediate Algebra/Coordinate Geometry based on 18 items, and a subscore in Plane Geometry/Trigonometry based on 18 items.

The Reading Test

The ACT Reading Test is a 40-item, 35-minute test that measures reading comprehension as a product of skill in referring and reasoning. That is, the test items require examinees to derive meaning from several texts by: (1) referring to what is explicitly stated and (2) reasoning to determine implicit meanings. Specifically, items ask examinees to use referring and reasoning skills to determine main ideas; locate and interpret significant details; understand sequences of events; make comparisons; comprehend cause-effect relationships; determine the meaning of context-dependent words, phrases, and statements; draw generalizations; and analyze the author's or narrator's voice or method. The test comprises four prose passages that are representative of the level and kinds of text commonly encountered in first-year college curricula; passages on topics in the social sciences, the natural sciences, prose fiction, and the humanities are included. Each passage is preceded by a heading that identifies what type of passage it is (e.g., "Prose Fiction"), names the author, and may include a brief note that helps in understanding the passage. Each passage is accompanied by a set of multiple-choice test items. These items focus on the complex of complementary and mutually supportive skills that readers must bring to bear in studying written materials across a range of subject areas. They do not test the rote recall of facts from outside the passage or rules of formal logic, nor do they contain isolated vocabulary questions.

Three scores are reported for the Reading Test: a total test score based on all 40 items, a subscore in Social Studies/Sciences reading skills (based on the 20 items in the social sciences and natural sciences sections of the test), and a subscore in Arts/Literature reading skills (based on the 20 items in the prose fiction and humanities sections of the test).

The Science Test

The ACT Science Test is a 40-item, 35-minute test that measures the interpretation, analysis, evaluation, reasoning, and problem-solving skills required in the natural sciences. The content of the Science Test is drawn from biology, chemistry, physics, and the Earth/space sciences, all of which are represented in the test. Examinees are assumed to have a minimum of two years of introductory science, which ACT's National Curriculum Studies have identified as typically one year of biology and one year of physical science and/or Earth science. Thus, it is expected that examinees have acquired the introductory content of biology, physical science, and Earth science, are familiar with the nature of scientific inquiry, and have been exposed to laboratory investigation.

The test presents seven sets of scientific information, each followed by a number of multiple-choice test items. The scientific information is conveyed in one of three different formats: data representation (graphs, tables, and other schematic forms), research summaries (descriptions of several related experiments), or conflicting viewpoints (expressions of several related hypotheses or views that are inconsistent with one another).

The items included in the Science Test cover three cognitive levels: understanding, analysis, and generalization. "Understanding" items require examinees to recognize and understand the basic features of, and concepts related to, the provided information. "Analysis" items require examinees to examine critically the relationships between the information provided and the conclusions drawn or hypotheses developed. "Generalization" items require examinees to generalize from given information to gain new information, draw conclusions, or make predictions.

One score is reported for the Science Test: a total test score based on all 40 items.

The Writing Test

The ACT Writing Test is a 30-minute essay test that measures examinees' writing skills—specifically those writing skills emphasized in high school English classes and in entry-level college composition courses. The test consists of one writing prompt that defines an issue and describes two points of view on that issue. The examinee is asked to respond to a question about his/her position on the issue described in the writing prompt. In doing so, the examinee may adopt one or the other of the perspectives described in the prompt, or they may present a different point of view on the issue. The essay score is not affected by the point of view taken on the issue.

Taking the Writing Test does **not** affect an examinee's score on the multiple-choice tests or the Composite score for those tests. Rather, two additional scores are provided: a Combined English/Writing score and a Writing subscore. Also provided are comments on the examinee's essay.

Test Development Procedures for the ACT Multiple-Choice Tests

This section describes the procedures that are used in developing the four multiple-choice tests described above. The test development cycle required to produce each new form of the ACT tests takes as long as two and one-half years and involves several stages, beginning with a review of the test specifications.

Reviewing Test Specifications

Two types of test specifications are used in developing the ACT tests: content specifications and statistical specifications.

Content specifications

Content specifications for the ACT tests were developed through the curricular analysis discussed above. While care is taken to ensure that the basic structure of the ACT tests remains the same from year to year so that the scale scores are comparable, the specific characteristics of the test items used in each specification category are reviewed regularly. Consultant panels are convened to review both the tryout versions and the new forms of each test to verify their content accuracy and the match of the content of the tests to the content specifications. At these panels, the characteristics of the items that fulfill the content specifications are also reviewed. While the general content of the test remains constant, the particular kinds of items in a

specification category may change slightly. The basic structure of the content specifications for each of the ACT multiple-choice tests is provided in Tables 3.1 through 3.4.

Statistical specifications

Statistical specifications for the tests indicate the level of difficulty (proportion correct) and minimum acceptable level of discrimination (biserial correlation) of the test items to be used.

The tests are constructed with a target mean item difficulty of about 0.58 for the ACT population and a range of difficulties from about 0.20 to 0.89. The distribution of item difficulties was selected so that the tests will effectively differentiate among examinees who vary widely in their level of achievement.

With respect to discrimination indices, items should have a biserial correlation of 0.20 or higher with test scores measuring comparable content. Thus, for example, performance on mathematics items should correlate 0.20 or higher with performance on the relevant Mathematics Test subscore.

Six elements of effective writing are included in the English Test. These elements and the approximate proportion of the test devoted to each are given in the table below.

Table 3.1. Content Specifications for the ACT English Test

Content/Skills	Proportion of test		Number of items	
Usage/Mechanics	0.53		40	
Punctuation[a]		0.13		10
Grammar and Usage[b]		0.16		12
Sentence Structure[c]		0.24		18
Rhetorical Skills	0.47		35	
Strategy[d]		0.16		12
Organization[e]		0.15		11
Style[f]		0.16		12
Total	**1.00**		**75**	

Scores reported: Usage/Mechanics
Rhetorical Skills
Total test score

[a]*Punctuation.* The items in this category test the student's knowledge of the conventions of internal and end-of-sentence punctuation, with emphasis on the relationship of punctuation to meaning (for example, avoiding ambiguity, indicating appositives).

[b]*Grammar and Usage.* The items in this category test the student's understanding of agreement between subject and verb, between pronoun and antecedent, and between modifiers and the words modified; verb formation; pronoun case; formation of comparative and superlative adjectives and adverbs; and idiomatic usage.

[c]*Sentence Structure.* The items in this category test the student's understanding of relationships between and among clauses, placement of modifiers, and shifts in construction.

[d]*Strategy.* The items in this category test the student's ability to develop a given topic by choosing expressions appropriate to an essay's audience and purpose; to judge the effect of adding, revising, or deleting supporting material; and to judge the relevancy of statements in context.

[e]*Organization.* The items in this category test the student's ability to organize ideas and to choose effective opening, transitional, and closing sentences.
[f]*Style.* The items in this category test the student's ability to select precise and appropriate words and images, to maintain the level of style and tone in an essay, to manage sentence elements for rhetorical effectiveness, and to avoid ambiguous pronoun references, wordiness, and redundancy.

The items in the Mathematics Test are classified with respect to six content areas. These areas and the approximate proportion of the test devoted to each are given below.

Table 3.2. Content Specifications for the ACT Mathematics Test

Content Area	Proportion of test	Number of items
Pre-Algebra[a]	0.23	14
Elementary Algebra[b]	0.17	10
Intermediate Algebra[c]	0.15	9
Coordinate Geometry[d]	0.15	9
Plane Geometry[e]	0.23	14
Trigonometry [f]	0.07	4
Total	**1.00**	**60**

Scores reported: Pre-Algebra/Elementary Algebra
Intermediate Algebra/Coordinate Geometry
Plane Geometry/Trigonometry
Total test score

[a]*Pre-Algebra.* Items in this content area are based on operations using whole numbers, decimals, fractions, and integers; place value; square roots and approximations; the concept of exponents; scientific notation; factors; ratio, proportion, and percent; linear equations in one variable; absolute value and ordering numbers by value; elementary counting techniques and simple probability; data collection, representation, and interpretation; and understanding simple descriptive statistics.

[b]*Elementary Algebra.* Items in this content area are based on properties of exponents and square roots, evaluation of algebraic expressions through substitution, using variables to express functional relationships, understanding algebraic operations, and the solution of quadratic equations by factoring.

[c]*Intermediate Algebra.* Items in this content area are based on an understanding of the quadratic formula, rational and radical expressions, absolute value equations and inequalities, sequences and patterns, systems of equations, quadratic inequalities, functions, modeling, matrices, roots of polynomials, and complex numbers.

[d]*Coordinate Geometry.* Items in this content area are based on graphing and the relations between equations and graphs, including points, lines, polynomials, circles, and other curves; graphing inequalities; slope; parallel and perpendicular lines; distance; midpoints; and conics.

[e]*Plane Geometry.* Items in this content area are based on the properties and relations of plane figures, including angles and relations among perpendicular and parallel lines; properties of circles, triangles, rectangles, parallelograms, and trapezoids; transformations; the concept of proof and proof techniques; volume; and applications of geometry to three dimensions.

[f]*Trigonometry.* Items in this content area are based on understanding trigonometric relations in right triangles; values and properties of trigonometric functions; graphing trigonometric functions; modeling using trigonometric functions; use of trigonometric identities; and solving trigonometric equations.

The items in the Reading Test are based on the prose passages that are representative of the kinds of writing commonly encountered in college freshman curricula, including prose fiction, the social sciences, the humanities, and the natural sciences. The four content areas and the approximate proportion of the test devoted to each are given below.

Table 3.3. Content Specifications for the ACT Reading Test

Reading passage content	Proportion of test	Number of items
Prose Fiction[a]	0.25	10
Social Science[b]	0.25	10
Humanities[c]	0.25	10
Natural Science[d]	0.25	10
Total	**1.00**	**40**

Scores reported: Social Studies/Sciences (Social Science, Natural Science)
Arts/Literature (Prose Fiction, Humanities)
Total test score

[a]*Prose Fiction*. The items in this category are based on short stories or excerpts from short stories or novels.

[b]*Social Science*. The items in this category are based on passages in the content areas of anthropology, archaeology, biography, business, economics, education, geography, history, political science, psychology, and sociology.

[c]*Humanities*. The items in this category are based on passages from memoirs and personal essays and in the content areas of architecture, art, dance, ethics, film, language, literary criticism, music, philosophy, radio, television, and theater.

[d]*Natural Science*. The items in this category are based on passages in the content areas of anatomy, astronomy, biology, botany, chemistry, ecology, geology, medicine, meteorology, microbiology, natural history, physiology, physics, technology, and zoology.

The Science Test is based on the type of content that is typically covered in high school science courses. Materials are drawn from the biological sciences, the Earth/space sciences, physics, and chemistry. The test emphasizes scientific reasoning skills rather than recall of specific scientific content, skill in mathematics, or skill in reading. Minimal arithmetic and algebraic computations may be required to answer some items. The three formats and the approximate proportion of the test devoted to each are given below.

Table 3.4. Content Specifications for the ACT Science Test

Content area[a]	Format	Proportion of test	Number of items
Biology Earth/Space Sciences Physics Chemistry	Data Representation[b]	0.38	15
	Research Summaries[c]	0.45	18
	Conflicting Viewpoints[d]	0.17	7
Total		**1.00**	**40**

Score reported: Total test score

[a]All four content areas are represented in the test. The content areas are distributed over the different formats in such a way that at least one passage, and no more than two passages, represents each content area.

[b]*Data Representation.* This format presents students with graphic and tabular material similar to that found in science journals and texts. The items associated with this format measure skills such as graph reading, interpretation of scatter plots, and interpretation of information presented in tables, diagrams, and figures.

[c]*Research Summaries.* This format provides students with descriptions of one or more related experiments. The items focus on the design of experiments and the interpretation of experimental results.

[d]*Conflicting Viewpoints.* This format presents students with expressions of several hypotheses or views that, being based on differing premises or on incomplete data, are inconsistent with one another. The items focus on the understanding, analysis, and comparison of alternative viewpoints or hypotheses.

Selection of Item Writers

Each year, ACT contracts with item writers to construct items for the ACT. The item writers are content specialists in the disciplines measured by the ACT tests. Most are actively engaged in teaching at various levels, from high school to university, and at a variety of institutions, from small private schools to large public institutions. ACT makes every attempt to include item writers who represent the diversity of the population of the United States with respect to ethnic background, gender, and geographic location.

Before being asked to write items for the ACT tests, potential item writers are required to submit a sample set of materials for review. Each item writer receives an item writer's guide that is specific to the content area. The guides include examples of items and provide item writers with the test specifications and ACT's requirements for content and style. Included are specifications for fair portrayal of all groups of individuals, avoidance of subject matter that may be unfamiliar to members of certain groups within society, and nonsexist use of language.

Each sample set submitted by a potential item writer is evaluated by ACT Test Development staff. A decision concerning whether to contract with the item writer is made on the basis of that evaluation.

Every item writer under contract is given an assignment to produce a small number of multiple-choice items. The small size of the assignment ensures production of a diversity of material and maintenance of the security of the testing program, since any item writer will know only a small proportion of the items produced. Item writers work closely with ACT test specialists, who assist them in producing items of high quality that meet the test specifications.

Item Construction

The item writers must create items that are educationally important and psychometrically sound. A large number of items must be constructed because, even with good writers, many items fail to meet ACT's standards.

Each item writer submits a set of items, called a *unit*, in a given content area. Most Mathematics Test items are discrete (not passage-based), but occasionally some may belong to sets composed of several items based on the same paragraph or chart. All items on the English and Reading Tests are related to prose passages. All items on the Science Test are related to passages and/or other stimulus material (such as graphs and tables).

Review of Items

After a unit is accepted, it is edited to meet ACT's specifications for content accuracy, word count, item classification, item format, and language. During the editing process, all test materials are reviewed for fair portrayal and balanced representation of groups within society and for nonsexist use of language. The unit is reviewed several times by ACT staff to ensure that it meets all of ACT's standards.

Copies of each unit are then submitted to content and fairness experts for external reviews prior to the pretest administration of these units. The content review panel consists of high school teachers, curriculum specialists, and college and university faculty members. The content panel reviews the unit for content accuracy, educational importance, and grade-level appropriateness. The fairness review panel consists of experts in diverse educational areas who represent both genders and a variety of racial and ethnic

backgrounds. The fairness panel reviews the unit to help ensure fairness to all examinees. Any comments on the units by the content consultants are discussed in a panel meeting with all the content consultants and ACT staff, and appropriate changes are made to the unit(s). All fairness consultants' comments are reviewed and discussed, and appropriate changes are made to the unit(s).

Item Tryouts

The items that are judged to be acceptable in the review process are assembled into tryout units for pretesting on samples from the national examinee population. These samples are carefully selected to be representative of the total examinee population. Each sample is administered a tryout unit from one of the four academic areas covered by the ACT tests. The time limits for the tryout units permit the majority of students to respond to all items.

Item Analysis of Tryout Units

Item analyses are performed on the tryout units. For a given unit the sample is divided into low-, medium-, and high-performing groups by the individuals' scores on the ACT test in the same content area (taken at the same time as the tryout unit). The cutoff scores for the three groups are the 27th and the 73rd percentile points in the distribution of those scores. These percentile points maximize the critical ratio of the difference between the mean scores of the upper and lower groups, assuming that the standard error of measurement in each group is the same and that the scores for the entire examinee population are normally distributed (Millman & Greene, 1989).

Proportions of examinees in each of the groups correctly answering each tryout item are tabulated, as well as the proportion in each group selecting each of the incorrect options. Biserial and point-biserial correlation coefficients between each item score (correct/incorrect) and the total score on the corresponding test of the regular (national) test form are also computed.

Item analyses serve to identify statistically effective test items. Items that are either too difficult or too easy, and items that fail to discriminate between examinees of high and low educational achievement as measured by their corresponding ACT test scores, are eliminated or revised for future item tryouts. The biserial and point-biserial correlation coefficients, as well as the differences between proportions of examinees answering the item correctly in each of the three groups, are used as indices of the discriminating power of the tryout items.

Each item is reviewed following the item analysis. ACT staff members scrutinize items flagged for statistical reasons to identify possible problems. Some items are revised and placed in new tryout units following further review. The review process also provides feedback that helps decrease the incidence of poor quality items in the future.

Assembly of New Forms

Items that are judged acceptable in the review process are placed in an item pool. Preliminary forms of the ACT tests are constructed by selecting from this pool items that match the content and statistical specifications for the tests.

For each test in the battery, items for the new forms are selected to match the content distribution for the tests shown in Tables 3.1 through 3.4. Items are also selected to comply with the statistical specifications described in a previous section of this chapter. The distributions of item difficulty levels obtained on recent

forms of the four tests are displayed in Table 3.5. The data in Table 3.5 are taken from random samples of approximately 2,000 examinees from each of the six national test dates during the 2010–2011 academic year. In addition to the item difficulty distributions, item discrimination indices in the form of observed mean biserial correlations and completion rates are reported.

Table 3.5. Difficulty[a] Distributions and Mean Discrimination[b] Indices for ACT Test Items, 2010–2011

	Observed difficulty distributions (frequencies)			
	English	**Mathematics**	**Reading**	**Science**
Difficulty range				
0.00–0.09	0	0	0	0
0.10–0.19	0	9	0	0
0.20–0.29	3	38	8	18
0.30–0.39	22	45	13	33
0.40–0.49	45	59	42	42
0.50–0.59	81	53	45	42
0.60–0.69	100	63	51	42
0.70–0.79	83	50	60	42
0.80–0.89	102	38	20	20
0.90–1.00	14	5	1	1
Number of items[c]	450	360	240	240
Mean difficulty	0.66	0.54	0.61	0.55
Mean discrimination	0.58	0.62	0.59	0.51
Avg. completion rate[d]	92	91	93	94

[a]Difficulty is the proportion of examinees correctly answering the item.

[b]Discrimination is the item-total score biserial correlation coefficient.

[c]Six forms consisting of the following number of items per test: English 75, Mathematics 60, Reading 40, Science 40.

[d]Mean proportion of examinees who answered each of the last five items.

The completion rate is an indication of how "speeded" a test is for a group of examinees. A test is considered to be speeded if most examinees do not have sufficient time to answer the items in the time allotted. The completion rate reported in Table 3.5 for each test is the average completion rate for the six national test dates during 2010–2011 academic year. The completion rate for each test is computed as the average proportion of examinees who answered each of the last five items.

Content and Fairness Review of Test Forms

The preliminary versions of the test forms are subjected to several reviews to ensure that the items are accurate and that the overall test forms are fair and conform to good test construction practice. The first review is performed by ACT staff. Items are checked for content accuracy and conformity to ACT style. The items are also reviewed to ensure that they are free of clues that could allow testwise examinees to answer the item correctly even though they lack knowledge in the subject areas or the required skills.

The preliminary versions of the test forms are then submitted to content and fairness experts for external review before the operational administration of the test forms. These experts are different individuals from those consulted for the content and fairness reviews of tryout units.

Two panels, a content review panel and a fairness review panel, are then convened to discuss with ACT staff the consultants' reviews of the forms. The content review panel consists of high school teachers, curriculum specialists, and college and university faculty members. The content panel reviews the forms for content accuracy, educational importance, and grade-level appropriateness. The fairness review panel consists of experts in diverse areas of education who represent both genders and a variety of racial and ethnic backgrounds. The fairness panel reviews the forms to help ensure fairness to all examinees.

After the panels complete their reviews, ACT summarizes the results. All comments from the consultants are reviewed by ACT staff members, and appropriate changes are made to the test forms. Whenever significant changes are made, the revised components are again reviewed by the appropriate consultants and by ACT staff. If no further corrections are needed, the test forms are prepared for printing.

In all, at least sixteen independent reviews are made of each test item before it appears on a national form of the ACT. The many reviews are performed to help ensure that each examinee's level of achievement is accurately and fairly evaluated.

Review Following Operational Administration

After each operational administration, item analysis results are reviewed for any anomalies such as substantial changes in item difficulty and discrimination indices between tryout and national administrations. Only after all anomalies have been thoroughly checked and the final scoring key approved are score reports produced. Examinees may challenge any items that they feel are questionable. Once a challenge to an item is raised and reported, the item is reviewed by content specialists in the content area assessed by the item. In the event that a problem is found with an item, actions are taken to eliminate or minimize the influence of the problem item as necessary. In all cases, the person who challenges an item is sent a letter indicating the results of the review.

Also, after each operational administration, DIF (differential item functioning) analysis procedures are conducted on the test data. DIF can be described as a statistical difference between the probability of the specific population group (the "focal" group) getting the item right and the comparison population group (the "base" group) getting the item right given that both groups have the same level of achievement with respect to the content being tested. The procedures currently used for the analysis include the standardized difference in proportion-correct (STD) procedure and the Mantel-Haenszel common odds-ratio (MH) procedure.

Both the STD and MH techniques are designed for use with multiple-choice items, and both require data from significant numbers of examinees to provide reliable results. For a description of these statistics and their performance overall in detecting DIF, see the ACT Research Report entitled *Performance of Three Conditional DIF Statistics in Detecting Differential Item Functioning on Simulated Tests* (Spray, 1989). In the analysis of items in an ACT form, large samples representing examinee groups of interest (e.g., males and females) are selected from the total number of examinees taking the test. The examinees' responses to each item on the test are analyzed using the STD and MH procedures. Compared with preestablished criteria, the items with STD or MH values exceeding the tolerance level are flagged. The flagged items are then further reviewed by the content specialists for possible explanations of the unusual STD or MH results. In the event that a problem is found with an item, actions will be taken as necessary to eliminate or minimize the influence of the problem item.

Test Development Procedures for the ACT Writing Test

This section describes the procedures that are used in developing essay prompts for the ACT Writing Test. These include many of the same stages as those used to develop the multiple-choice tests.

Selection and Training of Prompt Writers

ACT holds a prompt writing workshop each year in which new essay prompts are developed. The participants invited to take part in this prompt development process are both high school and postsecondary teachers who are specialists in writing, and who represent the diversity of the U.S. population in ethnic background, gender, and geographic location.

Prompt Construction

Prompts developed for the Writing Test provide topics that not only offer adequate complexity and depth so that examinees can write a thoughtful and engaging essay, but also are within the common experiences of high school students. Topics are carefully chosen so that they are neither too vast nor simplistic, and so that they do not require specialized prior knowledge. The topics are designed so that an examinee should be able to respond to a topic within the 30-minute time constraint of the test.

Content and Fairness Review of Prompts

After Writing Test prompts are developed and then refined by ACT writing specialists, the prompts go through a rigorous review process by external experts. These fairness and bias experts carefully review each prompt to ensure that neither the language nor the content of a prompt will be offensive to an examinee, and that no prompt will disadvantage any examinee from any geographic, socioeconomic, or cultural background.

Field Testing of Prompts

New Writing Test prompts are field-tested throughout the United States every year. Students from rural and urban settings, small and large schools, and both public and private schools write responses to the new prompts, which are then read and scored by trained ACT readers.

Review of Field Tests and Operational Administration

Once scoring of the new Writing Test prompts has been completed, the prompts are analyzed for acceptability, validity, and accessibility. The new field-tested prompts are also reviewed to ensure that they are compatible with previous operational prompts, that they function in the same way as previous prompts, and that they adhere to ACT's rigorous standards.

ACT Scoring Procedures

For each of the four multiple-choice tests in the ACT (English, Mathematics, Reading, and Science), the raw scores (number of correct responses) are converted to scale scores ranging from 1 to 36.

The Composite score is the average of the four scale scores rounded to the nearest whole number (fractions of 0.5 or greater round up). The minimum Composite score is 1; the maximum is 36.

In addition to the four ACT test scores and Composite score, seven subscores are reported: two each for the English Test and the Reading Test and three for the Mathematics Test. As is done for each of the four tests, the raw scores for the subscore items are converted to scale scores. These subscores are reported on a score scale ranging from 1 to 18. The four test scores and seven subscores are derived independently of one another. The subscores in a content area do not necessarily add to the test score in that area.

In addition to the above scores, if the examinee took the Writing Test, the examinee's essay is read and scored independently by two trained readers using a six-point scoring rubric. Essays are evaluated on the evidence they demonstrate of examinee ability to make and articulate judgments; develop and sustain a position on an issue; organize and present ideas in a logical way; and communicate clearly and effectively using the conventions of standard written English. Essays are scored holistically—that is, on the basis of the overall impression created by all the elements of the writing. Each reader rates an essay on a scale ranging from 1 to 6. The sum of the readers' ratings is an examinee's Writing Test subscore on a scale ranging from 2 to 12. An examinee who takes the Writing Test also receives a Combined English/Writing score on a score scale ranging from 1 to 36. Writing Test results do not affect an examinee's Composite score.

Electronic scanning devices are used to score the four multiple-choice tests of the ACT, thus minimizing the potential for scoring errors. If an examinee believes that a scoring error has been made, ACT hand-scores the answer document (for a fee) upon receipt of a written request from the examinee. An examinee may arrange to be present for hand-scoring by contacting one of ACT's regional offices, but must pay whatever extra costs may be incurred in providing this special service. Strict confidentiality of each examinee's record is maintained.

If an examinee believes that a Writing Test essay has been incorrectly scored, that score may be appealed, and the essay will be reviewed and rescored (for a fee) by two new expert readers. The two new readers score the appealed essay without knowledge of the original score, and the new score is adjudicated by ACT staff writing specialists before being finalized.

For certain test dates (specified in the current year's booklet *Registering for the ACT* and on ACT's website at **www.act.org**, examinees may obtain (upon payment of an additional fee) a copy of the test items used in determining their scores, the correct answers, a list of their answers, and a table to convert raw scores to the reported scale scores. For an additional fee, an examinee may also obtain a copy of his or her answer document. These materials are available only to examinees who test during regular administrations of the ACT on specified national test dates. If for any reason ACT must replace the test form scheduled for use at a test center, this offer is withdrawn and the examinee's fee for this optional service is refunded.

ACT reserves the right to cancel test scores when there is reason to believe the scores are invalid. Cases of irregularities in the test administration process—falsifying one's identity, impersonating another examinee (surrogate testing), unusual similarities in answers of examinees at the same test center, or other indicators that the test scores may not accurately reflect the examinee's level of educational achievement, including but not limited to examinee misconduct—may result in ACT's canceling the test scores. When ACT plans to cancel an examinee's test scores, it always notifies the examinee prior to taking this action. This notification includes information about the options available regarding the planned score cancellation, including procedures for appealing this decision. In all instances, the final and exclusive remedy available to examinees who want to appeal or otherwise challenge a decision by ACT to cancel their test scores is binding arbitration through written submissions to the American Arbitration Association. The issue for arbitration shall be whether ACT acted reasonably and in good faith in deciding to cancel the scores.

Technical Characteristics of the ACT Tests

ACT has conducted extensive analyses on the technical characteristics in the ACT – the score scale, norms, equating, and reliability of the tests. A carefully selected sample of examinees from one of the six national test dates each year is used as an equating sample. Scores on the alternate forms are equated to the score scale using equipercentile equating methodology. Summary statistics, based on the six national ACT administrations in 2005–2006, for scale score reliability coefficients and average standard errors of measurement for the ACT tests and subscores are given in Table 3.6. The technical characteristics of the ACT test are thoroughly documented in the ACT Technical Manual (ACT, 2007b). The ACT Technical Manual can be acquired from ACT's website at **www.act.org**.

Table 3.6. Scale Score Reliability and Average Standard Error of Measurement Summary Statistics for the Six National ACT Administrations in 2005–2006

	Scale score reliability			**Average SEM**		
Test/Subtest	**Median**	**Minimum**	**Maximum**	**Median**	**Minimum**	**Maximum**
English	.91	.89	.91	1.71	1.65	1.79
Usage/Mechanics	.86	.84	.88	1.36	1.25	1.39
Rhetorical Skills	.84	.81	.85	1.19	1.14	1.25
Mathematics	.91	.89	.92	1.47	1.43	1.56
Pre-Algebra/Elementary Algebra	.82	.81	.83	1.37	1.30	1.44
Intermediate Algebra/Coordinate Geometry	.72	.70	.75	1.47	1.38	1.54
Plane Geometry/Trigonometry	.74	.69	.78	1.52	1.34	1.66
Reading	.85	.85	.87	2.18	2.11	2.26
Social Studies/Sciences	.75	.73	.77	1.65	1.57	1.73
Arts/Literature	.77	.76	.78	1.75	1.67	1.89
Science	.80	.74	.83	2.00	1.90	2.12
Composite	.96	.95	.96	0.94	0.91	0.96

The *WorkKeys* Assessments Components: Reading for Information, Applied Mathematics, and Locating Information

In recent years, members of the business community as well as the general public have indicated concern that American workers, both current and future, lack the workplace skills needed to meet the challenges of rapidly evolving technical advances, organizational restructuring, and global economic competition. New jobs often require workers coming from high schools or postsecondary programs to have strong problem-solving and communication skills. Current trends in basic skill deficiencies indicate that American businesses will soon be spending more than $25 billion a year on remedial training programs for new employees.

ACT designed *WorkKeys* to address this problem. The system serves businesses, workers, educators, and learners. As part of the development process, ACT listened to employers, educators, and experts in employment and training requirements to find out which employability skills are crucial in most jobs. Based

on their insights, ACT developed the first nine *WorkKeys* skill areas: *Applied Technology*, *Applied Mathematics*, *Business Writing*, *Listening*, *Locating Information*, *Workplace Observation*, *Reading for Information*, *Teamwork*, and *Writing*. Personal skills assessments are also available.

Each skill area has its own skill scale that measures both the skill requirements of specified jobs and the employability skills of individuals. Before *WorkKeys*, there were no scales that could measure both the skills a person has and the skills a job needs. Each *WorkKeys* skill scale describes a set of skill levels. This makes it possible to determine the proficiency levels students and workers already have and to design job-training programs that can help them meet the demands of the jobs they want. The *WorkKeys* system is based on the assumption that people who want to improve their skills can do so if they have enough time and appropriate instruction. Showing a direct connection between job requirements and education and training has a positive effect on learner persistence and achievement.

The WorkKeys Assessment Development Process

WorkKeys assessments are designed to cover a range of skills that is not too narrow and not too wide. If too narrow, a huge battery of tests would be needed to measure skills accurately; and if too wide, the number of items needed for validation would make the assessment too long and time-consuming. Thus, the *WorkKeys* assessments are designed to meet the following criteria:

- The way a skill is assessed is generally congruent with the way the skill is used in the workplace.
- The lowest level assessed is at approximately the lowest level for which an employer would be interested in setting a standard.
- The highest level assessed is at approximately the level beyond which specialized training would be required.
- The steps between the lowest and highest levels are large enough to be distinguished and small enough to have practical value in documenting workplace skills.
- The assessments are sufficiently reliable for high-stakes decision making.
- The assessments can be validated against empirical criteria.
- The assessments are feasible with respect to cost, administration time, and complexity.

The development process for a *WorkKeys* assessment consists of five phases: skill definition, test specifications development, prototyping, pretesting, and construction of operational forms. The process used to develop the *WorkKeys* multiple-choice test items is similar to that used for many standardized assessments including others developed by ACT (Anastasi, 1982; Crocker & Algina, 1986). Both stimuli and response alternatives meet basic requirements associated with high-quality skills.

<u>Skill Definition</u>

Before constructing the *WorkKeys* assessments, ACT defines the content domains and develops hierarchical *WorkKeys* skill descriptions. This process typically begins with a panel made up of employers, educators, and ACT staff. The panel first develops a broad definition of a skill area and identifies the lowest and highest level of the skill that is worthwhile to measure. The panel then identifies examples of tasks within this broadly defined skill domain and narrows that domain to those examples that are important for job performance across a wide range of jobs. Next, the tasks are organized into “strands,” which are aspects of the general skill domain, or skill area that pertain to a singular concept to be measured. The strands assessed in *Reading for Information*, for example, include “choosing main ideas or details,” “understanding word meanings,” “applying instructions,” and “applying information and reasoning.”

The strands are also divided into levels based on the variables believed to cause a task to be more or less difficult. In general, at the low end of a strand a few simple things must be attended to, whereas at the high end, many things must be attended to and a person must process information to apply it to more complex situations. In the "applying instructions" strand of *Reading for Information*, for example, employees need only apply instructions to clearly described situations at the lower levels. At the higher levels, however, employees must not only understand instructions in which the wording is more complex, meanings are more subtle, and multiple steps and conditionals are involved, but must also apply these instructions to new situations.

Test Specifications

Using the skill definitions described above, the ACT *WorkKeys* development team refines the specifications, outlining in more detail the skills the assessment will measure and how the items will become more complex as the skill levels increase. Each level is defined in terms of its characteristics, and exemplar test items are created to illustrate it. While it is sometimes appropriate to assign content to a unique level, in most cases the complexity of the stimulus and question determines the level to which a particular test item is assigned.

WorkKeys test specifications for the multiple-choice assessments are unlike the test blueprints used in education. They are not a list of the content topics or objectives to be covered and the number of test items to be assigned to each. Rather, they are more like scoring rubrics used for holistic scoring of constructed-response assessments (White, E. M., 1994). Similarly, the alternatives for a single multiple-choice question may include multiple content classifications, modeling a well-integrated curriculum, yet making the typical approach to test blueprints, which assume that each item measures only one objective, inappropriate.

Prototyping

After development of the general test specifications, ACT test development associates (TDAs) begin writing items for the prototype test. All the items must be written to meet the test specifications and must correspond to the respective skill levels of the test. A number of prototype test items sufficient to create long test form (75 items for RFI and AM, and 50 items for LI) for the skill area are produced.

Each prototype test form (one per skill area) is administered to at least two groups of high school students and two groups of employees. Typically, one group of students and one of employees will be from the same city. The second groups of students and employees will be found in another state with a different situation (for example, if the first groups are from a suburban setting, the second may be from an inner city). The number of examinees varies according to the test format, with more being used for multiple-choice tests than for constructed-response tests. Typically, at least 200 students and 60 employees are divided across the two administration sites for each multiple-choice prototype test form.

During the prototype process, TDAs interview the examinees to gather their reactions to the test instrument, which helps ACT evaluate the functioning of the test specifications. Questions such as whether the prototype items were too hard, too easy, or tested skills outside the realm of the specifications must be answered before development can move to the pretesting stage. The examinees are asked to provide comments and suggestions about the prototype test form, and educators and employers are also invited to review and comment on it. Based on all the information from prototype testing, the test specifications are adjusted if necessary, and additional prototype studies may be conducted. When the prototype process is completed satisfactorily, a written guide for item writers is prepared.

Pretesting

For the pretesting phase, ACT contracts with numerous freelance item writers who produce a large number of items, which ACT staff edit to meet the content, cognitive, and format standards. *WorkKeys* item writers must be familiar with various work situations and have insight into the use of a particular skill in different employment settings because both content and contextual accuracy are critically important for *WorkKeys*. A test question containing inaccurate content may be distracting even if the specific content does not affect the examinee's ability to respond correctly to the skills portion of the question. Inaccurate facts, improbable circumstances, or unlikely consequences of a series of procedures or actions are not acceptable. An examinee who knows about a particular workplace should not identify any of the assessment content, circumstances, procedures, or keyed responses as unlikely, inappropriate, or otherwise inaccurate.

Given the wide range of employability skills assessed, verifying content accuracy for *WorkKeys* is challenging. To help *WorkKeys* staff detect any possible problems, the item writers write a justification for the best response and for each distractor (incorrect response) for each test item. Both the items and the justifications are checked and, if necessary, the test items are modified.

After the test questions and stimuli have been created and edited, and before administration of the pretesting forms, all items are submitted to external consultants for content and fairness reviews. Qualified experts in the specific skill area being assessed, usually persons using the skills regularly on the job, check for content and contextual accuracy. Members of minority groups review the items to make sure they will not be biased against, or offensive to, racial, ethnic, and gender groups. ACT provides all the reviewers with written guidelines and receives written evaluations back from them.

To provide the data required for both classical and item response theory (IRT)–based statistics, each multiple-choice item is administered to a sample of about 2,000 examinees. For practical reasons, most of these examinees are students, although smaller samples of employees are also assessed for each pretest. Then ACT researchers evaluate the psychometric properties (such as reliability and scalability) of each item.

Additionally, statistical, differential item functioning (DIF) analyses of the items are carried out to determine whether items function differently for various groups of individuals (by seeing if responses to items can be correlated with the gender or ethnicity of the examinees). Items that show DIF are eliminated from the item pool. Based on the data collected during pretesting for each skill area, no items in the *WorkKeys* tests show DIF. Statistical studies can also locate problem items, which are identified during the analysis and are reevaluated by staff and, if necessary, outside experts.

Operational Forms

Pretest item analyses are considered carefully when constructing the forms for operational testing. Alternate and equivalent test forms for each assessment are developed from the pool of items that meet all the content, statistical, and fairness criteria. ACT staff construct at least two equivalent test forms for each assessment. In these forms, both the overall characteristics of the test and the within-level characteristics for content, complexity, and psychometric characteristics are made as similar as possible.

In addition to developing the job-profiling procedure to link the content of the *WorkKeys* assessments to a specific job, ACT achieves validity through creating well-designed tests. During the development of the assessments, ACT works to minimize the likelihood of adverse impact resulting from use of the *WorkKeys* tests. Specifically, the assessments are designed to be job-related and fair by ensuring that the items go through a series of screens before they are made available to employers:

- The assessments are criterion-referenced (they use job requirements as the scoring reference, rather than population norms);
- The test specifications are well-defined;
- Items are written by people with employment experience in the workplace and thus the items tap a domain of workplace skill;
- Items measure a particular workplace skill;
- Content and fairness experts review the items to determine possible differences in responses among racial groups and gender; and
- Statistical analyses (for example, differential item functioning) at the item and test level are conducted to monitor the performance of various subgroups.

WorkKeys Assessment Descriptions

Applied Mathematics

The *Applied Mathematics* skill involves the application of mathematical reasoning to work-related problems. The assessment requires the examinee to set up and solve the types of problems and do the types of calculations that actually occur in the workplace. This assessment is designed to be taken with a calculator. As on the job, the calculator serves as a tool for problem solving. A formula sheet that includes, but is not limited to, all formulas required for the assessment is provided. There are five skill levels, with Level 7 requiring the most complex and Level 3 requiring the least complex mathematical concepts and calculations. The details of different level descriptions can be found in the table below.

Table 3.7. Skill Definition for *Applied Mathematics*

Level	Characteristics of Items	Skills
3	• Translate easily from a word problem to a mathematics equation • All needed information is presented in logical order • No extra information	• Solve problems that require a single type of mathematics operation (addition, subtraction, multiplication, and division) using whole numbers • Add or subtract negative numbers • Change numbers from one form to another using whole numbers, fractions, decimals, or percentages • Convert simple money and time units (e.g., hours to minutes)
4	• Information may be presented out of order • May include extra, unnecessary information • May include simple charts, diagrams, or graphs	• Solve problems that require one or two operations • Multiply negative numbers • Calculate averages, simple ratios, simple proportions, or rates using whole numbers and decimals • Add commonly known fractions, decimals, or percentages (e.g., 1/2, .75, 25%) • Add three fractions that share a common denominator • Multiply a mixed number by a whole number or decimal • Put the information in the right order before performing calculations

5	• Problems require several steps of logic and calculation (e.g., problem may involve completing an order form by totaling the order and then computing tax)	• Decide what information, calculations, or unit conversions to use to solve the problem • Look up a formula and perform single-step conversions within or between systems of measurement • Calculate using mixed units (e.g., 3.5 hours and 4 hours 30 minutes) • Divide negative numbers • Find the best deal using one- and two-step calculations and then comparing results • Calculate perimeters and areas of basic shapes (rectangles and circles) • Calculate percentage discounts or markups
6	• May require considerable translation from verbal form to mathematical expression • Generally require considerable setup and involve multiple-step calculations	• Use fractions, negative numbers, ratios, percentages, or mixed numbers • Rearrange a formula before solving a problem • Use two formulas to change from one unit to another within the same system of measurement • Use two formulas to change from one unit in one system of measurement to a unit in another system of measurement • Find mistakes in items that belong at Levels 3, 4, and 5 • Find the best deal and use the result for another calculation • Find areas of basic shapes when it may be necessary to rearrange the formula, convert units of measurement in the calculations, or use the result in further calculations • Find the volume of rectangular solids • Calculate multiple rates
7	• Content or format may be unusual • Information may be incomplete or implicit • Problems often involve multiple steps of logic and calculation	• Solve problems that include nonlinear functions and/or that involve more than one unknown • Find mistakes in Level 6 items • Convert between systems of measurement that involve fractions, mixed numbers, decimals, and/or percentages • Calculate multiple areas and volumes of spheres, cylinders, or cones • Set up and manipulate complex ratios or proportions • Find the best deal when there are several choices • Apply basic statistical concepts

Reading for Information

The *Reading for Information* skill involves reading and understanding work-related instructions and policies. The reading passages and questions in the assessment are based on the actual demands of the workplace. Passages take the form of memos, bulletins, notices, letters, policy manuals, and governmental regulations. Such materials differ from the expository and narrative texts used in most reading instruction, which are usually written to facilitate reading. Workplace communication is not necessarily well-written or targeted to the appropriate audience. Because the *Reading for Information* assessment uses workplace texts, the assessment is more reflective of actual workplace conditions. There are five skill levels, with Level 7 being the most complex and Level 3 the least complex. The details of different level descriptions can be found in the following table.

Table 3.8. Skill Definition for *Reading for Information*

Level	Characteristics of Stimuli and Items	Skills
3	• Reading materials include basic company policies, procedures, and announcements • Reading materials are short and simple, with no extra information • Reading materials tell readers what they should do • All needed information is stated clearly and directly • Items focus on the main points of the passages • Wording of the questions and answers is similar or identical to the wording used in the reading materials	• Identify main ideas and clearly stated details • Choose the correct meaning of a word that is clearly defined in the reading • Choose the correct meaning of common, everyday and workplace words • Choose when to perform each step in a short series of steps • Apply instructions to a situation that is the same as the one in the reading materials
4	• Reading materials include company policies, procedures, and notices • Reading materials are straightforward, but have longer sentences and contain a number of details • Reading materials use common words, but do have some harder words, too • Reading materials describe procedures that include several steps • When following the procedures, individuals must think about changing conditions that affect what they should do • Questions and answers are often paraphrased from the passage	• Identify important details that may not be clearly stated • Use the reading material to figure out the meaning of words that are not defined • Apply instructions with several steps to a situation that is the same as the situation in the reading materials • Choose what to do when changing conditions call for a different action (follow directions that include "if-then" statements)

5	• Policies, procedures, and announcements include all of the information needed to finish a task • Information is stated clearly and directly, but the materials have many details • Materials also include jargon, technical terms, acronyms, or words that have several meanings • Application of information given in the passage to a situation that is not specifically described in the passage • There are several considerations to be taken into account in order to choose the correct actions	• Figure out the correct meaning of a word based on how the word is used • Identify the correct meaning of an acronym that is defined in the document • Identify the paraphrased definition of a technical term or jargon that is defined in the document • Apply technical terms and jargon and relate them to stated situations • Apply straightforward instructions to a new situation that is similar to the one described in the material • Apply complex instructions that include conditionals to situations described in the materials
6	• Reading materials include elaborate procedures, complicated information, and legal regulations found in all kinds of workplace documents • Complicated sentences with difficult words, jargon, and technical terms • Most of the information needed to answer the items is not clearly stated	• Identify implied details • Use technical terms and jargon in new situations • Figure out the less common meaning of a word based on the context • Apply complicated instructions to new situations • Figure out the principles behind policies, rules, and procedures • Apply general principles from the materials to similar and new situations • Explain the rationale behind a procedure, policy, or communication
7	• Very complex reading materials • Information includes a lot of details • Complicated concepts • Difficult vocabulary • Unusual jargon and technical terms are used, but not defined • Writing often lacks clarity and direction • Readers must draw conclusions from some parts of the reading and apply them to other parts	• Figure out the definitions of difficult, uncommon words based on how they are used • Figure out the meaning of jargon or technical terms based on how they are used • Figure out the general principles behind the policies and apply them to situations that are quite different from any described in the materials

Locating Information

The *Locating Information* skill involves the locating, comparative, summarization, and analytic skills people use when they work with work-related graphics. The types of graphics used as stimuli include tables, data graphs, forms, charts, flowcharts, diagrams, maps, floor plans, instrument gauges, and blueprints. These graphics are based on materials that reflect the range of locating information demands found in the workplace. Because the *Locating Information* assessment uses workplace graphics, the assessment is more reflective of actual workplace conditions. There are four skill levels, with Level 6 being the most complex and Level 3 the least complex. The details of different level descriptions can be found in the table below.

Table 3.9. Skill Definition for *Locating Information*

Level	Characteristics of Graphics	Skills
3	• Elementary graphics • Simple order forms, bar graphs, tables, flowcharts, maps, instrument gauges, and floor plans • One graphic used at a time	• Find one or two pieces of information in a graphic • Fill in one or two pieces of information that are missing from a graphic
4	• Straightforward graphics • Basic order forms, diagrams, line graphs, tables, flowcharts, instrument gauges, and maps • One or more graphics are used at a time	• Find several pieces of information in graphics • Notice how graphics are related to each other • Sum up information shown in straightforward graphics • Identify trends shown in straightforward graphics • Compare information and trends shown in straightforward graphics
5	• Complicated graphics with possibly unusual formats • Detailed forms, tables, graphs, diagrams, maps, and instrument gauges • One or more graphics are used at a time	• Sort through distracting information • Sum up information shown in detailed graphics • Identify trends shown in detailed graphics • Compare information and trends shown in detailed graphics
6	• Complicated graphics containing large amounts of information; may also have challenging formats, technical terms, or symbols • Very detailed graphs, charts, tables, forms, maps, and diagrams • One or more graphics are used at a time	• Analyze data in one complicated graphic or several related graphics • Apply the information to specific situations • Use the information to make decisions • Use the information to draw conclusions

Technical Characteristics of the WorkKeys Tests

ACT has conducted extensive psychometric analyses on the *WorkKeys* tests, including scaling and equating, reliability, and validity studies. Different equating methods are used in *WorkKeys*; the common-item nonequivalent groups equating method is used in MME-related work. As an important reliability index, internal consistency reliability was found to be high for the *Reading for Information* and *Applied Mathematics* tests or moderately high for the *Locating Information* test. ACT has used a multi-faceted approach to collect validity evidence of the *WorkKeys* tests. As part of criterion-related validity evidence, the studies showed positive correlations between the test scores of the three tests and job performance ratings ranging from 0.12 to 0.86, which compares favorably with the correlations found in the general research literature on criterion-related validity of employment tests. The technical characteristics—the score scale, equating, reliability, and validity—of the *WorkKeys* Tests is thoroughly documented in the *WorkKeys* Technical Manuals of respective tests (ACT, 2008a, 2008b, and 2008c). The *WorkKeys* Technical Manuals can be requested by calling 1-800/WORKKEY (967-5539) or from ACT's website at **www.act.org.**

Test Development for Day 3

Test Specifications

As noted in the previous chapter, all MME Day 3 subject tests are based on the high school content standards. A general description of development activities for all MME Day 3 subject tests is provided below, followed by subject-specific descriptions.

MDE staff, contractors, and Michigan educators worked together to develop the tests. The test development cycle included the following steps:

- Specification Development
- Item Writer Training
- Item Development
- Item Review
- Field Testing
- Field Test Item Review
- Operational Test Construction

In addition to assessing student knowledge of subject-specific content, the MME tests also assess student thinking skills in each of the three components (i.e., Day 1 ACT Plus Writing, Day 2 *WorkKeys* and Day 3 Michigan Component). Critical thinking skills are a primary focus of each of the three components of the MME. These skills are assessed through both multiple choice (MC) and constructed-response (CR) items. CR items only appeared in ACT Writing section in the Spring 2011 test cycle. The blueprints included in the subject-matter sections of this document reflect the crossing of content with process.

Step 1: Specification Development

Following the yearly alignment process undertaken by MDE (see Chapter 5 for more information on the 2011 alignment process), MDE and its contractors develop Michigan Component test specifications. The test specifications identify the content and types of items to be included. These specifications include the High School Content Standards, general indicators of difficulty, and other psychometric characteristics as well as general physical indicators such as artwork parameters. Test item specifications are very detailed

and identify content limits, item formats, and similar aspects of test items, typically including sample items of each format.

All MDE tests are designed to assess higher order thinking skills. Most items in all subject areas focus more on comprehension and application than on simple recall or recognition. Indeed, guidelines for item writing for each test clearly include admonitions for item writers to avoid simple recall of trivial or unrelated facts, and specific attention is given to ensure that tests include adequate higher-order thinking skills. The ways in which higher order thinking skills are included in each subject test is addressed by content area in the subsequent, content-specific sections.

Step 2: Item Writer Training

For Michigan-developed components, all item writers are Michigan educators who have curriculum and instruction expertise and who have been recommended by their administrators. All have relevant degrees and experience, and many have previous experience in MME-specific item writing.

Once test and item specifications are written, contractors and content consultants from the BAA use these materials to train item writers to produce items specifically for MDE. The item-writing process begins over the summer. Pearson holds one 3 day training in June for item writers, In some cases, veteran item writers are given their item writing assignments as much as a month before the actual face-to-face meeting. Teachers are trained by experienced Pearson Content Specialist team members. Teachers engage in peer review of the items and continue working on their items once they leave the training. Pearson Content Specialists provide extensive feedback to the writers and there is much back and forth in Pearson's web-based Item Authoring tool. Once the items are in-house, Pearson reviews them during various stages, to include up to five internal rounds (Content 1, Editorial 1, Content 2, Editorial 2, and Senior Review), schedule permitting.

Step 3: Item Development

The Michigan item writers draft test items as described above in accordance with specifications approved by the BAA. Once this is completed, Pearson prepares the items for the first BAA review in September. This internal review consists of items being evaluated using the following criteria:

Skill

- Item measures one skill level.
- Item measures skill in manner consistent with specification.
- Item uses appropriate (realistic) level of skill.
- Item makes clear the skill to be employed.

Content

- Item measures one benchmark.
- Item measures benchmark in manner consistent with specification.
- Item taps appropriate (important) aspect of content associated with benchmark.
- Item makes clear the benchmark or problem to be solved.

Relevance

- Item calls for a realistic application of process to content.
- Item is not contrived.

- Item is appropriate for the grade level to be tested.
- Item groups reflect instructional emphasis.

Accuracy
- Item is factually accurate.
- Item contains only one correct or best response.
- If item pertains to disputed content, context for correct answer is clearly defined (e.g., "According to... the correct solution is...").
- Item is unambiguously worded.

Format
- Item contains no extraneous material except as required by the benchmark.
- Vocabulary is grade-appropriate and clear.
- Item contains no errors of grammar, spelling, or mechanics.
- Item is clearly and conveniently placed on the page.
- Item contains adequate white space for calculations as needed.
- Physical arrangement of item is consistent with benchmark or common practice (e.g., horizontal vs. vertical addition and subtraction, slash vs. horizontal fraction bar, notation, symbols, etc.).
- Keys for sets of MC items are balanced (i.e., equal numbers of A's, B's, C's, and D's).

Bias
- Item is free of race and sex stereotypes.
- Item contains no material known or suspected to give advantage to any group.
- Item is free of insensitive language.
- Item sets that identify race or sex either directly or indirectly are balanced with reference to race and sex.
- Item content and format are accessible to students with disabilities.
- Item content and format are accessible to students with limited English proficiency.

Step 4: Item Review

After the internal review takes place, all items are reviewed by committees of Michigan educators and Michigan citizens. This consists of bias/sensitivity review meetings (involving 10-15 Michigan educators on-site, the Bias and Sensitivity Review Committee [BSC]) and content review meetings (involving roughly the same number of educators, the Content Advisory Committee [CAC]). This allows grade-level educators to spend more time focusing on the nuances of each item and adjusting the items when necessary. Contractor staff trains the CAC and BSC and monitors the reviews. All items are first reviewed by the BSC and then the CAC.

Any item rejected by the BSC does not get passed on the CAC for review. Each review is led by MDE and contractor staff, using prescribed guidelines and forms to indicate the final status of each item:
- **Accept:** Each of the following eight category conditions (importance, thematic, grammar, clarity, accuracy, validity, sound measurement, grade-appropriate) has been met or exceeded and the item appears suitable for field testing.
- **Modify:** One or more of the category conditions have not been met or the item needs minor changes to make it acceptable. Reviewers provide recommendations on changes to be made to the item that will make the item suitable for field testing.

- **Reject:** Several category conditions have not been met, or are suspect, or need radical changes to make the item acceptable. In such cases, the item may be vague or ambiguous, inappropriate, or not clearly related to the text or to the standard. Without severe modifications it is unlikely to be salvaged. Reviewers provide comments as to why the item should be rejected.

Step 5: Contractor Review

After this first round of reviews, the contractor incorporates all changes into the items and the items are ready to be field tested. They are placed on forms and are composed by the BAA composition team. The items and forms are reviewed by the BAA content leads and other internal BAA staff and the composition team makes any necessary revisions. The items and forms are returned again to the BAA content lead and test development manager for overall quality control and sign-off before they are sent for printing.

Step 6: Field Testing

Items that have passed bias/sensitivity and content review are field tested. MME field testing is done in embedded operational test forms. All test forms consist of a certain number of operational items, along with a number of field test items. Field test items are distributed amongst the test forms. The process for field test item review is described in detail below.

Step 7: Field Test Item Review

After field testing, contractor staff analyzes item results and presents them to the same groups listed under Item Review above, which gives committee members the opportunity to review the items with field test statistics. Formerly, the Merit Award Board had final review authority over all test items. This function has recently shifted to the Department of Education. Once items and their field test results have been presented to the Department and the Department has accepted them, they go into banks of items from which future operational tests may be constructed. The processes for field test item reviews are presented in greater detail below, and results of field test item reviews that occurred during the 2010–2011 administration cycle are discussed later in this chapter.

Field Testing Procedures: Item Development, Review, Field Test Design, and Statistics

This section provides an overview of the field testing procedures, conducted by the development contractor. The specific item review process at various test development stages is described in other sections of Chapter 3.

Field Testing Design

The BAA conducts field testing by embedding matrix-sampled field-test items across multiple forms of operational assessments such that in general each field-test item appears on only one operational form. The total numbers of unique field test items and field test items embedded across forms are given in Table 3.10.

Table 3.10. Number of Forms and Field Test Items by Subject

Subject	Number of Forms	Total Number of Unique Field-test Items	Total Number of Field-test Items
Mathematics	12	99	108
Science	12	204	204
Social Studies	12	176	192

Field Test Sampling

It is critical that field test items be calibrated with operational items in such a way that the obtained item parameters represent those parameters that would result were the field test items administered to all students. For the MME, each form (1-10) is spiraled within each classroom in each school. Therefore, every school gets every form for some of its students, which helps to ensure that the field test item parameters are representative of those which would be obtained if the items were administered to all students.

Item Specifications

MDE employs *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 1999) as a primary source of guidance in the construction, field testing, and documentation of the tests. The introduction to the 1999 *Standards* best describes how those *Standards* are and were used in the development and evaluation of tests: "Evaluating the acceptability of a test or test application does not rest on the literal satisfaction of every standard in this document, and acceptability cannot be determined by using a checklist." (*Standards*, p. 4).

Thus, the terms 'target' and 'goal' are used when referring to various psychometric properties of the tests. For example, while it is a goal of test development for each high school test to have a reliability coefficient of .90 or greater, it is not our intention to eliminate a test with a reliability coefficient of .89. Instead, the test results would be published, along with the reliability coefficient and associated standard error of measurement.

Item Statistics

Because the MME tests are used in making individual decisions about students, they must be very reliable, particularly at cut points (the score points that separate adjacent achievement categories). Due to the fact that eligibility for Michigan Promise Scholarships[1] is involved at the high school level, the reliability at the scholarship cut score must be very high. Target reliability coefficients of .90 (or higher) are therefore set for each test. Other psychometric properties include item difficulty, item discrimination, and differential item functioning. General statistical targets are provided below.

For Multiple-Choice (MC) Items

- Percent correct: between 30 and 90 percent

[1] Please note that, as of the spring of 2011, funding is not available for the Michigan Promise Scholarship. However, funding was available at the time of the 2009 test administration. Additionally, students are still encouraged to take the complete MME to establish their eligibility for the Promise Scholarship should funds become available.

- Point biserial correlation with total score: .25 or greater
- Mantel-Haenszel: Few Category C items

It should be pointed out that the point biserial correlations for MC items assume embedded field testing and employ the base test total score, which is independent of the field tested item.

Differential Item Functioning

Items that disadvantage any identifiable subgroup of students are considered biased and detract from the validity of the tests. While only human judges can determine whether or not an item is biased, item statistics can serve as a tool to help judges in their decisions. After field testing, the BSC reviews item statistics that detect differential item functioning (DIF). Specifically, Mantel-Haenszel statistics are used as measure of DIF. Mantel-Haenszel statistics are the industry standard methodology for DIF analyses, and correspond well with the categories used by ETS. These analyses are conducted after field testing. The Mantel-Haenszel statistics are generated for each item, which alert the BAA and the BSC committee to the possible presence of DIF. At this point, the BSC reviews the item further to substantiate the item statistic flag. If the item is found to indeed have DIF, it is not used in its current form in further assessments.

Field Testing Embedding

No released items are made for MME and pre-equating is not employed for the MME.

Post-Field-Test Item Review

After field-test administration, an item review process is undertaken to evaluate the items for further use. This section describes the steps of that process, which include: (1) Preparing item statistics for internal use and for review committees, (2) internal and contractor review of statistics, (3) item review, including item statistics, by bias and content committees, and (4) potential item revisions.

Field Test Item Statistics and 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. 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: MME 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 versus female comparison and white versus 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; and
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, 5th, 25th, 50th, 75th, 95th percentile for creating Box-and-Whisker plot for each student group (total, male, female, white, and black students) for each option and each score point.

The process of generating item statistics is as follows:
For Days 1 and 2, ACT complete all scoring and produce raw scores, which they send to the BAA. For Day 3, Measurement Incorporated completes the scoring, and provides the BAA with raw scores and with any necessary erasure analyses. The BAA then creates a matched file, with data from Days 1, 2, and 3 and returns this to ACT. ACT calibrates the tests and calculates the scale scores, conducts IRT analyses and produces the statistics listed above, which they then provide to the BAA for further analyses and use by review committees and BAA psychometricians. Finally, Measurement Incorporated produces the final score reports, using the scale scores and other information generated by ACT.

Statistics Prepared for Review Committees

From the analyses listed above, the following statistics were used to create item labels for the post-field-test reviews. Different sets of statistics were prepared for MC for review committee. Figure 3.1 displays all the statistics prepared for MC items for 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 descriptive statistics 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 versus female and white versus black; and
6. Review decision.

When the p-value for an MC item was out of the desired range, a difficulty flag was shown. When point-biserial correlation for an MC item was out of range, a point-biserial or item-total correlation flag was shown. If the DIF level for male versus female or white versus black comparison was higher than moderate, a DIF flag was turned on. When options did not function well or score point distribution was abnormal, a miskey flag was on. The criteria used for flagging an MC item are presented in Table 3.11.

MME Grade: 11 Subject: Science Admin: Spring 2010

ID: 100000323646
Form: 1005
Position: 36
Scenario: NA

GLCE: P2.01E

☐ Accept as is
☐ Reject

Table 1. Item Information

Type:	MC	**P-value:**	0.30	**Difficulty Flag:**			
Key:	B	**N-count:**	10523	**PB Correlation:**	0.18	**PB Correlation Flag:**	CL
		Maturity:	FT			**Option Quality Flag:**	H P

Table 2. Breakout Group Descriptives and Option Analysis

		N-count	Percent of Students Selected Option				
			A	B *	C	D	Omit
Group	All	10523	13	30	31	26	0
	Male	5140	15	37	26	22	0
	Female	5383	11	24	36	29	0
	White	8103	13	32	31	24	0
	Black	1589	12	23	32	33	0
Option PB Correlations			0.01	0.18	-0.09	-0.10	-0.04

Table 3. Differential Item Functioning

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

Figure 3.1**.** Item label for a MC item.

**Note: IRT statistics were not provided during the most recent data review (July 2010).*

Table 3.11. Flagging Criteria

Statistic	Desired	Definition	Comments
P-value	>0.3 and < 0.9	The percentage of students who answered the item correctly.	Outside this range and the item may be too difficult or too easy. The desired overall mean P-Value on a test is around 0.6.
PB Correlation	> 0.25	The relationship between students' performance on an item and their performance overall on the test.	Any less than 0.25 and the item may be unreliable in discriminating well between the high and low achievers; if less, consider the response distribution and the item content.
DIF Flag	No Flag	DIF refers to the differences in performance on a studied item between the reference and the focal groups after the two groups have been matched by ability.	DIF only indicates that the examinees of equal proficiency from different subgroups have an unequal probability of responding correctly to an item. The items that exhibit DIF should be carefully examined for potential bias against particular groups.
Option Analysis	The option of key has the highest percentage	Option analysis (Score point distribution) shows the percentage of the total students and those in the gender and ethnicity subgroups who chose each option.	The keyed option should usually have the highest percentage. The keyed option point-biserial correlation should be larger than 0 while the non-key option should be smaller than 0.

Notes: Explanation of Flags:

PL … p-value low
PH … p-value high
CL … correlation low between item and total
B … moderate DIF
C … substantial DIF
H … highest percentage is not a keyed option
L … low percentage of any option *(less than or equal to 2%*)
P … positive pb-correlation for any non-keyed option
N … negative pb-correlation for the keyed option
O … omit has a positive pb-correlation greater than .03

Item Reviews

Bias/Sensitivity and Content Committee Review

Pearson planned and conducted Bias/Sensitivity committees (BSC) followed by Content Advisory committees (CAC) on field tested items that were flagged either because of Differential Item Functioning (DIF) or any other content related item property (see flagging criteria on Table 3.11). The goal of these committees was to identify items that are eligible to be used as scorable items in future operational assessments. In these meetings, items may be either: a) accepted "as is' or b) discarded. The BAA identified the members of the BSCs and CACs using members previously involved in the development contractor committees.

Pearson prepared items following field testing for reviews by a Bias/Sensitivity Committee and a Content Advisory Committee. The administration contractor (Measurement, Inc.) provided the field test statistics necessary for Pearson to create the data labels necessary for BSC and CAC meetings. Pearson assembled all materials for the meetings including the items, data and analyses of the items, agenda, training materials, security agreements, sign in sheets, and the necessary records for committee sign off on each item. Each committee met for one or more days depending on the number of flagged items. The reviews were guided by checklists to ensure that the items meet the criteria for inclusion in the item bank and for potential use on future examinations. Pearson reviewed the flagging and review criteria with the BAA to be sure that all nuances of acceptability are captured correctly. The review panel examined each item and determined if it is of high quality and matches the intended assessment objective. The items were reviewed to ensure they are appropriate for the grade level. The determination of accuracy of all material, checking that each question has only a single right answer, was part of the review process. The item statistics for each item were presented, along with a general orientation to interpretation and use of the data in item approval. The bias and sensitivity committee focused additionally on issues that ensure that the items have no stereotypical statements, present no unfair advantage or disadvantage to any group, and are free of bias for race, ethnicity, gender, age, disability status, and any other category of individuals for whom the item may be unfair. These reviews occurred in face-to-face group meetings. Separate review sessions were created for content and bias/sensitivity review. Items were organized by content area and high school content expectation. BAA staff had access to all items and comments during the review period. Reports were generated that detail the reviewer's comments, and they have been evaluated by Pearson assessment specialists in conjunction with the BAA. Pearson worked with BAA staff to ensure that committee decisions were captured accurately.

Item Revision Procedures

It is Pearson's policy to leave post-field test items as intact as possible since those items have data attached to them (i.e. item statistics, etc.). Making major changes can negate that data. However, there are circumstances where the item may be revised.

Generally, the field test data review committee examines items and either accepts them or rejects them. Occasionally, committee members suggest minor revisions that could improve the clarity or quality of the item. The BAA must approve of any changes to the item, and if the committee or the BAA believes significant changes are required to improve the item, it is rejected as ready for operational use.

The committee's recommendations are brought back to Pearson and entered into the system. At this time, the field tested items are available for use on operational forms. Items selected for operational use will be composed and reviewed by Pearson staff, then sent to the BAA for review. Minor changes may be requested

as the items are considered as a group. Once Pearson has made these revisions and the BAA has approved the form, the Pearson quality assurance team reviews the form one final time before it is sent to the printer.

Item Banking

Procedures

After items have undergone the field testing procedures described above and are ready to become operational items, they are entered into an item bank. Pearson developed a secure item bank for grades K through 12 in the areas of Mathematics, English Language Arts, Science and Social Studies for the BAA. The BAA desires an electronic item bank software application that includes search, preview, and reporting functionality. They also desire this application to be secure (maintained by a password and other controls to be used for statewide high-risk assessment preparation).

Existing Architecture

File Maker Pro 7.0 software runs the Item Bank application. A Pearson developed Item Bank prototype has been modified to meet these requirements. The application runs as a secure web site. File Maker Pro 7.0 is a relational database with multiple tables. Each item and its assets are stored as an individual record.

Populate the Item Bank

A unique, seven-digit Corporate Identification Code (CID) was assigned to each item. All passages were assigned a unique passage code. The source word document was a compilation of each item, and associated passage if applicable, at a designated grade level and status. To prepare for the population into the item bank, each item was saved as an individual word document.

The individual word documents were converted to three formats: .pdf, .gif, and .txt. The .pdf file is used to build reports. The .gif format displays the item as it appeared in the most recent test administration. The .txt document is stored, but not visible to the user, in order to perform a search on any character within the item.

Data is imported in the Item Bank via Microsoft Excel spreadsheet. A unique relationship for all assets is defined by the Corporate Identification Number (CID). Image files are imported directly from the folders which they are grouped.

Data Verification Process:

The following process is followed by Pearson, Inc. to verify data entered into the item bank:

- Check for modifiable fields (exception is Comments field).
- Export data from FileMaker tables (into Excel or Access).
  - Look for blank fields.
  - Look for duplicates.
  - Look for anomalies in data.
  - Assign a second reviewer to check data.
- Compare counts to prepared list.

Quality Control Procedures for Item Bank

Quality Control of Items

- A hard copy of every item should be available, typically from the test booklet or final PDF.
- A comparison of each item should be done by comparing the test booklet data and preview with the item text on the data screen and preview in the Item Bank for each item.
- The hard copy from the book will only show the reviewer the stem, art, and options.
- The item text (on the Item Data screen) will not contain any formatting. The actual content should be verified as correct.
- Test Define will be used to obtain the Metadata information for each item.
- The form number will be stored in the statistics table and will be checked during the QC of the statistics.
- The Item Status and the "Item Not Appropriate" field information will be verified against the "Released Position" field in the Test Define.

Quality Control of Passages

- A hard copy of each passage should be available (test booklet or final PDF).
- A comparison of each passage should be done by comparing the hard copy and preview with the text on the passage data screen and preview in the Item Bank for each passage.
- The passage text (on the passage data screen) will not contain any formatting. The actual content will be verified as accurate. Text should include titles; subtitles; by-lines; captions for photos, illustrations and drawings; titles and subtitles for photos, illustrations and drawings; footnotes; numbers; text and numbers within a graph, table or chart.
- The metadata that is associated with a passage (passage title, author, word count, etc.) should be verified against the source document that was used to populate the item bank. This will be accomplished by comparing the source document with an exported list of these fields from the item bank for accuracy and completeness.

Quality Control of Passages and Related Items

- The Test Define is the source document to verify that all passages in the item bank are related to the appropriate items.
- View each passage in the bank to verify that all appropriate items are linked to their respective passage.

Quality Control of Metadata

- Export item metadata out of the item bank, import the data into an external database (Access), and perform queries (null, unmatched, duplicate, and cross tab) to verify the data is linked together properly and that all data is accounted for.
- Missing data must be populated before a final stats QC can take place. Once missing data has been accounted for and statistics have been populated, perform searches based on administration (year and season), grade, and form number. Verify counts against Test Defines for actual count of item type, item designation, and sequence numbers.
- Compare the test booklets to the found set in the item bank. Test defines may also be cross-referenced at this time. In comparing the representation of the item in the item bank to the test booklets, the contractor should focus on:
  - Year, Season, Form, CID, Item Code, Grade, Item Type, Art, Passages, Item Designation (CR or FT), Item Status, Item Not Appropriate, Item Preview, and Correct Answer

(Item vs. Item Preview screens, Item vs. Stats screen).

Quality Control of Item Statistics

- Use the file from Psychometrics that contains all statistical information to QC the statistics information. This is the statistical data received from Pearson Educational Measurement, Inc.
- Export the records from the item bank (CID and all stats) into an external database (Access) and compare it with the file provided by Psychometrics. Any missing data, duplicate data, or anomaly in any of the fields is an indication that further review and verification is needed.

Quality Control of Functionality

- Verify that the item bank functions as required:
  - Fields that should be searchable or modifiable should have that functionality.
  - Linking between tables of the item bank.
  - Built-in search functions, such as viewing items associated with a passage, etc.
  - Verify that selected sets can be saved and exported from the bank.
- A test-run of all reports should be executed to verify accurate performance.
- Verify that the saved forms, selected items, and found sets have been cleared before the final compile.

Data Included in Item Bank

The data included in the item bank is the same as the data prepared after field test administration. Please refer to the section on post field test item review to get detailed information.

An item bank performs three broad functions: (1) it acts as a repository for test items, passages, statistics, and associated attributes, such as metadata and art; (2) it permits the selection of items for test construction by multiple search criteria; and (3) it can generate a number of basic reports.

Construction of Operational Test Forms

The Michigan Department of Education's Bureau of Assessment and Accountability (BAA), formerly, Office and Educational Assessment and Accountability (OEAA), Measurement Incorporated (Measurement Inc. or M.I.) and Pearson Educational Measurement (PEM) work collaboratively to develop and construct the operational test forms used to support the MME program.

Test form development entails the following steps:

- Review the assessment blueprints for the operational assessments.
- Select assessment items to meet the content and process specifications of the assessment blueprints.
- Assess the statistical characteristics of the selected assessment items.
- Review and approve test forms.

The following sections discuss essential aspects, include guidelines, and identify important references to follow through the four-step process.

Assessment Blueprints

As the name implies, the assessment blueprints identify the content and types of items to be included on the operational forms. These specifications include benchmark and content targets (limits), general indicators of

difficulty and other psychometric characteristics, as well as general physical indicators such as passage length and artwork parameters.

All MME assessments are designed to assess higher order thinking skills. Most items in all subject areas focus more on comprehension and application than on simple recall or recognition. Indeed, specifications for each assessment clearly include admonitions to avoid simple recall of trivial or unrelated facts.

For 2011, the MME (Day 3 Michigan-developed) assessment used multiple-choice (MC) items only. Each item is aligned to a specific domain, standard, and objective. The alignment information is used during the forms construction process to help ensure the forms meet the blueprints.

This section provides an overview of the test blueprints for each subject, accommodated materials, and item specifications that guide the building of the operational test forms. The 2011 MME (Day 3 Michigan-developed) test contains three subject area tests: mathematics, science, and social studies. The test structures are summarized in this section.

Mathematics

The MME Mathematics Assessment is based on the Michigan High School Content Expectations. For 2011, each mathematics form includes a common set of two MC items per Standard (maximum of 10 points from common items), plus a matrix of items (one item per standard), and Field Test items (as needed). Ten unique initial forms, one makeup form, and an accommodated form were constructed. In order to ensure comparability across all forms, each form is developed based on the carefully constructed test specifications and test development principles, outlined previously in this chapter. Equating methodologies are then used to ensure that the scales are on comparable levels (see Chapter 9 for more information on scaling and equating). These forms are spiraled within each classroom, so that all ten initial forms are distributed across schools and students. The test structure for MME mathematics assessment is summarized in Table 3.12.

Table 3.12. Test Structure for the Spring 2011 MME Mathematics Core Test

Subject	# Common Operational	#Matrix	#Field Test	Total Operational Items
Mathematics	10	6	9	16

Science

For the 2011 MME Science test, each form consists of a common set of HSCEs (one item per Standard, for a maximum of 16 points from common items), plus a matrix of items that cover the other HSCEs (one item per Standard), and Field Test items. Ten unique initial forms, one makeup form, and one accommodated form were constructed. As described in the mathematics section above, each form is comparable due to the test specifications and test development principles, and is then equated and scaled using the methodologies outlined in Chapter 9. The test structure for science tests is summarized in Table 3.13.

Table 3.13. Test Structure for the Spring 2011 MME Science Core Test

Subject	#Common Operational	#Matrix	#Field Test	Total Operational Items
Science	16	16	17	32

Social Studies

For the 2011 MME Social Studies tests, ten initial forms, one makeup form, and one accommodated Form were constructed. As described in the mathematics section above, each form is comparable due to the test specifications and test development principles, and is then equated and scaled using the methodologies outlined in Chapter 9. The test structure for social studies tests is summarized in Table 3.14.

Table 3.14. Test Structure for the Spring 2011 MME Social Studies Core Test

Subject	#MC Operational	#Matrix	#Field Test	Total Operational Items
Social Studies	32	N/A	16	32

Accommodated Formats

Each operational test is available to students who require accommodations according to their IEP, section 504 plan, or ELL instructional plan. Tests are available in Braille, large print, audio cassette, audio DVD, and video DVD. Form 12 is generally a unique form for accommodation for all the three components of the MME. Students testing with accommodations take the MME in sequence within a two-week accommodated testing window. For more detailed information regarding accommodated formats of the MME, see the *Spring 2011 MME Day 3 Administration Manual for Students Testing with Accommodations*.

Item Selection

In addition to the content coverage requirements, the forms must also meet certain statistical targets. These targets are outlined in the next three sections below.

Select Assessment Items to Meet the Assessment Blueprints

Following field testing, the items are submitted for review to both the Bias Review Committees (BRCs) and the Content Advisory Committees (CACs). These committees, composed of Michigan educators and Michigan citizens, sort the field tested items and identify which items are eligible for inclusion in the operational item pool. There is a separate pool for each subject assessed. It is from these pools that items are selected to meet the requirements outlined in the assessment blueprints.

Test forms are developed using the selected items. In addition to overarching content requirements for each test form developed, content experts and psychometricians consider requirements related to subdomains, graphics and other visual representations, passage and content dependent items, and clueing concerns.

Assess the Statistical Characteristics of the Selected Assessment Items

The statistical process begins with the work of the Content Advisory Committees and the Bias Review Committees following the field test. The committees evaluate the field test items using item statistics from classical measurement theory and item response theory models. From the work of these committees, a pool of items that are eligible to be used in constructing the operational forms is identified.

Because the MME assessments are used in making individual decisions about students, they must be very reliable, particularly at cut points (the score points that separate adjacent achievement categories).

Particularly at the high school level, because Merit scholarships are involved, the reliability at the scholarship cut score must be very high. The targeted reliability coefficient is .90 (or higher) for each assessment. Other psychometric properties include item difficulty, item discrimination, and differential item functioning. General item and form level statistical targets are provided below:

For Multiple-Choice (MC) Items

- Percent correct: $.25 \leq$ p-value $\leq .95$.
- Point biserial: $\geq .25$.
- Mantel-Haenszel: Few Category C items[2].

To help ensure adequate coverage of a full range of achievement on the operational assessments, the draft forms are evaluated to see whether the following targets are met (see Table 3.15). As necessary, items are replaced on the draft forms until this distribution is approximately approached.

Table 3.15. Desired Range of Item Difficulty Distribution

Rasch Item Difficulty	% of items
-2.00 to -1.00	25
-0.99 to 0.00	25
0.01 to 1.00	25
1.01 *to* 2.00	25

Even with careful test form development, it is usually not possible to create alternate forms that are exactly equal with respect to difficulty. The MME assessments are being analyzed using Item Response Theory (IRT).

As the MME test forms are assembled, spreadsheets are used to track the statistics and other metadata (e.g., alignment) for the selected assessment items. Both classical and IRT statistics are included. The statistics listed on the spreadsheets include item p-values, correlations, and IRT item difficulties for multiple-choice items and item means, standard deviations, correlations, and IRT step difficulty estimates for constructed-response items.

The above two steps require an iterative process to create test forms that are a combination of the content and statistical information. Working together, Pearson psychometricians and content experts replace items until both groups are satisfied with the forms. Through this iterative process of item selection, item content takes precedence over statistical characteristics.

Review and Approve Test Forms

Once Pearson staff have reached consensus on a test form, the form and associated information is submitted to the BAA staff for review and approval. Included in the test matrices are open slots for embedded field test items. The BAA reviews the test forms to determine whether both content and statistical requirements are met.

Guidelines for test forms review include:

- Confirm that all assessment items were accepted by the BAA and the committees;

[2]For category C items, D's absolute value is significantly greater than or equal to 1.5.

- Confirm that all blueprint requirements are met;
- Confirm that all content considerations including content/skill/topic balance, correct keys, no clueing, and correct graphics are met;
- Confirm that the item and mean difficulty levels are accurate and meet requirements; and
- Confirm that the assessments cover a full range of achievement levels.

As necessary, the BAA and Pearson replace items that are identified by the BAA as problematic, either from a content or psychometric perspective. As items are replaced, the match of the newly revised test form to the specifications is updated and reviewed. This process continues until the BAA has approved each form.

Accommodated Test Forms

A testing accommodation is a change to the testing environment to assist a student with special needs so that assessment can mirror instruction as much as is possible without invalidating test results. District and campus testing coordinators are responsible for communicating information about testing accommodations to test administrators and other interested individuals. Information about testing accommodations is also included in the test administrator manuals.

The decision to use a particular accommodation with a student should be made on an individual basis and should take into consideration the needs of the student and whether the student routinely receives the accommodation in classroom instruction and testing. If a student receives special education services, all accommodations must be documented in the student's individualized education program (IEP), section 504 plan, or ELL instructional plan.

Typically, accommodations allow for a change in one or more of the following areas:
- Presentation format
- Test setting
- Scheduling or timing
- Response format

The following accommodated testing materials are provided for the MME: Braille, Large Print, Oral Administration and Bilingual.

Accommodated Format Production: Day 1 ACT Plus Writing

For the MME Day 1 materials for the ACT Plus Writing, the Braille version is created from the unique accommodated form. This same form will be used for all students who take the ACT Plus Writing with accommodations. ACT test forms are designed from the outset according to principles of universal design, so that the tests are amenable to accommodations across the range of testing populations, conditions, and formats. ACT keeps tests as simple and straightforward as possible, consistent with curricular requirements—and this applies equally to vocabulary, graphics, typographic design, page layout, and the interrelationships among all these elements.

The accommodated form is provided to National Braille Press (NBP) for production of the Braille version. NBP is responsible for Braille transcription and creation of the raised line drawings included in the booklet. ACT does an additional proof of the Raised Line Drawings, but otherwise, NBP is responsible for all quality control checks.

Large Print

The Large Print format is developed from the unique accommodated form. ACT generally maintains the item layout of the regular type test booklet where possible; sometimes the layout of certain enlarged graphics must be adjusted so that the graphics do not cross over the binding and become obscured. ACT standard is 18-point font for large type. ACT produces and proofs the copy in-house before delivering it to The Brandt Company for printing. Brandt performs quality control checks in addition to the ones at ACT.

Oral Presentation

Students approved for oral presentation have the tests read to them either by a live reader or from a recording, in the three formats outlined below.

Reader Scripts

Reader Scripts are used when a student will have the test read by a qualified member of the testing staff individually in a separate room. The Reader Script is created from the tapescript (see below) once the audiocassette masters have been approved. These scripts include detailed instructions to the reader on administration procedures, how items are to be read, and guidelines to ensure a standardized administration no matter who is reading the tests. Reader Scripts are created from the unique accommodated form by ACT Test Development and proofed extensively before being delivered to RR Donnelley for printing. Reader Scripts are currently scanned directly from final camera-ready copy.

Cassettes and Audio DVDs

The audio recordings for cassettes and Audio DVDs are created from the unique accommodated form, using a tapescript written by ACT Test Development. The narrator is chosen by ACT and the same recording is used for both cassette and audio DVD formats. The audio recordings include a recitation of each item, as well as a recitation of all directions (stop, turn the page, etc.). They also include instructions for students on how to recheck their work or refer to passages in the test booklet students follow along with as needed. The cassettes are created first, and once the masters are approved, a digital file is delivered to the audio DVD vendor to perform "tracking" that is unique to the audio DVDs. Tracking the discs enables students to efficiently refer back to items and recheck their work. Cassettes and Audio DVDs are only available in English for the ACT Plus Writing.

Translated and Video Formats: State-Allowed Administrations.

The ACT Plus Writing is also available as an English Video DVD. The video component consists of a test booklet on the screen, intertitles preceding the questions, and prominent arrows that follow the screen text in sync with the audio component, as a visual aid to students. Additionally, the ACT Mathematics and Science Tests, along with the directions for all tests, are translated into Spanish and Arabic, the top language groups represented in the state after English.

This translation is done from the unique accommodated form and is also presented as a video DVD using the same English video component as described above. ACT's subcontractor, Metro Studios, contracts out the translations, and is also responsible for synchronizing the English audio, translated audio, and English video components together. The translation team consists of a primary translator who also narrates the tests, a spotter who ensures the translated test is narrated exactly as shown in the Reader Script, and a proofer who compares the finished recording to the English version of the test and identifies any translation errors or questions. Metro Studios facilitates any necessary discussion between the original translator and the proofer, and revisions are made as needed. Metro Studios has primary responsibility for translation accuracy and performs quality control checks for all three video formats. Students testing with a translated format receive the accommodated test booklet, printed in English, with which to follow along. This booklet matches the one displayed on-screen in the video.

Accommodated Format Production: Day 2 *WorkKeys*

In a particular administration, initial testing, make-up testing, and accommodated testing typically have different sets of questions. The test forms, however, are built to identical specifications and are fully equated to the other test forms administered in that testing situation as well as to forms used by the general population. Regardless of the accommodation, the same test form is used for the translated forms, Braille forms, large print forms, reader scripts, and other accommodations.

Translations

The International Test Commission has developed guidelines for test adaption, especially across cultures. The guidelines reference *Adapting Educational and Psychological Tests for Cross-Cultural Assessment* by Ronald K. Hambleton and colleagues (2005) regarding advantages and disadvantages of various translation methods. ACT's *WorkKeys* Development team has chosen to use the back translation method of quality control.

ACT uses an outside company for the Spanish translations and a different source for the Arabic translations. ACT works with a local company that produces the translated materials of the WorkKeys accommodated forms. Back translations are done by different personnel than those doing the original translation, for quality control. The *WorkKeys* editor (MA, Foreign Language Education) then compares the back translation with the original reader script done by *WorkKeys* personnel.

The company doing the Spanish translations has three staff members whose native language is Spanish, two of those with college degrees in communications and languages. The Arabic translation is also done by a staff member whose native language is Arabic, and who is a certified Arabic teacher and chair of a department for English Language Learners.

WorkKeys Video DVDs are made available in English, Arabic, and Spanish. The same approach is used for these as for the ACT Plus Writing Video DVDs. However, in the case of the *WorkKeys* test, *Reading for Information* is considered the language arts segment of the test and it is presented entirely in English. *The Applied Mathematics* and *Locating Information* segments (test questions, answer choices and Test Directions) are translated into either Arabic or Spanish for the audio component of each respective DVD.

Braille

ACT currently uses two vendors for the *WorkKeys* assessments. Both vendors follow the codes set forth by the Braille Authority of North America (BANA) and the guidelines for proofreading as used for the National Library Service (NLS). You can find the NLS specs and guidelines at its website: http://www.loc.gov/nls.

In short the procedures for proofreading Braille are to have the document read after translation by a team consisting of a sighted person and a blind person. The blind person reads aloud to the sighted person who follows the print. When that is completed, a correction sheet is returned to the translator. After those corrections are made, it goes to a different team to be re-read. All the text is re-read, keeping in mind the corrections from the first reading. After the test is corrected from the second reading and the corrections are approved, it goes to a QC person for review. The QC person compares the Braille to the hardcopy checking for any possible errors, which sometimes might be formatting Braille. When the document is approved, it is sent to have TA notes written, if applicable. The TA notes are then checked by the translator. When the test

goes to the production floor, a percentage of the tests are checked by the same QC person before the final copies are produced.

It should also be noted that both vendors use transcribers and proofreaders that have been certified by the Library of Congress.

ACT also receives a proof copy of the Braille document. We review all tactile graphics and visually compare them to the original art to make sure everything is included.

Other media

Reader scripts are prepared by ACT for appropriate *WorkKeys* assessments, indicating how each item should be read aloud (e.g., pronunciations of names, format of numbers, mathematical statements). Audio recordings are created using the reader script with a "spotter" following the script during the recording, and a "proof" copy is created for further checking by a *WorkKeys* editor for exact match. For large print materials, text is typically enlarged 130%. As noted, all materials undergo several quality control checks.

Accommodated Format Production: Day 3 Michigan Components

Braille

For the MME Day 3 materials, the Braille test is created from the accommodated form. Items for the accommodated form are selected specifically because of their adaptability to Braille (in addition to meeting test specifications). Doing this ensures that items do not have to be dropped from the Braille form and replaced with other items, which helps ensure the comparability of the Braille form to the accommodated form, and thus to the other test forms.

Once the accommodated form is produced, it is provided to an independent subcontractor, Cheeney Media Corporation, who is responsible for the production of all accommodated formats. Cheeney in turn subcontracts the Brailling and production of the Braille form to the American Printing House. After the American Printing House finishes translating the form into Braille, Cheeney Media conducts the appropriate quality control checks.

Large Print

Like the Braille format, the Large Print format is developed from the accommodated form. The items on this form are screened for adaptability to large print. Text is enlarged to one of four font sizes based on the degree of visual impairment. The font sizes offered reflect the sizes of print being used in current instructional situations. Mathematics diagrams requiring measurement are not enlarged.

Cheeney Media Corporation subcontracts the production of this format to the American Printing House as well, and performs the appropriate quality control checks after the American Printing House produces the Large Print forms.

Oral Administration

Students may have oral administrations by having a test administrator read the script aloud or by using a pre-recorded audio version of the scripted test.

Reader Scripts
Reader Scripts are created for each test component for each day, indicating exactly how each item should be read aloud without compromising the quality of the item. For example, if a problem requires students to indicate the largest number, the answers would not be read aloud. These scripts include detailed instructions to the reader regarding how to administer the assessment fairly. They also include phonetic spelling and other guidelines to ensure that each Reader reads the script in exactly the same way, as this is important for a standardized administration. Reader Scripts are produced from the accommodated format, and are carefully checked by Cheeney Media Corporation and by the BAA for accuracy (i.e. are all of the items the same on the accommodated form and in the Reader Script? Are there any errors in the spoken specification?)

Audio Recording
The audio recordings are created from the accommodated form, using the Reader Script as a script. The audio recordings include a recitation of each item, as well as a recitation of all direction (stop, turn the page, etc.). They also include instructions regarding how to review items if necessary. Audio recordings are available in English, Spanish and Arabic.

Bilingual Tests

The MME is printed in English, and is translated into Spanish and Arabic, the top language groups represented in the state after English. This translation is done from the accommodated form. For MME Day 3, Cheeney Media Corporation uses an independent subcontractor to perform the initial translations. These translations are then re-translated by a separate independent subcontractor to ensure accuracy. If there are any discrepancies, Cheeney Media facilitates the discussion between translators and produces a maximally accurate translation.

For the test administration, students receive an accommodated form with the questions printed in English, but are then provided with a DVD with the translation in Spanish or Arabic. Students may have the test interpreted on the day of testing for languages where a recorded bilingual version is not available.

Chapter 4: Administration

A valid and reliable MME assessment requires that assessments are aligned with the Michigan High School Content Expectations (HSCEs) and then administered and scored according to sound measurement principles. The MME is composed of three primary elements:

- Day 1 – The ACT Plus Writing;
- Day 2 – *WorkKeys* (*Reading for Information*, *Applied Mathematics*, and *Locating Information*); and
- Day 3 – Michigan developed mathematics, science, and social studies.

Sound assessment practices require that schools administer all assessments in a standardized and consistent manner across the state to ensure that all students have a fair and equitable opportunity to attain MME scores that accurately reflect their achievement in each of the MME content areas:

- Reading,
- Writing,
- Mathematics,
- Science, and
- Social Studies.

The schools play a key role in administering the MME assessment in a manner consistent with established procedures, monitoring the fair administration of the assessment, and working with the BAA to address deviations from required assessment administration procedures. School Test Supervisors, Backup Test Supervisors, Test Accommodations Coordinators, Room Supervisors, and Proctors play a key role in the fair and equitable administration of the MME Assessment.

Each public school and participating non-public school must designate the following Key MME testing staff: Test Supervisor, Backup Test Supervisor, and Test Accommodations Coordinator who meet the operational eligibility criteria to administer the MME. Each school building (or offsite testing facility) that administers the assessments must become established as an MME test center as approved by ACT. Key MME Staff were required to attend one of a series of MME administration workshops conducted in the fall and designed to provide a standardized presentation of all the information necessary for administering both the Standard Time and Accommodated Spring 2011 MME assessment. The following test administration manuals provided the basis for the Power Point used in the workshops:

- *Spring 2011 – ACT Supervisor's Manual – State Testing*,
- *Spring 2011 – ACT Supervisor's Manual – State Special Testing*,
- *Spring 2011 – Administration Instructions State – Allowed Accommodations*,
- *WorkKeys – Supervisor's Manual for State Testing*,
- *WorkKeys – Supervisor's Manual for State Testing – Special Testing*, and
- *MME Administration Manual Spring 2011*.

In addition to the Key MME Testing Staff, Room Supervisors and Proctors are used to actually adminsitjer the tests. School staff eligible to administer the MME Assessments must meet the following criteria:

- Test (and Backup) Supervisors – could NOT be related to any examinee taking the MME in 2010–2011 anywhere in Michigan.
- Room Supervisors and Proctors could NOT assist in a room where any relative was being tested.
- Test Accommodations Coordinators – could NOT be related to or be a guardian of any student taking the MME accommodations anywhere in Michigan during the testing year.
- Test Accommodations Coordinators—could Not be an athletic coach when any student athlete was tested with accommodations.
- Room Supervisors—could NOT be an athletic coach and administer the MME to a student athlete in a one-one-one situation (either Standard Time or Accommodated testing).

Relatives or wards include children/stepchildren, grandchildren, nieces, nephews, siblings, in-laws, spouses and persons under Key Staff's guardianship.

Depending on the number of students in each room, trained Room Supervisors and Proctors were assigned to assist Test Supervisors or Test Accommodations Coordinators. The following staffing guidelines were required: A Proctor may be used to assist a Room Supervisor or the Test Supervisor if fewer than 25 examinees are testing. A Proctor is required (in addition to the Room Supervisor) for every 25 examinees (or portion thereof) after the first 25 in a room. A Proctor is required (in addition to the Room Supervisor) for every 10 examinees testing with accommodations, after the first 10 in a room.

Roles and responsibilities of the Test Supervisor, Backup Test Supervisor, Test Accommodations Coordinator, Room Supervisor, and Proctor are specified in the *Spring 2011–ACT Supervisor's Manual – State Testing, Spring 2011– ACT Supervisor's Manual – State Special Testing*, *Spring 2011 Administration Instructions State-Allowed Accommodations*, *WorkKeys Supervisor's Manual for State Testing*, *WorkKeys Supervisor's Manual for State Testing – Special Testing,* and the Day 3 *MME Administration Manual for Standard Time Testing Spring* 2011, the Day 3 *MME Administration Manual Spring 2011* and the Day3 *MME Administration Manual Students Testing with Accommodations Spring 2011.*

Michigan is committed to assessing all public school students as required by state policy and federal law and has provided the opportunity for non-public schools to participate on a voluntary basis. Consequently all Michigan students identified as 11th graders in the Michigan School Data System (MSDS) were either required to test or given the opportunity to take the MME. All eligible 12th grade students were offered the opportunity to take MME.

Preparation for Test Administration

The Test Supervisor, Backup Test Supervisor, and Test Accommodations Coordinator are required to protect the integrity of all secure test m The Test Supervisor, Backup Test Supervisor, and Test Accommodations Coordinator are required to protect the integrity of all secure test materials and ensure that all examinees at their school are tested under the same conditions as examinees at every other school administering the MME.

<u>Qualifications and Requirements for Test Supervisors and Backup Supervisors include:</u>

1. **Not be related to or guardian of any examinee participating in State Testing with standard time anywhere in Michigan on either the initial or makeup test date this year. (Relatives or wards include children, stepchildren, grandchildren, nieces, nephews, siblings, in-laws, spouses, and persons under their guardianship.)**
2. Be proficient in English.

3. Be experienced in testing and measurement.
4. Be a staff member of the school.
5. Have control over locked, limited-access storage at the school to secure the test materials.
6. Ensure that the tests are administered in strict compliance with all policies and procedures as documented in each Supervisor's Manual (one for each day of testing).
7. Not be engaged in test preparation activities for the ACT or *WorkKeys* at any time during the current testing year (September through August), except as specifically required by school contract. The normal duties of a counselor or teacher are **not** a conflict of interest, provided they are part of job responsibilities specifically defined by one's employer and the employer is not a commercial enterprise.

Primary Responsibilities

1. Newly appointed Test Supervisors and Backup Test Supervisors must participate in a mandatory training session conducted by ACT, Measurement Incorporated and Michigan Department of Education staff. Workshop dates and times are provided to appropriate staff in correspondence from ACT.
2. Read and follow exactly all policies and procedures in each Supervisor's Manual (one for each day).
3. Arrange for all students to complete pre-test sections of their answer folders in a supervised session at school **before** test day. Affix barcode labels to examinee answer folders prior to test day.
4. Arrange for all students to test on the designated test dates with testing as the first activity of the morning. All room supervisors must begin reading the Verbal Instructions **no later than 9:00 a.m.**
5. Make arrangements for test rooms that meet standard testing requirements, including uncrowded seating facing the same direction, manageable security, good lighting and ventilation, adequate writing surfaces, and required space between examinees.
6. Ensure test rooms are free from distractions during the test session(s) (bells, public address system turned off, etc.) and separated from regular school activities.
7. Ensure standard time enrollment figures have been provided as directed.
8. Receive, check-in, and ensure security of test materials upon receipt until return. Take steps to protect materials from damage, theft, or loss, and from conditions that could allow prior access to the tests.
9. Identify a sufficient number of qualified assistants to serve as room supervisors and proctors. One room supervisor is required per room, plus one proctor for every 25 examinees in the room after the first 25. All testing staff must be proficient in English, may **not** be involved in commercial test preparation activities outside of normal school duties, and may not be enrolled in high school. No room supervisor or proctor may assist in a room where a relative is testing.
10. Conduct training for all room supervisors and proctors before the test dates, including a complete review of each Supervisor's Manual (one for each day).
11. Ensure all testing staff remain attentive to testing responsibilities throughout the entire administration, including accurate timing of test sections and breaks and monitoring for prohibited behavior.
12. Complete, verify, and return all required reports, seating diagrams, forms, answer folders, test booklets, and accommodated testing materials immediately after testing.
13. Document all irregularities and consult directly, and immediately if required, with ACT, the BAA, and Measurement, Inc., as appropriate, regarding actions to be taken.
14. Cooperate fully with ACT, the BAA, and Measurement, Inc., if applicable, to investigate and resolve suspected or documented irregularities.

Qualifications and Responsibilities for Test Accommodations Coordinators:

1. **Not be related to or guardian of any examinee participating in State Testing with Accommodations anywhere in Michigan this year during the two week testing window for accommodations. (Relatives or wards include children, stepchildren, grandchildren, nieces, nephews, siblings, in-laws, spouses, and persons under their guardianship.)**
2. Be proficient in English.
3. Be experienced in testing and measurement.
4. Be a staff member of the school.
5. Have control over locked, limited-access storage at the school to secure test materials.
6. Ensure that the tests are administered in strict compliance with all policies and procedures as documented in each Supervisor's Manual for accommodated testing (one for each day of testing).

To avoid the appearance of a conflict of interest and to protect both the examinee and testing staff from allegations of impropriety, the Test Accommodations Coordinator must also:

1. Not be a private consultant or individual tutor whose fees are paid by a student (or the student's family) for whom accommodations are requested.
2. Not be engaged in commercial test preparation activities at any time during the current testing year (September through August), except as specifically required by school contract. The normal duties of a counselor or teacher are **not** a conflict of interest, provided they are part of job responsibilities specifically defined by one's employer and the employer is not a commercial enterprise.
3. Not be involved in coaching high school or college athletics (applicable only if any student requesting accommodations participates in athletics). This qualification is in place to protect testing staff who receive and handle secure test materials and who administer the test to students individually or in very small groups without other testing staff present.

Primary Responsibilities

1. Determine which students need ACT-Approved Accommodations or State-Allowed Accommodations. Complete and submit ACT-Approved Accommodations applications and request State-Allowed Accommodations. The ACT-Approved Accommodations application requiresgathering required signatures and compiling supporting documentation. Determine accommodations needed for Day 2 and Day 3 and provide the information to the Test Supervisor who then orders the materials from the BAA Secure Site.
2. Provide timely response to requests from ACT for additional information about individual students.
3. Newly appointed Test Accommodations Coordinators must participate in a mandatory training session conducted by ACT, Measurement, Inc., and Michigan Department of Education staff. Workshop dates and times are provided to appropriate staff by ACT.
4. Train room supervisors and proctors who will assist with the administration of tests to students approved for accommodations.
5. Check-in all secure accommodated test materials and, in consultation with the Test Supervisor, maintain security of all accommodated testing materials.
6. Arrange for all students to complete pre-test sections of their answer folders in a supervised session at school **before** test day. Affix barcode labels to examinees answer folders prior to test day.
7. Arrange for all students to test within the designated accommodations testing window using only the authorized accommodations and materials assigned to each student.
8. Assign examinees to test rooms, separated by timing code with a room supervisor for each room. Students with different timing codes cannot test in the same room.
9. Complete, verify, and return all required reports, seating diagrams, forms, answer documents, and test booklets/alternate formats as directed immediately following the end of the accommodated testing timeframe.

10. Document all irregularities and consult directly with ACT, the BAA, or Measurement, Inc., as appropriate, regarding actions to be taken.
11. Cooperate fully with ACT, the BAA, or Measurement, Inc. to investigate and resolve suspected or documented irregularities.

MME Spring 2011 Test Administration Days

MME was administered to all eleventh graders and any twelfth graders who were eligible for a retest. For initial tests, Day 1 administration of the ACT Plus Writing was March 1, 2011, *WorkKeys* Day 2 administration was March 2, 2011, and Day 3 Michigan-developed Components administration was March 3, 2011. For makeup tests, Day 1 ACT Plus Writing was administered on March 15, 2011, *WorkKeys* Day 2 tests was administered on March 16, 2011 and Day 3 Michigan Components were administered on March 17, 2011. The dates for testing students with accommodations were as follows: Day 1 ACT Plus Writing: March 1-15, 2011; Day 2 WorkKeys: March 2-16; Day 3 Michigan-developed components: March 3-17.

MME Day 1 and Day 2: Materials Processing

Materials Orders—Day 1 and Day 2

ACT used enrollment numbers provided by BAA to calculate standard materials quantities for Day 1 and Day 2 of the MME. Materials quantities for accommodated students were produced using historical data from the previous year's accommodations application and request process for Day 1 and through the online ordering system for Day 2.

Shipping—Day 1 and Day 2

To provide the BAA with secure and dependable services for the shipping of Michigan assessment materials, ACT's Distribution Center maintains the quality and security of material distribution and return by using such methods as sealed trailers and hiring reputable carriers with the ability to immediately trace shipments. ACT uses all available tracking capabilities to provide status information and early opportunities for corrective action.

Standard time materials are packaged by school and addressed to the Test Supervisors. Accommodations materials are packaged by school and addressed to the Test Accommodations Coordinator. Each shipment to a school contains a shipping document listing each school's materials.

Final standard time Day 1 and Day 2 materials quantities are packaged using information provided by the Test Supervisors when ordering test materials through the BAA's secure website. Michigan educators also provide ACT with the Pre-Identification information needed to print barcode labels which are affixed to each answer document. Bar-coding of all secure materials during the pre-packaging effort allows the accurate tracking of these materials through the entire packing, delivery, and return process. It also enables ACT to inventory all materials throughout the packaging and delivery process and thus provide the BAA with status updates at any time. For the Spring 2011 testing, secure and nonsecure materials shipped for Day 1 and Day 2 are summarized in Tables 4.1 through 4.4.

Day 1 Processing

Table 4.1. Total Secure Documents Shipped in Spring 2011 Administration

Test Booklets	Quantities
ACT Multiple Choice-Initial	143,515
ACT Multiple Choice-Makeup	8,609
ACT Writing-Initial	148,220
ACT Writing-Makeup	8,609
Total Accommodations	11,249
Total	320,202

Table 4.2. Total Nonsecure Documents Shipped in Spring 2011 Administration

Answer Folders	Quantities
ACT Plus Writing-Initial	155,355
ACT Plus Writing-Makeup	8,826
Total	164,181

Day 2 Processing

Table 4.3. Total Secure Documents Shipped in Spring 2011 Administration

Test Booklets	Quantities
WorkKeys Initial	141,010
WorkKeys Makeup	9,184
WorkKeys Accommodations	12,800
Total	162,994

Table 4.4. Total Nonsecure Documents Shipped in Spring 2011 Administration

Answer Folders	Quantities
WorkKeys Initial	154,862
WorkKeys Makeup	8,931
Total	163,793

Receipt and Processing—Day 1 and Day 2

Each school's shipment included a copy of the packing list along with other shipping information to permit the accurate inventory of materials upon receipt by the Test Supervisor or the Test Accommodations Coordinator. Day 1 and Day 2 materials were shipped via a secure carrier, with traceable means, to pre-specified shipping addresses provided by the Test Supervisors and Test Accommodations Coordinators from each school. ACT standard shipping process does not allow for shipment to districts. These shipments were comprised of non-secure shipments followed by shipments of secure test materials. The nonsecure shipments included administration manuals, pre-printed barcode labels, and the answer documents students needed prior to the Day 1 and Day 2 assessments to complete the noncognitive sections of the ACT and *WorkKeys* in a supervised in-school pre-test session.

ACT requested each school's Test Supervisor or Test Accommodations Coordinator inventory the materials sent, verify the shipping contents and call ACT's toll-free number to report any shipping problems or materials shortages. Instructions were provided for secure storage of materials until test day.

Table 4.5. Number of Students Tested in Spring 2011 (ACT / *WorkKeys* / MME)

	MME Testing – Spring 2011
Day 1 - ACT	119,150
Day 2 - *WorkKeys*	118,546
Day 3 - MME	120,938

Test Security—Days 1 and 2

Secure test materials include all ACT and *WorkKeys* test booklets and used answer folders. The Test Supervisor is responsible for the security of all test materials from the time the carrier delivers them to the school to the time they are in the return carrier's possession. The Test Supervisor must protect the materials from damage, theft, or loss, and from conditions that could allow prior access to the tests.

Test materials must be kept in a locked, secure area, such as a vault or non-portable cabinet in a locked, limited-access room. Only the Test Supervisor, Backup Test Supervisor, Test Accommodations Coordinator, and possibly a few specifically authorized persons may have access to the area. If the security of test materials is compromised, ACT will not report scores.

ACT and *WorkKeys* test booklets are copyrighted and cannot be photocopied or used for any purpose other than testing. Under no circumstances is a test booklet seal to be broken by anyone other than the examinee as instructed on test day. Testing staff and examinees are prohibited from disclosing test questions, essay topics, or response choices to anyone.

Directions in the manuals note that testing staff who observe a student engaging in one or more prohibited behaviors should mark the student's answer folder VOID and complete an Irregularity Report. The Room Supervisor is instructed to immediately notify the Test Supervisor of the suspected prohibited behavior. Adequate Yearly Progress (AYP) requires the use of a valid assessment score. A student without a valid assessment score is considered "not assessed" for AYP purposes.

Materials Return—Day 1 and Day 2

Schools were provided with instructions and documentation for returning their materials.

The tracking numbers of the FedEx return labels provided to each school and those numbers were entered into our internal tracking system database.

Materials were prepared for return by the Test Supervisor. They packaged the materials and used the cartons pre-labeled for FedEx pick up. On the day after the initial test day, FedEx was dispatched to each school that had been sent Day 1 and 2 materials to retrieve test materials. This process was repeated for each school on the day after make-up testing.

For accommodated materials, all materials must be returned after the close of the accommodated testing window. Test Accommodations Coordinators are provided with specific return instructions similar to those provided to Test Supervisors for non-accommodated materials.

Supervisor manuals provide clear instructions on how to assemble, box, and return testing materials after test administration. Because of the criticality of used test materials and quantities often involved, safety is also a

major concern, not only for the materials but for the people moving them. Only single column boxes are used to distribute and collect test materials, so the weight of each carton is kept to a reasonable and manageable limit.

Preaddressed, prepaid labels are provided. The labels facilitate accurate and efficient sorting of each carton and its contents upon receipt.

Day 1 and Day 2 materials were returned directly to ACT.

Materials Discrepancy Checking Process—Day 1 and Day 2

ACT logged in the returned assessment materials from Day 1 and Day 2 of the MME during the check-in process. A check-in database was created for 2011-2012 to facilitate this process. The database tracked Day 1 and Day 2 standard time and accommodations materials. ACT followed up with schools to assure timely return of those testing materials, as well as tracked schools who did not return all testing materials.

Schools that have not returned any material
Detailed status reports are generated as test materials are received and checked in. These reports are monitored daily for missing or incomplete shipments, and follow-up occurs with schools missing materials.

Schools that have returned incomplete shipments
Detailed status reports, listing the number of boxes received from each school, are reviewed daily. An ACT team member will follow up with a phone call on quantities appearing to be less than expected as compared with FedEx tracking information.

Schools with missing secure test materials
After secure materials are scanned, reports indicating missing materials are generated. These reports identify materials and serial numbers and are provided to the ACT team for follow-up with the affected schools.

Schools Returning Answer Documents After Established 3/25/11 Cutoff Date
Documents were processed in accordance with late receipts processing guidelines mutually agreed upon by ACT and the BAA.

Processing Assessment Materials Returned by Schools—Day 1 and Day 2

ACT logged in the returned assessment materials from Day 1 and Day 2 of the MME during the check-in process within 24 hours of receipt and the answer folders were prepared for scanning within 72 hours of receipt. ACT followed up with schools to assure timely return of all testing materials, as well as to track schools that had not returned all Day 1 and Day 2 testing materials. The status of each school was readily discernable from the database updated by check-in staff. ACT utilized standard processing procedures for the Day 1 and Day 2 assessments, in terms of transferring the documents from the check-in process to the scanning process.

MME Day 3 Michigan Components: Materials Processing

Materials Orders—Day 3

Schools ordered all Day 3 materials through the BAA secure site. During the initial orders of materials, schools identified the number of standard-time assessment students and the number of accommodated students taking each format of the test.

For approved schools (type 1 only) that did not place initial orders, BAA used the greater of either the number of grade 11 students enrolled in the September SRSD file or the pre-ID count to place an order.

Appropriate quantities of materials for each initial order were packed and shipped to each school in three shipments, one containing non-secure materials; one containing secure standard-print materials; and one containing secure materials for students testing with accommodations.

The initial non-secure shipment to the Test Supervisor included: packing lists; box lists; pre-ID student barcode labels; blank labels; answer documents (only in shrink-wrapped packs of 10); standard-time administration manuals (1 per 15 students); and administration manuals for testing students with accommodations (1 per 10 extended time only students and 1 for every school ordering any accommodated materials).

The initial secure shipment of standard-print materials to the Test Supervisor included: packing lists; box lists; security lists; school header sheets; class/group ID sheets; test books; Administration Forms and Irregularity Forms envelopes and the materials return kits.

The initial secure shipment of accommodated materials included all accommodated format materials ordered, including standard-print Form 12 test books for students testing with extended time.

Schools placed additional orders for specific quantities of specific items, rather than the count of students testing. This was also true for orders for makeup materials. All additional orders were shipped to the Test Supervisor at the schools.

For all orders, Measurement Incorporated combined the pull of data and the processing of that data into one step. This provided immediate feedback to the BAA secure site about any orders that could not be filled immediately and eliminated duplicates of process data.

Picking and Packing—Day 3

Measurement Incorporated warehouse staff utilized a packaging application to generate an on-demand pick list. The pick list documented the specific materials and quantities to be included in an order, but not the exact barcode ranges for secure materials. After warehouse staff picked the materials for the order, the materials and pick list were delivered to a packing station. The staff member at the packing station initiated the packing process by scanning the order number on the pick list. The packing station employee used a hand scanner to capture the barcode value on each secure material. As each material was scanned into the order, the packing application verified that it was the correct material and kept a running count of the

quantity. After the packing station employee entered all order material information into the system, a validation check verified that the proper quantity of the correct materials had been packed.

All validation failures were displayed on the packing station screen. The packing station employee corrected any errors by scanning more materials into the order or removing materials by scanning the barcode of the material that needed to be removed. Once all validation failures had been corrected and the material types and quantities matched the order information, the packaging application printed a packing list to be included in the shipment. An additional list of secure items and their barcode numbers was included in each shipment containing secure materials.

Each packing station included a shipper tracking label printer to maintain order accuracy. By doing so, each order remained independent of other orders during packaging and sealing. The tracking label contained the order number and address of the recipient for verification against the packing list. After verifying the shipper tracking label against the packing list, each box was sealed with heavy-duty plastic tape and the shipper tracking label applied. The shipper tracking label contained the order number, school number, school name to which the materials were being delivered, and a Box n of X identifier to indicate the number of boxes shipped. At the time the shipping tracking number was created, the application created a corresponding entry in the tblFedExTracking table in the Measurement Incorporated database.

Shipping—Day 3

Measurement Incorporated monitored the distribution of materials to schools by FedEx. Test Supervisors were instructed to inventory all test materials sent, in order to assure that they received an adequate supply of assessment materials. Test Supervisors were asked to call Measurement Incorporated's Call Center via a toll-free telephone number to report any problems. Additional orders for materials were placed via the BAA secure site.

Table 4.6. Total Secure Documents Shipped in Spring 2011 Administration

Test Booklets	Quantities
MME Initial	139,121
MME Makeup	8,672
MME Accommodations	29,034
Total	176,827

Table 4.7. Total NonSecure Documents Shipped in Spring 2011 Administration

Answer Folders	Quantities
MME Initial and Makeup	153,040
MME Accommodations	23,500
Total	165,540

Materials Return—Day 3

Schools were provided with "Materials Return Kits" containing all of the necessary labels and documentation for returning the materials.

Materials were prepared for return by the Test Supervisors/Co-coordinators. They packaged the materials and applied the self-adhesive return labels that were supplied in the "Return Kits" from their original secure shipments. In order to retrieve materials immediately after testing, Measurement Incorporated used a pre-paid FedEx service for the return of all assessment materials.

On the day after the initial test day, FedEx was dispatched to each school that had been sent Day 3 materials to retrieve test materials. This process was repeated for each school on the day after make-up testing.

Note: the tracking numbers of each school's FedEx return labels were documented at the time of "Return Kit" production and entered into the Measurement Incorporated internal tracking system database. This process offered an accurate, expedient method of logging in materials upon return to Measurement Incorporated.

Materials Receipt and Processing—Day 3

Upon arrival at Measurement Incorporated, all boxes were scanned into the tracking system database where they were logged-in and checked against the pre-assigned school tracking numbers. This provided immediate information about the number of boxes received and their points of origin. Next, the boxes marked with a "Scorable" label were separated from the boxes marked with a "Non-Scorable" label. Boxes without either label were processed as "Scorable."

Scorable Materials—Day 3

The boxes labeled "Scorable" were opened first to remove used answer documents. These answer documents, along with school headers and any class/group ID sheets, were then placed into bar-coded tote boxes for IT Operations. (If there was no school header sheet, the Warehouse generated one, using information from the shipping label or from the answer documents.)

As the materials were transferred to IT Operations for scanning, the tracking number from the shipping box was scanned, along with the barcode for each associated tote box. This procedure provided a permanent link between the school's box from which the answer documents were received and the tote box that contained the answer documents from that box. When the tote box was full, a scan batch ID sheet was placed on top and was linked to the scan batch ID label on the tote box.

Non-Scorable Materials—Day 3

The boxes containing non-scorable materials were then examined for removal of any scorable materials that may have been incorrectly returned. A separate, or "redundancy," check was performed on each box by a second individual to assure that all scorable materials were discovered. Any scorable materials located during these searches were placed immediately into the appropriate tote boxes according to the procedure outlined for other scorable materials. These tote boxes of used answer documents were then forwarded to the IT Department for scanning and processing.

The security check-in process for the secure materials from the non-scorable boxes captured the security barcode number for each returned test booklet or accommodated format item. These materials were unpacked and then scanned at a workstation equipped with a barcode reader and a PC. The scanned secure materials were then packed in barcoded boxes that linked the barcodes of each secure item to that box. All items were scanned twice by two different teams in order to ensure quality of the process and results.

Note: if any boxes, scorable or non-scorable, contained MME Day 1 (ACT) or Day 2 (*WorkKeys*) materials, the worker first ensured that all Day 3 materials were removed from the box. The box was then sent to ACT, following procedures in the "Process for Handling Misdirected MME Materials" document.

A report that listed security barcodes present in the master database, but not found during check-in, was produced.

Overall, 99.73% of secure materials sent were returned and checked-in. More specifically, of 176,827 secure materials sent, 176,353 were checked-in. This left 474 items "missing" for further investigation at the school level by BAA and Measurement Incorporated.

Test Security—Day 3

Procedures related to Day 3 test security are identical to those on Days 1 and 2. Test materials must be kept in a locked, secure area, such as a vault or non-portable cabinet in a locked, limited-access room. Only the Test Supervisor, Back-up Test Supervisor, Test Accommodations Coordinator, and, possibly, a few specifically authorized persons may have access to the area. Test booklets cannot be photocopied or used for any purpose other than testing. Under no circumstances is a test booklet seal to be broken by anyone other than the examinee as instructed on test day. Testing staff and examinees are prohibited from disclosing test questions, essay topics, or response choices to anyone.

One difference between Days 1/2 and Day 3 is the procedure followed should a Room Supervisor observe a student engaging in unethical practices. For Days 1/2 testing, prohibited behavior results in voided answer documents. For Day 3 testing, the Room Supervisors mark the "Prohibited Behavior" circle on the answer document and return the answer document. Measurement Incorporated will scan, but not score, this answer document.

Scanning/Scoring—Day 3

Once logged into the Operations Department, the scan bins were shifted to the cutting area, where one scan bin at a time was removed from the cart for cutting. The cutting operation converted the multi-page answer document into a stack of single sheets ready for scanning.

Note: when the answer documents were printed, each sheet was imprinted with a lithocode value unique to that document. Both a scannable and human-readable version of the lithocode were printed on every sheet of every answer document. In the unlikely event that a scan bin was dropped at the cutting or pre-scanning stage, the unique lithocode allowed the answer documents to be reassembled, and the answer document integrity to be verified at the scanner and project database once the data was transferred. Software validations at the scanner ensured that all pages of each student's answer document were accounted for; thus, any pages that were out of order could be easily corrected prior to any further processing.

Measurement Incorporated image scanned all pages of a student's answer document at the same time using BancTec XDS IntelliScan color image scanners. The BancTec XDS IntelliScan is rated to scan 190 sheets per minute at an optical resolution of 240 dots per inch (DPI) and creates both JPEG and TIFF images for every page. These scanners utilize precision camera assemblies pressurized to minimize dust. This, plus low maintenance LED camera illumination, reduced the need for rescans. The scanner features a completely open paper path to dramatically improve document throughput. This paper path reduced the time to recover from paper jams and other common complications for scanners with more restrictive paper paths. Both sonic and vacuum double-sheet detection technology ensured that every sheet was scanned. In addition, BancTec has designed custom document integrity software for Measurement Incorporated. This application detects

out-of-sequence pages. The scanner stops to allow operator correction before imaging, thus eliminating post scanning corrective action.

To ensure that all sheets in the scan bin were scanned, the last sheet in every bin was an "End of Batch" sheet. If the End of Batch record did not appear in the data file, an error alert was generated, and a technician made a visual check of the scan bin to verify that all answer documents had been scanned. If necessary, the data file was re-opened, and any missing sheet(s) appended to the file, thereby creating a complete data file.

Data Correction

Once all of the scanned data was combined to create the student records, data validation routines were executed. These routines analyzed the data and created error tables for answer documents containing questionable data. Common error detection routines included checks for the following situations:

- Inconsistencies in school, grade, or form
- Inconsistencies in headers and answer documents
- Duplicate student barcodes within the same bin or another bin of answer documents
- Missing student barcodes
- Missing or incomplete demographics (such as a blank name)
- Double marks in the demographic and/or multiple-choice grids

Measurement Incorporated utilized a double data correction process. Data correction operators used the Measurement Incorporated data correction application that retrieved flagged data records and highlighted the problem field on a computer screen so it could be resolved. The operator compared the highlighted data to the scanned image of the answer document, making any necessary corrections. Once an operator corrected a flagged record, the same flagged record was routed to a second data correction operator who repeated the data correction process. After a flagged record was edited by two operators, the data correction application checked that both operators made identical corrections. In the event that the two corrections differed, the record was routed to a supervisory staff member for a third and final resolution. This process continued until all flagged records were examined.

To ensure accuracy, once a correction was written to the database, the document was validated again to ensure that the corrected edit had not created another error. All edits were recorded and tracked in Measurement Incorporated databases, along with the user ID of the staff member making the edits.

Multiple-Choice Scoring

After all flagged data was reviewed and corrected, student selected responses were scored against the item answer keys. The Test Maps table, called tblTestMaps, was pulled from the BAA database in Michigan. That data was converted into a set of 36 records, each with its own set of correct answers (or answer key strings). Then, those answer keys were applied to student responses to produce a string of ones and zeroes, indicating right and wrong answers. A validation process (key check) was used to detect any potential answer key problems. The students' selected responses and correct answer indicators were transmitted to MDE in a data file.

In all, 117,963 scored MME Day 3 student records were in the data files transmitted to MDE.

Score Reporting

The basis for all the report data was a student data "Match File" transmitted to Measurement Incorporated from BAA.

The reports included: Individual Student Report, Parent Report, Student Roster, Student Record Label, ISD Comprehensive Report and District Comprehensive Report, State Demographic Report, ISD Demographic Report, District Demographic Report, School Demographic Report, State Summary Report, District Summary Report, and School Summary Report.

Measurement Incorporated provided each of the reports as a static or dynamic Adobe Acrobat PDF on the Electronic Report Hosting website. These PDF files were electronically transmitted to the Michigan MME. The PDF files were divided into batches based on the report type, and the PDF files of each batch were placed in their own sub-directories. HOVS produced PDF files separated by school. In addition to the electronic distribution of reports, the PDF of reports was printed and distributed to the schools.

For schools in districts that selected the green option for reporting, the Individual Student Reports, Parent Reports, and Student Record Labels were printed. All other reports were available only online as PDFs.

The PDFs were extensively reviewed before the preliminary reports were printed and mailed. Labels and district reports were printed inline and sorted with the related school reports. Labels were printed on inventory label material. The reports were segmented by color card stock.

Reports packages were shrink-wrapped and packaged for traceable ground delivery throughout the state. Depending on the size of the report, the reports were placed in appropriate shipping boxes or envelopes. The packages were matched against a distribution list for accuracy and completeness of the cycle run.

Description of Reports

Parent Report

The Parent Report presented individual test results for all students in grades 11 and 12 who tested in a subject.

The Parent Report contained the following information:

- Scale score for each subject area
- Performance level for each subject area
- Subscore values for each subscore strand for each subject area, including the number of points the student earned, the number of points possible, and the percent correct
- Text, including a letter from the superintendent, performance level definitions, subject descriptions, assessment descriptions, and ACT and *WorkKeys* descriptions
- Scale score graphs
- ACT test scores
- Work Keys level scores

The Parent Report provided information for the following subjects: MME Reading, MME Writing, MME Mathematics, MME Science, and MME Social Studies.

Individual Student Report

The Individual Student Report (ISR) provided a detailed description of each student's performance in the subject areas assessed by the MME. This report was designed to help educators identify the academic strengths of their students and the areas that may need improvement. Schools may include these reports in student record files.

The Individual Student Report contained the following information:

- Scale score for each subject area
- Performance level for each subject area
- Subscore values for each subscore strand for each subject area, including the number of points the student earned, the number of points possible, and the percent correct
- ACT test scores
- Work Keys level scores

The Individual Student Report provided information for the following subjects: MME Reading, MME Writing, MME Mathematics, MME Science, and MME Social Studies.

Student Roster

The Student Roster presented individual test results for all students in grades 11 and 12 who tested in a subject. It listed those students by class/group who took the test in the subject - regardless of what form they took.

The Student Roster contained the following information for each subject area:

- Scale Score
- Performance Level
- Subscore values for each subscore strand, including the number of points the student earned and the number of points possible

The last line of each subject of the report showed the number of students assessed, defined as the number of students reported on the roster for that group.

The Student Roster provided information for the following subjects: MME Reading, MME Writing, MME Mathematics, MME Science, and MME Social Studies.

Student Record Labels

The Student Record Label provided a summary description of each student's performance in the subject areas assessed on the MME.

The Student Record Label consisted of the following information for each student:

- Demographic information

- Scale score and performance level for subjects tested

Student Labels provided student information in different subjects in the following order:

- Reading
- Writing
- Mathematics
- Science
- Social Studies

State Demographic Report, ISD Demographic Report, District Demographic Report, School Demographic Report

The Demographic Report was a statistical summary of twenty student demographic areas for all the subjects in a grade, aggregated in a student group. There were eighteen types of student groups, arrived at by combining the following modes, populations, and grades:

Modes:
- State
- District
- School

Student Populations:
- All Students
- Students with Disabilities
- All Except Students with Disabilities

Grades:
- 11
- 12

The Demographic Report provided data for the following subjects: MME Reading, MME Writing, MME Mathematics, MME Science, and MME Social Studies.

In calculating the percentage of students with scale scores at a certain performance level, both the numerator and denominator were expressed as a float, and the result of that calculation was rounded in the manner of the SQL function ROUND (numeric_expression, length).

State Summary Report, District Summary Report, School Summary Report

The Summary Report consisted of two pages:
1 - A summary of performance levels achieved, compared to previous years
2 - A distribution of scores by subject and strand in a grade

Both of the report pages were produced for all eighteen types of student groups reported, which were arrived at by combining the following modes, populations, and grades:

Modes:
- State
- District
- School

Student Populations:
- All Students
- Students with Disabilities
- All Except Students with Disabilities

Grades:
- 11
- 12

The Summary Reports provided data for the following subjects: MME Reading, MME Writing, MME Mathematics, MME Science, and MME Social Studies.

Any mean score in this report was the average score calculated by summing the applicable scores and dividing that sum by the total number of those scores. Percentages were calculated by dividing the number in a category by the total number of students assessed. In any division calculation, both the numerator and denominator were expressed as a float, and the result of that calculation was rounded in the manner of the SQL function ROUND (numeric_expression, length).

ISD Comprehensive Report and District Comprehensive Report

The Comprehensive Report provided summary score data by subject and grade for public schools, aggregated in a student group. The District Comprehensive Report listed data for the district, followed by data for each school within the district. The ISD Comprehensive Report listed data for the ISD, followed by data for each district.

There were twelve types of student groups, arrived at by combining the following modes, populations, and grades:

Modes:
- ISD
- District

Student Populations:
- All Students
- Students with Disabilities
- All Except Students with Disabilities

Grades:
- 11
- 12

The Comprehensive Report provided data for the following subjects: MME Reading, MME Writing, MME Mathematics, MME Science, and MME Social Studies.

In calculating the mean scale score or the percentage of students with scale scores at a certain performance level, both the numerator and denominator were expressed as a float, and the result of that calculation was rounded in the manner of the SQL function ROUND (numeric_expression, length).

Accommodations for Students with Disabilities (SWD) and English Language Learners (ELL)

All students are to participate in the assessment programs approved by the State Board of Education. For some students, accommodations that are customarily used during routine classroom activities may be considered for use during the administration of the MME assessments. The State Board of Education has approved standard and nonstandard assessment accommodations for the Michigan Educational Assessment System including MME, MI-Access, and ELPA.

The MME Accommodation Summary Table (Table 4.8 beginning next page) identifies standard and nonstandard accommodations for students with disabilities, Section 504 students, and/or students with limited English proficiency (also referred to as English language learners, or ELL). Standard accommodations do ***not*** change the construct that the assessment is measuring and ***do*** provide a valid score. Nonstandard accommodations change the construct that the assessment is measuring, rendering scores that are not valid. Accommodations not listed in the table are considered nonstandard.

The Michigan Merit Examination (MME) consists of three major components administered over three days: the ACT Plus Writing, three *WorkKeys* tests (*Reading for Information*, *Applied Mathematics*, and *Locating Information*), and Michigan developed items for mathematics, science and social studies. Table 4.8 outlines the Spring 2011 test organization.

Table 4.9 outlines which components contribute to each MME score. The MME scores will play a role in qualifying for the Michigan Promise scholarship and will be the foundation for the No Child Left Behind (NCLB) calculation of Adequate Yearly Progress (AYP) and EdYES! accountability reports for high schools.

Spring 2011
Michigan Merit Examination (MME) Accommodations Summary Table

The Michigan Merit Examination (MME) consists of three major components administered over three days: the ACT Plus Writing, three *WorkKeys* tests (Reading for Information, Applied Mathematics, and Locating Information), and Michigan-developed items for mathematics, science, and social studies. The chart below outlines the Spring 2011 test organization.

Spring 2011 Test Organization

Day*	Assessment	Subject Session	Number of Parts	Total Items	Testing Time (minutes)	Estimated Time Required for Administration
Day 1 March 1 (Makeup March 15)	**ACT Plus Writing**	English	5	75 MC items	45	Total test administration time - including check in, instructions, breaks, and collection of materials – 5 hours
		Mathematics		60 MC items	60	
		Reading		40 MC items	35	
		Science		40 MC items	35	
		Writing		1 Prompt	30	
Day 1 Standard Testing Time 205 minutes (3 hours / 25 minutes)						
Day 2 March 2 (Makeup March 16)	***WorkKeys***	Reading for Information	3	33 MC Items	45	Total test administration time - including check in, instructions, breaks, and collection of materials – 3.5 hours
		Applied Mathematics		33 MC Items	45	
		Locating Information		38 MC items	45	
Day 2 Standard Testing Time 135 minutes (2 hour / 15 minutes)						
Day 3 March 3 (Makeup March 17)	**Michigan Components**	Mathematics	3	25 MC items	30	Total test administration time - including check in, instructions, breaks, and collection of materials – 2.5 hours
		Science		49 MC items	35	
		Social Studies		48 MC items	35	
Day 3 Standard Testing Time 100 minutes (1 hour / 40 minutes)						
				TOTAL MINUTES	440	
				TOTAL HOURS	7.33	

*More detailed information about this schedule and the MME program is available on the MME Web page at www.michigan.gov/mme.

The chart below outlines which components contribute to each MME score. The MME scores are required for the No Child Left Behind (NCLB) calculation of Adequate Yearly Progress (AYP) and accountability reports for high schools.

			Table 4.9 Components Contributing to MME Scores				
Day	**Test**	**Subject Session**	**Reading**	**Writing**	**Mathematics**	**Science**	**Social Studies**
Day 1	ACT Plus Writing	English		Selected items			
		Mathematics			Selected items		
		Reading	Selected items				
		Science				Selected items	
		Writing		ALL			
Day 2	*WorkKeys*	Reading for Information	Selected items				
		Applied Mathematics			Selected items		
		Locating Information			Selected items		Selected items
Day 3	Michigan Components	Mathematics			ALL		
		Science				ALL	
		Social Studies					ALL

MME Test Accommodations Window and Testing Staff Requirements

All accommodated testing must be administered within the two-week window that begins on the initial test date for that component of the MME and ends on the makeup date for that component. Testing may be scheduled on any days during the window, but each student must take the tests in prescribed order – all of Day 1 (the ACT Plus Writing) in order, followed by the Day 2 *WorkKeys* tests in order, followed by the Day 3 Michigan components in order. All testing staff must meet ACT's requirements. If testing occurs outside the authorized window, or with procedures that conflict with ACT directions, or under supervision of testing staff who do not meet ACT's requirements, then the answer documents will not be scored. If the misadministration is discovered after scoring, then the scores will be canceled.

ACT-Approved vs. State-Allowed Accommodations on the ACT (Day 1 of the MME)

ACT is committed to ensuring that official ACT scores reported to colleges and other entities from MME testing are comparable to scores earned through other forms of ACT testing involving the application of ACT's test accommodations policies. Therefore, ACT supports the following two forms of accommodations on the ACT when it is administered as Day 1 of the MME:

1) **ACT-Approved accommodations** result in ACT scores that are fully reportable to colleges, scholarships, and other entities *in addition to* being used for MME scores. Only students with professionally diagnosed and documented disabilities who receive accommodations in school should apply for ACT-approved accommodations.
2) **"State-Allowed" accommodations** result in ACT scores that are not college reportable; they are used only for MME scores. English language learners who do not have a disability but receive accommodations in school should request State-Allowed accommodations.

Requesting Accommodations on the ACT (Day 1 of the MME)

In general, all accommodations on the ACT must be requested and reviewed by ACT. However, there are limited exceptions. For example, because testing will normally occur at the local school rather than a separate test center, some arrangements do **not** require review or prior approval from ACT (e.g., placement at the front of the room). Such arrangements are noted on the attached accommodations summary table as "local decision" meaning they do **not** require ACT review or approval.

All schools must appoint a Test Accommodations Coordinator (TAC) who will submit requests for accommodations to ACT. The TAC has access to two different forms specifically designed for the MME administration of the ACT:

1) ***ACT-Approved Accommodations*** – This form is used to request ACT approval of accommodations on the MME for students who meet ACT eligibility requirements. (See information about ACT's review of these requests in the next section below.)
2) ***State-Allowed Accommodations*** – This form is used to order test materials for students who will test with "State-Allowed" accommodations. These students are those who do not meet ACT's eligibility requirements (e.g., English language learners with no disabilities) or whose requests for ACT approval have been denied. ACT will ship the materials ordered for each student; no review or approval process will be conducted.

ACT Review of Requests for ACT-Approved Accommodations on the ACT (Day 1 of the MME)

ACT will review requests for ACT-Approved Accommodations by applying the Americans with Disabilities Act (ADA) standards that are used for all such requests. Not every request for an accommodation listed on the attached accommodations summary table as available will be approved. Approval is dependent on submission of all required documentation by the stipulated deadline and review by ACT. It is possible for ACT to approve an accommodation for one student, while the same accommodation may be denied for a different student. ACT's decision whether to approve the requested accommodations under the ADA will determine whether resulting ACT scores can be reported to colleges *in addition to* being used for MME scores.

Ordering State-Allowed Accommodations Materials for the ACT (Day 1 of the MME)

Students who do not meet ACT eligibility requirements (e.g., English language learners with no disabilities) or whose requested accommodations are denied by ACT have two options: 1) Test under standard conditions and receive college reportable ACT scores, or 2) submit an order for "State-Allowed" accommodations materials resulting in ACT scores that are NOT college reportable. IMPORTANT NOTE: TACs must submit an order for "State-Allowed" accommodations for each applicable student so that ACT can ship the correct ACT test materials – which are *different* from those used by examinees testing with ACT-Approved accommodations.

ACT scores resulting from testing with "State-Allowed" accommodations are **not** college reportable, but will be used for MME scores. Thus, some students will achieve ACT scores that are college reportable because their accommodations have been approved by ACT, while others using the same accommodations will achieve ACT scores that are *not* college reportable because their use of those accommodations was not approved by ACT.

Ordering test materials for students testing with accommodations on MME Day 2 (*WorkKeys*) and Day 3 (Michigan Components)
All accommodated test materials, including extended time test booklets, for MME Day 2 and Day 3 must be ordered from the BAA Secure Site December 1, 2010 – January 11, 2011. There is no request or approval form for accommodations on Day 2 and Day 3. Testing with accommodations on Day 2 and/or Day 3 is a local decision based on the student's regular instruction and supported by the student's IEP, 504 Plan, or ELL instruction. ACT's approval of accommodations applies **only** to materials for and the administration of the ACT Plus Writing (Day 1). Because there is no issue of reporting scores to colleges, schools may provide accommodations on the *WorkKeys* and Michigan components of the MME consistent with the accommodations listed in the "MME Day 2 or Day 3" columns of the attached accommodations summary table, even if the student tests without those accommodations on the ACT. Accommodated test materials for MME Day 2 and Day 3 must be ordered on the BAA Secure Site. Please pay close attention to whether the accommodation is standard or non-standard, and what impact the accommodation may have on student eligibility for National Career Readiness Certification (NCRC) and/or *WorkKeys* score results.

WorkKeys National Career Readiness Certificate (NCRC) Eligible Scores
WorkKeys scores achieved during Day 2 of the MME **may** be eligible for the NCRC provided a student satisfies the criteria associated with one of the 4 achievement levels. Four levels of achievement are possible: Bronze, Silver, Gold, or Platinum, based on scores earned on the three *WorkKeys* tests.

WorkKeys-Ineligible Accommodations on Day 2 *WorkKeys*
WorkKeys scores will not be issued for students using accommodations marked as "No" under the column "MME Day 2." Use of these accommodations will also result in ineligibility for the National Career Readiness Certificate (NCRC). HOWEVER, the student will receive valid MME scores if the accommodation is designated as "MME eligible". Accommodations eligible for MME scores are also indicated by a standard accommodation in the IEP/504 or ELL column, and the accommodation is supported by the student's IEP, 504 Plan, or ELL instruction. Students testing with accommodations identified as "MME ineligible" will not be counted as assessed and will not receive MME scores in the affected subject areas.

Standard/Nonstandard Accommodations on MME Day 3
It is important to know whether an assessment accommodation is considered standard or nonstandard for the MME since it can have an impact whether a school or district meets the No Child Left Behind (NCLB) Adequate Yearly Progress (AYP). At the district, school, and subgroup (ethnicity, economically disadvantaged, English Language Learners, and Students with Disabilities) levels, a minimum of 95% of the students enrolled in Grade 11 must be assessed, either on the MME or MI-Access.

A standard assessment accommodation is one that does not change what the specific assessment is measuring. The score achieved by a student using a standard assessment accommodation does count when calculating NCLB participation and proficiency rates. A nonstandard assessment accommodation changes what the assessment is measuring and results in an invalid score. **If a student takes the MME Day 2 or Day 3 test using a nonstandard accommodation, the student will not count as assessed and will not receive MME scores in the affected subjects.** Please see the attached MME accommodations summary table to determine if the student's accommodation is standard or nonstandard for MME Day 2 or Day 3. There is one column for accommodations supported by an IEP or 504 Plan. There is a separate column for accommodations supported by a student's ELL instruction.

Any assessment accommodations not listed in the MME Accommodations Summary Table are considered nonstandard.

Use of accommodations on any section of the MME, Day 1, Day 2, and/or Day 3 must be recorded on the student answer document for that day, following instructions in the MME Day 1 and Day 2 Answer Folder Supplements, *WorkKeys Supervisor's Manual for State Testing – Special Testing* and the *MME Day 3 Administration Manual for Students Testing with Accommodations*.

Important Day 2 notes for TACs –

1. Students testing with accommodations must use the *WorkKeys* test books with blue covers. They will have three test form numbers in the lower left hand side, each beginning with an "R". Test forms beginning with any other letter are **not** to be used for testing with accommodations.
2. For each student tested with accommodations, the TAC must be sure that an admin code is written and gridded on Page 3 of the student's answer folder for each test administered with accommodations.
3. For each student tested with accommodations, the TAC must provide responses to the Local Items in Box 26 on Page 2 of the answer folder. BAA cannot determine eligibility/ineligibility for MME scores without a response to Local Items in Box 26.
4. Because only the instructions are translated on the Reading for Information test, a "700" admin code should not be entered for this test. TACs must record the proper Admin Code from the 500 or 600 range. Admin codes in the 700 range are valid for the Applied Mathematics and Locating Information tests.

Reading the MME Accommodations Summary Table

The attached Spring 2011 Michigan Merit Examination (MME) Accommodations Summary Table is arranged in columns, explained below.

Column	Explanation
Accommodation	Each accommodation that appears on the Assessment Accommodations Summary Table approved by the Michigan State Board of Education is listed.
MME Day 1 (The ACT Plus Writing)	
May Request	ACT has indicated whether or not each accommodation may be requested for the ACT Plus Writing (Day 1 of the MME), or whether State-Allowed accommodated formats may be ordered instead. • Accommodations for which local decisions may be made without a request to ACT are specifically noted. • Some formats or accommodations are noted as State-Allowed only. • A few accommodations are not permitted for the ACT. If any "not permitted" accommodations are used, ACT scores will not be issued. • Some accommodations do not apply to the ACT. Additional details about some accommodations may be needed before a decision can be made for an individual request.
ACT Comments	These comments clarify ACT's understanding of each accommodation and any associated restrictions for Day 1.
College Reportable ACT Scores	ACT has noted whether each accommodation that requires approval will result in ACT scores that are fully reportable to colleges and other entities ***when approved by ACT for an individual student with disabilities***. If specific restrictions must be met or documentation from the test administration provided, these are also noted. The use of accommodations that require approval and which have not been approved by ACT for an individual student are eligible for State-Allowed accommodations testing. Taking the ACT Plus Writing with State-Allowed accommodations will result in ACT scores that are reportable ***only*** for MME scores. If a student uses a combination of accommodations and ***any*** of those accommodations are State-Allowed (not ACT-Approved), the resulting scores will ***not*** be college reportable but can be used for MME Scores. NOTE 1: State-Allowed accommodations must be requested (ordered) from ACT for that student so that the school will receive Day 1 accommodated test materials assigned to that student. NOTE 2: The use of accommodations considered Standard (S) for MME Day 2 or Day 3 will result in valid MME scores that may be used for school accountability. This is true in combination with both ACT-Approved and State-Allowed accommodations.
MME Day 2 (*WorkKeys*)	
WorkKeys Eligible Scores	ACT has indicated whether each accommodation is eligible (Yes) or not eligible (No) for *WorkKeys* scores. Some formats or accommodations do not apply to *WorkKeys*, and some require additional details before a determination can be made. The use of a *WorkKeys*-ineligible accommodation must be reported on the *WorkKeys* State Testing Answer Folder through the use of specific 3-digit Administration Codes. Additionally, MDE will review administrators' responses to the *WorkKeys* Local Items, as recorded in Box 26 of the *WorkKeys* State Testing Answer Folders, to determine eligibility for MME scores. For a list of those Admin codes and further instruction please see page 5 of the Day 2 *WorkKeys State*

	Testing Supervisor's Manual for Special State Testing. Use of *WorkKeys*-ineligible accommodations must be described in detail on the Testing Time Verification Form and Administration Report on page 89 of the Day 2 *WorkKeys Supervisor's Manual for State Testing – Special Testing*. NOTE: The use of accommodations considered Standard (S) will result in valid MME scores that may be used for school accountability.
MME Day 3 Standard/Nonstandard	
	With respect to valid MME scores, MDE has indicated whether each accommodation is considered Standard (S) or Nonstandard (NS) for Day 3 of the MME. Separate notations have been made for two groups – IEP/504 and ELL. Assessment accommodations not listed in the MME Accommodations Summary Table are considered nonstandard. A standard assessment accommodation is one that does not change what the specific assessment is measuring. The score achieved by a student using a standard assessment accommodation does count when calculating NCLB participation and proficiency rates. A nonstandard assessment accommodation changes what the assessment is measuring and results in an invalid score. **If a student takes the MME Day 3 using a nonstandard accommodation, the student will not count as assessed and will not receive MME scores in the affected subjects.** Please see the attached MME accommodations summary table to determine if the student's accommodation is standard or nonstandard for MME Day 3. There is one column for accommodations supported by an IEP or 504 Plan. There is a separate column for accommodations supported by a student's ELL instruction. Use of a nonstandard accommodation on any section of MME Day 3 must be reported on the student answer folder following instructions in the *MME Day 3 Administration Manual for Students Testing with Accommodations*.

Spring 2011 Michigan Merit Examination (MME) Accommodations Summary Table						
Accommodation	MME Day 1 (The ACT Plus Writing)			MME Day 2	MME Day 3 Standard/Nonstandard	
	May Request	ACT Comments	College Reportable ACT Scores [3]	*WorkKeys* Eligible Scores[2]	IEP/504	ELL
A. Timing/Scheduling						
1. Administration of the assessment at a time most beneficial to the student, with appropriate supervision	Yes	Must be within the designated two-week window that begins on initial state test day for ACT Plus Writing and ends on the makeup day for ACT Plus Writing. Testing may be scheduled for any days during the window, but each student must take the sections of the ACT Plus Writing in prescribed order: English, Mathematics, Reading, Science, and finally, Writing.	Yes	Yes Must be within the designated two-week window for Day 2. Day 2 testing may not begin until Day 1 testing is complete.	S	S

1. "Yes" in the "College Reportable ACT Scores" column means ACT scores will be college reportable ONLY IF that accommodation: 1) is shown as "local decision" or 2) was approved by ACT for an individual student **with a disability.** If the accommodation requires approval, but was not approved by ACT, the ACT scores achieved using that accommodation will not be college reportable. ACT scores are college reportable only if ALL accommodations that require ACT approval are ACT-Approved for that student.
2. Under the "MME Day 2" column, ACT has indicated whether each accommodation is eligible (Yes) or not eligible (No) for WorkKeys score results. Accommodations designated "No" are also ineligible for the National Career Readiness Certificate (NCRC). HOWEVER, the student will receive valid MME scores if the accommodation is designated "MME eligible." Accommodations eligible for MME scores are also indicated by a standard accommodation in the IEP/504 or ELL column, and the accommodation is supported by the student's IEP, 504 Plan, or ELL instruction. Accommodations identified as "MME ineligible" will not count as assessed and will not receive MME scores in the affected subjects.

Spring 2011 Michigan Merit Examination (MME) Accommodations Summary Table

Accommodation	MME Day 1 (The ACT Plus Writing)			MME Day 2	MME Day 3 Standard/Nonstandard	
	May Request	**ACT Comments**	**College Reportable ACT Scores [3]**	***WorkKeys* Eligible Scores[2]**	**IEP/504**	**ELL**
2. Administer the parts within a content area in any order	No	ACT tests must **always** be administered in prescribed sequence.	No	No *WorkKeys* tests must **always** be administered in prescribed sequence. MME Eligible -- TACs are required to enter the appropriate Admin code from the range of 900-954	S	S
3. Extended assessment time NOTE: All MME tests are timed. Timing codes are assigned by ACT for Day 1. For Day 2 and Day 3, schools may allow time-and-a-half, double time, or a maximum of 3 hours for each test.	Yes	Normally time-and-a-half in single self-paced session using regular or large-print. For certain formats and disabilities, ACT will assign a timing code for the ACT based on the test format and disability, up to triple time (and testing over multiple days, one test per day). Oral presentation (e.g., cassette, audio DVD, or reader), and Braille normally *require* triple time.	Yes Only if testing complies with timing code assigned by ACT	Yes Only if *WorkKeys* timing guidelines are followed.	S	S
4. Frequent supervised breaks	Yes	Interpreted as "stop-the-clock" breaks; normally available only with standard time. If requested with extended time, must provide documentation to support need for "stop-the-clock" breaks *in addition to* extended time.	Yes	Yes	S	S

1. "Yes" in this column means ACT scores will be college reportable ONLY IF that accommodation: 1) is shown as "local decision" or 2) was approved by ACT for an individual student **with a disability.** If the accommodation requires approval, but was not approved by ACT, the ACT scores achieved using that accommodation will not be college reportable. ACT scores are college reportable only if ALL accommodations that require ACT approval are ACT-Approved for that student.
2. Under the "MME Day 2" column, ACT has indicated whether each accommodation is eligible (Yes) or not eligible (No) for WorkKeys score results. Accommodations designated "No" are also ineligible for the National Career Readiness Certificate (NCRC). HOWEVER, the student will receive valid MME scores if the accommodation is designated "MME eligible." Accommodations eligible for MME scores are also indicated by a standard accommodation in the IEP/504 or ELL column, and the accommodation is supported by the student's IEP, 504 Plan, or ELL instruction. Accommodations identified as "MME ineligible" will not count as assessed and will not receive MME scores in the affected subjects.

Spring 2011 Michigan Merit Examination (MME) Accommodations Summary Table						
Accommodation	MME Day 1 (The ACT Plus Writing)			MME Day 2	MME Day 3 Standard/Nonstandard	
	May Request	ACT Comments	College Reportable ACT Scores [3]	*WorkKeys* Eligible Scores[2]	IEP/504	ELL
5. Method of informing students of remaining time (e.g., clock or timer)	Local decision- but must adhere to all ACT directions	Five minutes remaining announcement routinely part of verbal instructions for all students on ACT. Students approved for time extensions on the ACT are given hourly announcements of time. **No other assistance in monitoring time is allowed.**	Yes	Yes Must adhere to directions in Manual.	S	S
B. Setting						

1. "Yes" in this column means ACT scores will be college reportable ONLY IF that accommodation: 1) is shown as "local decision" or 2) was approved by ACT for an individual student **with a disability.** If the accommodation requires approval, but was not approved by ACT, the ACT scores achieved using that accommodation will not be college reportable. ACT scores are college reportable only if ALL accommodations that require ACT approval are ACT-Approved for that student.
2. Under the "MME Day 2" column, ACT has indicated whether each accommodation is eligible (Yes) or not eligible (No) for WorkKeys score results. Accommodations designated "No" are also ineligible for the National Career Readiness Certificate (NCRC). HOWEVER, the student will receive valid MME scores if the accommodation is designated "MME eligible." Accommodations eligible for MME scores are also indicated by a standard accommodation in the IEP/504 or ELL column, and the accommodation is supported by the student's IEP, 504 Plan, or ELL instruction. Accommodations identified as "MME ineligible" will not count as assessed and will not receive MME scores in the affected subjects.

Spring 2011 Michigan Merit Examination (MME) Accommodations Summary Table

Accommodation	MME Day 1 (The ACT Plus Writing)			MME Day 2	MME Day 3 Standard/Nonstandard	
	May Request	**ACT Comments**	**College Reportable ACT Scores [3]**	***WorkKeys* Eligible Scores[2]**	**IEP/504**	**ELL**
6. Administration of the assessment in an alternate education setting (in school) with appropriate supervision e.g., • Bilingual/English as a Second Language (ESL) setting • Special education setting • In a distraction free space or alternate location such as separate room or location within the room	Local decision unless requesting off-site	If setting is off-site, appropriate off-site application must be approved by ACT.	Yes	Yes If off-site application for Day 1 is approved by ACT.	S	S
7. Administration of the assessment in an alternate education setting (out of school) with appropriate supervision e.g., • Home when student is homebound • Care facility when it is medically necessary	Yes	Appropriate off-site or home-bound application must be approved by ACT.	Yes	Yes If off-site or home-bound application for Day 1 is approved by ACT.	S If off-site or home-bound application for Day 1 is approved by ACT.	S If off-site or home-bound application for Day 1 is approved by ACT.

1. "Yes" in this column means ACT scores will be college reportable ONLY IF that accommodation: 1) is shown as "local decision" or 2) was approved by ACT for an individual student **with a disability.** If the accommodation requires approval, but was not approved by ACT, the ACT scores achieved using that accommodation will not be college reportable. ACT scores are college reportable only if ALL accommodations that require ACT approval are ACT-Approved for that student.
2. Under the "MME Day 2" column, ACT has indicated whether each accommodation is eligible (Yes) or not eligible (No) for WorkKeys score results. Accommodations designated "No" are also ineligible for the National Career Readiness Certificate (NCRC). HOWEVER, the student will receive valid MME scores if the accommodation is designated "MME eligible." Accommodations eligible for MME scores are also indicated by a standard accommodation in the IEP/504 or ELL column, and the accommodation is supported by the student's IEP, 504 Plan, or ELL instruction. Accommodations identified as "MME ineligible" will not count as assessed and will not receive MME scores in the affected subjects.

Spring 2011 Michigan Merit Examination (MME) Accommodations Summary Table

Accommodation	MME Day 1 (The ACT Plus Writing)			MME Day 2	MME Day 3 Standard/Nonstandard	
	May Request	**ACT Comments**	**College Reportable ACT Scores [3]**	***WorkKeys* Eligible Scores[2]**	**IEP/504**	**ELL**
8. Administration of the assessment in an **interim** alternative education setting (out of school) with appropriate supervision (e.g., juvenile facility)	Local decision unless requesting off-site	If setting is off-site, appropriate off-site application must be approved by ACT.	Yes	Yes If off-site application for Day 1 is approved by ACT.	S If off-site application for Day 1 is approved by ACT	S If off-site applicatio n for Day 1 is approved by ACT
9. Administration of the assessment individually or in a small group	Local decision unless requesting off-site or required by approved accommod ation	If setting is off-site, appropriate off-site application must be approved by ACT. Note that individual testing is **required** for selected accommodations (e.g., if approved accommodations could disturb others or if approved for a reader).	Yes	Yes If off-site application for Day 1 is approved by ACT.	S	S
10. Placement of student where he/she is most comfortable (e.g., front of the room, back of the room)	Local decision unless requesting off-site	If setting is off-site, appropriate off-site application must be approved by ACT.	Yes	Yes If off-site application for Day 1 is approved by ACT.	S	S
11. Use of accommodated seating, special lighting, or furniture	Local decision	Provided by the school.	Yes	Yes For ELL students this accommodation is not eligible for MME purposes. A response to Local Item #2 is required	S	NS

1. "Yes" in this column means ACT scores will be college reportable ONLY IF that accommodation: 1) is shown as "local decision" or 2) was approved by ACT for an individual student **with a disability.** If the accommodation requires approval, but was not approved by ACT, the ACT scores achieved using that accommodation will not be college reportable. ACT scores are college reportable only if ALL accommodations that require ACT approval are ACT-Approved for that student.
2. Under the "MME Day 2" column, ACT has indicated whether each accommodation is eligible (Yes) or not eligible (No) for WorkKeys score results. Accommodations designated "No" are also ineligible for the National Career Readiness Certificate (NCRC). HOWEVER, the student will receive valid MME scores if the accommodation is designated "MME eligible." Accommodations eligible for MME scores are also indicated by a standard accommodation in the IEP/504 or ELL column, and the accommodation is supported by the student's IEP, 504 Plan, or ELL instruction. Accommodations identified as "MME ineligible" will not count as assessed and will not receive MME scores in the affected subjects.

Spring 2011 Michigan Merit Examination (MME) Accommodations Summary Table

Accommodation	MME Day 1 (The ACT Plus Writing)			MME Day 2	MME Day 3 Standard/Nonstandard	
	May Request	**ACT Comments**	**College Reportable ACT Scores [3]**	***WorkKeys* Eligible Scores[2]**	**IEP/504**	**ELL**
12. Able to move, stand or pace during assessment in a manner where others' work cannot be seen and is not distracting to others (e.g., kneeling, constant movement)	Local decision		Yes	Yes	S	S
13 Use of concentration aids (e.g., stress balls, T-stools, background music or noise buffers)	Submit details with request	Requests considered individually based on documentation submitted. Music and noise buffers not normally approved. Approval and reportable status depend on detailed information about the tools/buffers proposed for use.	Depends on details	Yes If approved for Day 1, may also be used for Day 2. For ELL students this accommodation is not eligible for MME purposes. A response to Local Item #2 is required	S	NS
14. Placement of teacher/proctor near student	Local decision		Yes	Yes	S	S

1. "Yes" in this column means ACT scores will be college reportable ONLY IF that accommodation: 1) is shown as "local decision" or 2) was approved by ACT for an individual student **with a disability.** If the accommodation requires approval, but was not approved by ACT, the ACT scores achieved using that accommodation will not be college reportable. ACT scores are college reportable only if ALL accommodations that require ACT approval are ACT-Approved for that student.
2. Under the "MME Day 2" column, ACT has indicated whether each accommodation is eligible (Yes) or not eligible (No) for WorkKeys score results. Accommodations designated "No" are also ineligible for the National Career Readiness Certificate (NCRC). HOWEVER, the student will receive valid MME scores if the accommodation is designated "MME eligible." Accommodations eligible for MME scores are also indicated by a standard accommodation in the IEP/504 or ELL column, and the accommodation is supported by the student's IEP, 504 Plan, or ELL instruction. Accommodations identified as "MME ineligible" will not count as assessed and will not receive MME scores in the affected subjects.

Spring 2011 Michigan Merit Examination (MME) Accommodations Summary Table

Accommodation	MME Day 1 (The ACT Plus Writing)			MME Day 2	MME Day 3 Standard/Nonstandard	
	May Request	**ACT Comments**	**College Reportable ACT Scores [3]**	***WorkKeys* Eligible Scores[2]**	**IEP/504**	**ELL**
C. Presentation						
15. Reading all assessment directions in student's native language • Student must be dominant in that native language; **and** • Student's English proficiency is determined to be basic or lower intermediate; **and** • Student receives bilingual instruction in their native language for the maintenance of that language	Yes (State-Allowed only)	Includes spoken instructions <u>and</u> directions printed in the test booklets. If student's reason for accommodations is English language proficiency, student must order "State-Allowed" accommodations materials.	No	Yes TACs are required to enter the appropriate Admin code from the range of 550-554 or 600-640	S	S
16. Qualified person familiar to the student administers the assessment (e.g., Special Education Teacher, Bilingual/ESL staff)	Local decision-staff must meet all ACT requirements	Only if not a relative or athletic coach (if student is an athlete). Only if all directions for test administration are read verbatim **in English** with no clarifications in another language.	Yes	Yes Staff qualifications for Day 1 apply.	S	S

1. "Yes" in this column means ACT scores will be college reportable ONLY IF that accommodation: 1) is shown as "local decision" or 2) was approved by ACT for an individual student **with a disability.** If the accommodation requires approval, but was not approved by ACT, the ACT scores achieved using that accommodation will not be college reportable. ACT scores are college reportable only if ALL accommodations that require ACT approval are ACT-Approved for that student.
2. Under the "MME Day 2" column, ACT has indicated whether each accommodation is eligible (Yes) or not eligible (No) for WorkKeys score results. Accommodations designated "No" are also ineligible for the National Career Readiness Certificate (NCRC). HOWEVER, the student will receive valid MME scores if the accommodation is designated "MME eligible." Accommodations eligible for MME scores are also indicated by a standard accommodation in the IEP/504 or ELL column, and the accommodation is supported by the student's IEP, 504 Plan, or ELL instruction. Accommodations identified as "MME ineligible" will not count as assessed and will not receive MME scores in the affected subjects.

Spring 2011 Michigan Merit Examination (MME) Accommodations Summary Table

Accommodation	MME Day 1 (The ACT Plus Writing)			MME Day 2	MME Day 3 Standard/Nonstandard	
	May Request	**ACT Comments**	**College Reportable ACT Scores [3]**	***WorkKeys* Eligible Scores[2]**	**IEP/504**	**ELL**
17A. Assessment directions • Teacher may emphasize key words in directions • Teacher may repeat directions exactly as worded in administrator manual	Yes	Directions in the test booklet not normally read aloud can be read to students. Permitted for college reportable ACT scores only if approved for reader or audio version of test and directions are read **verbatim** in English. Emphasis only as marked in the printed directions; must be read verbatim without signals regarding right or wrong.	Yes	Yes	S	S
17B. Assessment directions • Student may restate directions in his/her own words • Student may ask for clarification of directions	Yes (State - Allowed only)	Only if tested individually.	No	Yes Only if tested individually.	S	S
18. Teacher provides visual, auditory, or physical cues to student to begin, maintain, or finish task	Yes	If cues will disturb other examinees, must test individually.	Yes	Yes If cues will disturb other examinees, must test individually.	S	S

1. "Yes" in this column means ACT scores will be college reportable ONLY IF that accommodation: 1) is shown as "local decision" or 2) was approved by ACT for an individual student **with a disability.** If the accommodation requires approval, but was not approved by ACT, the ACT scores achieved using that accommodation will not be college reportable. ACT scores are college reportable only if ALL accommodations that require ACT approval are ACT-Approved for that student.
2. Under the "MME Day 2" column, ACT has indicated whether each accommodation is eligible (Yes) or not eligible (No) for WorkKeys score results. Accommodations designated "No" are also ineligible for the National Career Readiness Certificate (NCRC). HOWEVER, the student will receive valid MME scores if the accommodation is designated "MME eligible." Accommodations eligible for MME scores are also indicated by a standard accommodation in the IEP/504 or ELL column, and the accommodation is supported by the student's IEP, 504 Plan, or ELL instruction. Accommodations identified as "MME ineligible" will not count as assessed and will not receive MME scores in the affected subjects.

Spring 2011 Michigan Merit Examination (MME) Accommodations Summary Table						
Accommodation	MME Day 1 (The ACT Plus Writing)			MME Day 2	MME Day 3 Standard/Nonstandard	
	May Request	ACT Comments	College Reportable ACT Scores [3]	*WorkKeys* Eligible Scores[2]	IEP/504	ELL
19. Reading aloud the English, Reading, and Writing assessments to an individual student using a readers script	Yes	ACT-produced audio version must be used or reader's script read verbatim in **English**. For college reportable ACT scores, student must test individually if not using audio version with headset (see #22 for audio version).	Yes	Yes ACT-produced audio version used or readers script read verbatim in English.	S	S
20. Reading aloud the Mathematics, Science, and Social Studies assessment to an individual student using a readers script	Yes	ACT-produced audio version must be used or reader's script read verbatim in **English**. For college reportable ACT scores, student must test individually if not using audio version with headset (see #22 for audio version).	Yes	Yes ACT-produced audio version used or readers script read verbatim in English.	S	S

1. "Yes" in this column means ACT scores will be college reportable ONLY IF that accommodation: 1) is shown as "local decision" or 2) was approved by ACT for an individual student **with a disability.** If the accommodation requires approval, but was not approved by ACT, the ACT scores achieved using that accommodation will not be college reportable. ACT scores are college reportable only if ALL accommodations that require ACT approval are ACT-Approved for that student.
2. Under the "MME Day 2" column, ACT has indicated whether each accommodation is eligible (Yes) or not eligible (No) for WorkKeys score results. Accommodations designated "No" are also ineligible for the National Career Readiness Certificate (NCRC). HOWEVER, the student will receive valid MME scores if the accommodation is designated "MME eligible." Accommodations eligible for MME scores are also indicated by a standard accommodation in the IEP/504 or ELL column, and the accommodation is supported by the student's IEP, 504 Plan, or ELL instruction. Accommodations identified as "MME ineligible" will not count as assessed and will not receive MME scores in the affected subjects.

Spring 2011 Michigan Merit Examination (MME) Accommodations Summary Table						
Accommodation	MME Day 1 (The ACT Plus Writing)			MME Day 2	MME Day 3 Standard/Nonstandard	
	May Request	ACT Comments	College Reportable ACT Scores [3]	*WorkKeys* Eligible Scores[2]	IEP/504	ELL
21. Reading content and questions in the student's native language (Mathematics, Social Studies, Science, and Writing) • Student must be dominant in a native language other than English; **and** • Student's English proficiency is determined to be basic or lower intermediate; **and** • Student receives bilingual instruction in that native language for the maintenance of that language	Yes (State-Allowed only)	If student's reason for accommodations is English language proficiency, student must order "State-Allowed" accommodations materials.	No	No MME eligible TACs are required to enter the appropriate Admin code from the range of 900-954	S	S
22. Use of state-produced audio versions of the assessments	Yes	Must use headset if testing in a group. ACT produces audio version.	Yes	Yes ACT produces audio version.	S	NA Refer to conditions listed in #23

1. "Yes" in this column means ACT scores will be college reportable ONLY IF that accommodation: 1) is shown as "local decision" or 2) was approved by ACT for an individual student **with a disability.** If the accommodation requires approval, but was not approved by ACT, the ACT scores achieved using that accommodation will not be college reportable. ACT scores are college reportable only if ALL accommodations that require ACT approval are ACT-Approved for that student.
2. Under the "MME Day 2" column, ACT has indicated whether each accommodation is eligible (Yes) or not eligible (No) for WorkKeys score results. Accommodations designated "No" are also ineligible for the National Career Readiness Certificate (NCRC). HOWEVER, the student will receive valid MME scores if the accommodation is designated "MME eligible." Accommodations eligible for MME scores are also indicated by a standard accommodation in the IEP/504 or ELL column, and the accommodation is supported by the student's IEP, 504 Plan, or ELL instruction. Accommodations identified as "MME ineligible" will not count as assessed and will not receive MME scores in the affected subjects.

Spring 2011 Michigan Merit Examination (MME) Accommodations Summary Table

Accommodation	MME Day 1 (The ACT Plus Writing)			MME Day 2	MME Day 3 Standard/Nonstandard	
	May Request	**ACT Comments**	**College Reportable ACT Scores [3]**	***WorkKeys* Eligible Scores[2]**	**IEP/504**	**ELL**
23. Use of state-produced video or audio versions of assessment in English, for English language learners • Student must be dominant in a native language other than English; **and** • Student's English proficiency is determined to be basic or lower intermediate.	Yes (State-Allowed only)	If student's reason for accommodations is English language proficiency, student must order "State-Allowed" accommodations materials. ACT produces video or audio versions for the state.	No	Yes ACT produces video or audio versions for the state.	S	S

1. "Yes" in this column means ACT scores will be college reportable ONLY IF that accommodation: 1) is shown as "local decision" or 2) was approved by ACT for an individual student **with a disability.** If the accommodation requires approval, but was not approved by ACT, the ACT scores achieved using that accommodation will not be college reportable. ACT scores are college reportable only if ALL accommodations that require ACT approval are ACT-Approved for that student.
2. Under the "MME Day 2" column, ACT has indicated whether each accommodation is eligible (Yes) or not eligible (No) for WorkKeys score results. Accommodations designated "No" are also ineligible for the National Career Readiness Certificate (NCRC). HOWEVER, the student will receive valid MME scores if the accommodation is designated "MME eligible." Accommodations eligible for MME scores are also indicated by a standard accommodation in the IEP/504 or ELL column, and the accommodation is supported by the student's IEP, 504 Plan, or ELL instruction. Accommodations identified as "MME ineligible" will not count as assessed and will not receive MME scores in the affected subjects.

Spring 2011 Michigan Merit Examination (MME) Accommodations Summary Table						
Accommodation	MME Day 1 (The ACT Plus Writing)			MME Day 2	MME Day 3 Standard/Nonstandard	
	May Request	ACT Comments	College Reportable ACT Scores [3]	*WorkKeys* Eligible Scores[2]	IEP/504	ELL
24. Use of state-produced video or audio versions of assessment in a language other than English for English language learners • Student must be dominant in that language; **and** • Student's English proficiency is determined to be basic or lower intermediate; **and** • Student receives bilingual instruction in that native language for the maintenance of that language.	Yes (State-Allowed only)	If student's reason for accommodations is English language proficiency, student must order "State-Allowed" accommodations materials. ACT produces video DVDs in Spanish and Arabic for the state. No other languages are offered.	No	Yes ACT produces video DVDs for the Applied Mathematics and Locating Information tests in Spanish and Arabic for the state. No other languages are offered. Student will receive valid *WorkKeys* scores but not be eligible for NCRC. Is eligible for MME TACs are required to enter the appropriate Admin code from the range of 730-742 for Applied Mathematics and Locating Information. For Reading for Information an Admin Code from the 500 or 600 range should be recorded. Reading score is eligible for NCRC.	S	S

1. "Yes" in this column means ACT scores will be college reportable ONLY IF that accommodation: 1) is shown as "local decision" or 2) was approved by ACT for an individual student **with a disability.** If the accommodation requires approval, but was not approved by ACT, the ACT scores achieved using that accommodation will not be college reportable. ACT scores are college reportable only if ALL accommodations that require ACT approval are ACT-Approved for that student.
2. Under the "MME Day 2" column, ACT has indicated whether each accommodation is eligible (Yes) or not eligible (No) for WorkKeys score results. Accommodations designated "No" are also ineligible for the National Career Readiness Certificate (NCRC). HOWEVER, the student will receive valid MME scores if the accommodation is designated "MME eligible." Accommodations eligible for MME scores are also indicated by a standard accommodation in the IEP/504 or ELL column, and the accommodation is supported by the student's IEP, 504 Plan, or ELL instruction. Accommodations identified as "MME ineligible" will not count as assessed and will not receive MME scores in the affected subjects.

Spring 2011 Michigan Merit Examination (MME) Accommodations Summary Table

Accommodation	MME Day 1 (The ACT Plus Writing)			MME Day 2	MME Day 3 Standard/Nonstandard	
	May Request	**ACT Comments**	**College Reportable ACT Scores [3]**	***WorkKeys* Eligible Scores[2]**	**IEP/504**	**ELL**
25. Directions provided using sign language (American Sign Language (ASL) or Exact English)	Yes	Applies only to spoken instructions exactly as provided in the administration manual.	Yes	Yes Day 1 restrictions apply.	S	S
26 Sign the Reading and Writing assessments in: • American Sign Language (ASL) • Exact English Signing (EES)	Yes – ASL (State-Allowed only) Yes – EES	Signing of items with American Sign Language (ASL) is not ACT-approved. Exact English Signing (EES) of test items may be requested and approved in specific cases for college reportable scores.	No – ASL Yes – only if EES approved by ACT	No ASL not permitted. Use of ASL will result in student not receiving *WorkKeys* scores or MME scores. TACs are required to enter the appropriate Admin code from the range of 900-954 and must provide a response to Local Item Question #2. Yes – only if EES. For ELL students, this accommodation is not eligible for MME purposes. A response to Local Item #2 is required	NS - ASL S - EES	NS – ASL NS – EES

1. "Yes" in this column means ACT scores will be college reportable ONLY IF that accommodation: 1) is shown as "local decision" or 2) was approved by ACT for an individual student **with a disability.** If the accommodation requires approval, but was not approved by ACT, the ACT scores achieved using that accommodation will not be college reportable. ACT scores are college reportable only if ALL accommodations that require ACT approval are ACT-Approved for that student.
2. Under the "MME Day 2" column, ACT has indicated whether each accommodation is eligible (Yes) or not eligible (No) for WorkKeys score results. Accommodations designated "No" are also ineligible for the National Career Readiness Certificate (NCRC). HOWEVER, the student will receive valid MME scores if the accommodation is designated "MME eligible." Accommodations eligible for MME scores are also indicated by a standard accommodation in the IEP/504 or ELL column, and the accommodation is supported by the student's IEP, 504 Plan, or ELL instruction. Accommodations identified as "MME ineligible" will not count as assessed and will not receive MME scores in the affected subjects.

Spring 2011 Michigan Merit Examination (MME) Accommodations Summary Table						
Accommodation	MME Day 1 (The ACT Plus Writing)			MME Day 2	MME Day 3 Standard/Nonstandard	
	May Request	ACT Comments	College Reportable ACT Scores [3]	*WorkKeys* Eligible Scores[2]	IEP/504	ELL
27. Sign the Listening and Speaking sections of ELPA (American Sign Language (ASL) or Exact English)	NA			NA		
28. Sign the Reading and Writing sections of the ELPA (American Sign Language (ASL) or Exact English)	NA			NA		

1. "Yes" in this column means ACT scores will be college reportable ONLY IF that accommodation: 1) is shown as "local decision" or 2) was approved by ACT for an individual student **with a disability.** If the accommodation requires approval, but was not approved by ACT, the ACT scores achieved using that accommodation will not be college reportable. ACT scores are college reportable only if ALL accommodations that require ACT approval are ACT-Approved for that student.
2. Under the "MME Day 2" column, ACT has indicated whether each accommodation is eligible (Yes) or not eligible (No) for WorkKeys score results. Accommodations designated "No" are also ineligible for the National Career Readiness Certificate (NCRC). HOWEVER, the student will receive valid MME scores if the accommodation is designated "MME eligible." Accommodations eligible for MME scores are also indicated by a standard accommodation in the IEP/504 or ELL column, and the accommodation is supported by the student's IEP, 504 Plan, or ELL instruction. Accommodations identified as "MME ineligible" will not count as assessed and will not receive MME scores in the affected subjects.

Spring 2011 Michigan Merit Examination (MME) Accommodations Summary Table

Accommodation	MME Day 1 (The ACT Plus Writing)			MME Day 2	MME Day 3 Standard/Nonstandard	
	May Request	**ACT Comments**	**College Reportable ACT Scores [3]**	***WorkKeys* Eligible Scores[2]**	**IEP/504**	**ELL**
29. Sign the Mathematics, Locating Information, Science and Social Studies assessments in: • American Sign Language (ASL) • Exact English Signing (EES)	Yes – ASL (State-Allowed only) Yes - EES	Signing of items with American Sign Language (ASL) is not ACT-approved. Exact English Signing (EES) of test items may be requested and approved in specific cases for college reportable scores.	No – ASL Yes – only if EES approved by ACT	No ASL not permitted Use of ASL will result in student not receiving *WorkKeys* scores or MME scores. TACs are required to enter the appropriate Admin code from the range of 900-954 and must provide a response to Local Item Question #2. Yes – only if EES For ELL students, this accommodation may not be eligible for MME purposes. A response to Local Item #2 is required	NS - ASL S - EES	NS - ASL NS - EES
30. Use of calculator/talking calculator on the noncalculator sections of the Mathematics assessment	NA	There are no "noncalculator" sections of the ACT Mathematics test.	NA	NA	NA	NA
31. Use of a calculator on the Science and Social Studies assessments	No	Calculators are permitted only on the ACT Mathematics, not any other tests.	No	NA	NS	NS

1. "Yes" in this column means ACT scores will be college reportable ONLY IF that accommodation: 1) is shown as "local decision" or 2) was approved by ACT for an individual student **with a disability.** If the accommodation requires approval, but was not approved by ACT, the ACT scores achieved using that accommodation will not be college reportable. ACT scores are college reportable only if ALL accommodations that require ACT approval are ACT-Approved for that student.
2. Under the "MME Day 2" column, ACT has indicated whether each accommodation is eligible (Yes) or not eligible (No) for WorkKeys score results. Accommodations designated "No" are also ineligible for the National Career Readiness Certificate (NCRC). HOWEVER, the student will receive valid MME scores if the accommodation is designated "MME eligible." Accommodations eligible for MME scores are also indicated by a standard accommodation in the IEP/504 or ELL column, and the accommodation is supported by the student's IEP, 504 Plan, or ELL instruction. Accommodations identified as "MME ineligible" will not count as assessed and will not receive MME scores in the affected subjects.

Spring 2011 Michigan Merit Examination (MME) Accommodations Summary Table

Accommodation	MME Day 1 (The ACT Plus Writing)			MME Day 2	MME Day 3 Standard/Nonstandard	
	May Request	**ACT Comments**	**College Reportable ACT Scores [3]**	***WorkKeys* Eligible Scores[2]**	**IEP/504**	**ELL**
32. Use of arithmetic tables	No	Arithmetic tables are not allowed for the ACT.	No	No All examinees may use the ACT-developed *WorkKeys* Formula Sheet provided only on the *WorkKeys* Applied Mathematics test. Use of any other table will result in student not receiving *WorkKeys* scores or MME scores. TACs are required to enter the appropriate Admin code from the range of 900-954 and must provide a response to Local Item Question #2.	NS	NS
33. Use of manipulatives (e.g., actual coins and bills, base 10 blocks, and concrete objects)	NA	Items do not involve this kind of manipulation.	NA	No For ELL students, this accommodation is not eligible for MME purposes. A response to Local Item #2 is required TACs are required to enter the appropriate Admin code from the range of 900-954 and provide a response to Local Item #2	S	NS

1. "Yes" in this column means ACT scores will be college reportable ONLY IF that accommodation: 1) is shown as "local decision" or 2) was approved by ACT for an individual student **with a disability.** If the accommodation requires approval, but was not approved by ACT, the ACT scores achieved using that accommodation will not be college reportable. ACT scores are college reportable only if ALL accommodations that require ACT approval are ACT-Approved for that student.
2. Under the "MME Day 2" column, ACT has indicated whether each accommodation is eligible (Yes) or not eligible (No) for WorkKeys score results. Accommodations designated "No" are also ineligible for the National Career Readiness Certificate (NCRC). HOWEVER, the student will receive valid MME scores if the accommodation is designated "MME eligible." Accommodations eligible for MME scores are also indicated by a standard accommodation in the IEP/504 or ELL column, and the accommodation is supported by the student's IEP, 504 Plan, or ELL instruction. Accommodations identified as "MME ineligible" will not count as assessed and will not receive MME scores in the affected subjects.

Spring 2011 Michigan Merit Examination (MME) Accommodations Summary Table

Accommodation	MME Day 1 (The ACT Plus Writing)			MME Day 2	MME Day 3 Standard/Nonstandard	
	May Request	**ACT Comments**	**College Reportable ACT Scores [3]**	***WorkKeys* Eligible Scores[2]**	**IEP/504**	**ELL**
34. Use of an abacus	Yes	Provided by school or student; student must test individually.	Yes	Yes For ELL students, this accommodation is not eligible for MME purposes. A response to Local Item #2 is required	S	NS
35. Use of rulers as provided by the State	NA	Items do not require rulers.	NA	NA	NA	NA
36. Use of adapted rulers, protractors, Braille and large print rulers and protractors.	NA	Items do not require rulers or protractors.	NA	NA	S	NS
37. Use of auditory amplification devices or special sound systems	Submit details with request	Used only for spoken instructions. Requests considered individually based on documentation submitted. Approval and reportable status depend on detailed information about proposed devices or systems.	Depends on details	Yes If approved for Day 1, may also be used for Day 2. For ELL students, this accommodation may not be eligible for MME purposes. A response to Local Item #2 is required	S	NS

1. "Yes" in this column means ACT scores will be college reportable ONLY IF that accommodation: 1) is shown as "local decision" or 2) was approved by ACT for an individual student **with a disability.** If the accommodation requires approval, but was not approved by ACT, the ACT scores achieved using that accommodation will not be college reportable. ACT scores are college reportable only if ALL accommodations that require ACT approval are ACT-Approved for that student.
2. Under the "MME Day 2" column, ACT has indicated whether each accommodation is eligible (Yes) or not eligible (No) for WorkKeys score results. Accommodations designated "No" are also ineligible for the National Career Readiness Certificate (NCRC). HOWEVER, the student will receive valid MME scores if the accommodation is designated "MME eligible." Accommodations eligible for MME scores are also indicated by a standard accommodation in the IEP/504 or ELL column, and the accommodation is supported by the student's IEP, 504 Plan, or ELL instruction. Accommodations identified as "MME ineligible" will not count as assessed and will not receive MME scores in the affected subjects.

Spring 2011 Michigan Merit Examination (MME) Accommodations Summary Table

Accommodation	MME Day 1 (The ACT Plus Writing)			MME Day 2	MME Day 3 Standard/Nonstandard	
	May Request	**ACT Comments**	**College Reportable ACT Scores [3]**	***WorkKeys* Eligible Scores[2]**	**IEP/504**	**ELL**
38. Use of visual aids (e.g., closed circuit television, magnification devices)	Yes	Provided by school or student. Depending on the device used, student may have to test individually.	Yes	Yes For ELL students, this accommodation is not eligible for MME purposes. A response to Local Item #2 is required	S	NS
39. Use of state-produced Braille and enlarged print versions of assessment	Yes	ACT produces Braille and enlarged print versions.	Yes	Yes ACT produces Braille and enlarged print versions. For ELL students, this accommodation is not eligible for MME purposes. A response to Local Item #2 is required	S	NS
40. Use of a page turner	Local decision-staff must meet all ACT requiremen ts	Page turner must meet same requirements as all testing staff.	Yes	Yes Staff qualifications for Day 1 apply. For ELL students, this accommodation is not eligible for MME purposes. A response to Local Item #2 is required	S	NS

1. "Yes" in this column means ACT scores will be college reportable ONLY IF that accommodation: 1) is shown as "local decision" or 2) was approved by ACT for an individual student **with a disability.** If the accommodation requires approval, but was not approved by ACT, the ACT scores achieved using that accommodation will not be college reportable. ACT scores are college reportable only if ALL accommodations that require ACT approval are ACT-Approved for that student.
2. Under the "MME Day 2" column, ACT has indicated whether each accommodation is eligible (Yes) or not eligible (No) for WorkKeys score results. Accommodations designated "No" are also ineligible for the National Career Readiness Certificate (NCRC). HOWEVER, the student will receive valid MME scores if the accommodation is designated "MME eligible." Accommodations eligible for MME scores are also indicated by a standard accommodation in the IEP/504 or ELL column, and the accommodation is supported by the student's IEP, 504 Plan, or ELL instruction. Accommodations identified as "MME ineligible" will not count as assessed and will not receive MME scores in the affected subjects.

Spring 2011 Michigan Merit Examination (MME) Accommodations Summary Table

Accommodation	MME Day 1 (The ACT Plus Writing)			MME Day 2	MME Day 3 Standard/Nonstandard	
	May Request	**ACT Comments**	**College Reportable ACT Scores [3]**	***WorkKeys* Eligible Scores[2]**	**IEP/504**	**ELL**
41. Use of non-skid surface that will not damage the answer document or scanning equipment (DO NOT use tape or other adhesive)	Yes	Provided by school or student.	Yes	Yes For ELL students, this accommodation is not eligible for MME purposes. A response to Local Item #2 is required	S	NS
42. Use of acetate colored shield, highlighters, highlighter tape, page flags, and reading guides on test booklets.	Local decision	Provided by school or student. "Reading guides" are interpreted as place-keepers. May require student to test individually (e.g., highlighters).	Yes	Yes	S	S
43. Use of bilingual dictionaries that define or explain words or terms	Yes (State-Allowed only)	Provided by school or student.	No	No This accommodation is not MME eligible TACs are required to enter the appropriate Admin code from the range of 900-954 and must provide a response to Local Item Question #2.	NS	NS

1. "Yes" in this column means ACT scores will be college reportable ONLY IF that accommodation: 1) is shown as "local decision" or 2) was approved by ACT for an individual student **with a disability.** If the accommodation requires approval, but was not approved by ACT, the ACT scores achieved using that accommodation will not be college reportable. ACT scores are college reportable only if ALL accommodations that require ACT approval are ACT-Approved for that student.
2. Under the "MME Day 2" column, ACT has indicated whether each accommodation is eligible (Yes) or not eligible (No) for WorkKeys score results. Accommodations designated "No" are also ineligible for the National Career Readiness Certificate (NCRC). HOWEVER, the student will receive valid MME scores if the accommodation is designated "MME eligible." Accommodations eligible for MME scores are also indicated by a standard accommodation in the IEP/504 or ELL column, and the accommodation is supported by the student's IEP, 504 Plan, or ELL instruction. Accommodations identified as "MME ineligible" will not count as assessed and will not receive MME scores in the affected subjects.

Spring 2011 Michigan Merit Examination (MME) Accommodations Summary Table							
Accommodation		MME Day 1 (The ACT Plus Writing)			MME Day 2	MME Day 3 Standard/Nonstandard	
		May Request	ACT Comments	College Reportable ACT Scores [3]	*WorkKeys* Eligible Scores[2]	IEP/504	ELL
44.	Use of dictionary, thesaurus, spelling book, or grammar book for Mathematics, Science, Social Studies, Reading, and Writing	Yes (State-Allowed only)	Provided by school or student.	No	No This accommodation is not MME eligible TACs are required to enter the appropriate Admin code from the range of 900-954 and must provide a response to Local Item Question #2.	NS	NS
45.	Use of bilingual word-for-word non-electronic translation glossary for English language learners	Yes (State-Allowed only)	Provided by school or student.	No	Yes	S	S

1. "Yes" in this column means ACT scores will be college reportable ONLY IF that accommodation: 1) is shown as "local decision" or 2) was approved by ACT for an individual student **with a disability.** If the accommodation requires approval, but was not approved by ACT, the ACT scores achieved using that accommodation will not be college reportable. ACT scores are college reportable only if ALL accommodations that require ACT approval are ACT-Approved for that student.
2. Under the "MME Day 2" column, ACT has indicated whether each accommodation is eligible (Yes) or not eligible (No) for WorkKeys score results. Accommodations designated "No" are also ineligible for the National Career Readiness Certificate (NCRC). HOWEVER, the student will receive valid MME scores if the accommodation is designated "MME eligible." Accommodations eligible for MME scores are also indicated by a standard accommodation in the IEP/504 or ELL column, and the accommodation is supported by the student's IEP, 504 Plan, or ELL instruction. Accommodations identified as "MME ineligible" will not count as assessed and will not receive MME scores in the affected subjects.

Spring 2011 Michigan Merit Examination (MME) Accommodations Summary Table

Accommodation	MME Day 1 (The ACT Plus Writing)			MME Day 2	MME Day 3 Standard/Nonstandard	
	May Request	**ACT Comments**	**College Reportable ACT Scores [3]**	***WorkKeys* Eligible Scores[2]**	**IEP/504**	**ELL**
46. Use of screen reader/text-to-speech on reading assessment	No	Not permitted	No	No This accommodation is not MME eligible TACs are required to enter the appropriate Admin code from the range of 900-954 and must provide a response to Local Item Question #2.	NS	NS
47. Use of screen reader/text-to-speech on ELPA • Listening, Writing and Speaking sections only	NA		NA	NA	NA	NA

1. "Yes" in this column means ACT scores will be college reportable ONLY IF that accommodation: 1) is shown as "local decision" or 2) was approved by ACT for an individual student **with a disability.** If the accommodation requires approval, but was not approved by ACT, the ACT scores achieved using that accommodation will not be college reportable. ACT scores are college reportable only if ALL accommodations that require ACT approval are ACT-Approved for that student.
2. Under the "MME Day 2" column, ACT has indicated whether each accommodation is eligible (Yes) or not eligible (No) for WorkKeys score results. Accommodations designated "No" are also ineligible for the National Career Readiness Certificate (NCRC). HOWEVER, the student will receive valid MME scores if the accommodation is designated "MME eligible." Accommodations eligible for MME scores are also indicated by a standard accommodation in the IEP/504 or ELL column, and the accommodation is supported by the student's IEP, 504 Plan, or ELL instruction. Accommodations identified as "MME ineligible" will not count as assessed and will not receive MME scores in the affected subjects.

Spring 2011 Michigan Merit Examination (MME) Accommodations Summary Table

Accommodation	MME Day 1 (The ACT Plus Writing)			MME Day 2	MME Day 3 Standard/Nonstandard	
	May Request	**ACT Comments**	**College Reportable ACT Scores [3]**	***WorkKeys* Eligible Scores[2]**	**IEP/504**	**ELL**
D. Response						
48. Student responds in his/her native language to the constructed response items on assessments.	No	The only constructed response is the ACT Writing Test, and it must be written in **English**. If student is approved for oral responses, responses must be in **English**. (See #54.)	NA	NA	NA	NA
49. Student responds in sign language for Reading and Writing or Functional Independence Assessing Print and Expressing Ideas (Exact English only)	Yes	Only if tested individually and responses marked on scannable document by testing staff. For college reportable ACT scores, video documentation of test session must be returned to ACT. Sign language response to ACT Writing Test must be Exact English Signing (EES).	Yes – only if recording of test session returned to ACT and Writing Test signed EES	Yes Only if tested individually and responses marked on scannable document by testing staff.	NA	NA
50. Student responds in sign language for the ELPA Listening and Speaking sections (American Sign Language (ASL) or Exact English)	NA		NA	NA	NA	NA
51. Student responds in sign language for the ELPA Reading and Writing sections (Exact English only)	NA		NA	NA	NA	NA

1. "Yes" in this column means ACT scores will be college reportable ONLY IF that accommodation: 1) is shown as "local decision" or 2) was approved by ACT for an individual student **with a disability.** If the accommodation requires approval, but was not approved by ACT, the ACT scores achieved using that accommodation will not be college reportable. ACT scores are college reportable only if ALL accommodations that require ACT approval are ACT-Approved for that student.
2. Under the "MME Day 2" column, ACT has indicated whether each accommodation is eligible (Yes) or not eligible (No) for WorkKeys score results. Accommodations designated "No" are also ineligible for the National Career Readiness Certificate (NCRC). HOWEVER, the student will receive valid MME scores if the accommodation is designated "MME eligible." Accommodations eligible for MME scores are also indicated by a standard accommodation in the IEP/504 or ELL column, and the accommodation is supported by the student's IEP, 504 Plan, or ELL instruction. Accommodations identified as "MME ineligible" will not count as assessed and will not receive MME scores in the affected subjects.

Spring 2011 Michigan Merit Examination (MME) Accommodations Summary Table

Accommodation		MME Day 1 (The ACT Plus Writing)			MME Day 2	MME Day 3 Standard/Nonstandard	
		May Request	**ACT Comments**	**College Reportable ACT Scores [3]**	***WorkKeys* Eligible Scores[2]**	**IEP/504**	**ELL**
52.	Student responds in sign language for Mathematics, Science and Social Studies assessments (American Sign Language (ASL) or Exact English)	Yes	Only if tested individually and responses marked on scannable document by testing staff. For college reportable ACT scores, video documentation of test session must be returned to ACT.	Yes – only if recording of test session returned to ACT	Yes Only if tested individually and responses marked on scannable document by testing staff. For ELL students, this accommodation may not be eligible for MME purposes. A response to Local Item #2 is required	S	NS
53.	Student points to answers or writes directly in assessment booklet (transferred to answer document by teacher)	Yes	If student points to answers, student must test individually. Responses must be transcribed to scannable answer document by testing staff while examinee observes.	Yes	Yes If student points to answers, student must test individually. Responses must be transcribed to scannable answer document by testing staff while examinee observes	S	S

1. "Yes" in this column means ACT scores will be college reportable ONLY IF that accommodation: 1) is shown as "local decision" or 2) was approved by ACT for an individual student **with a disability.** If the accommodation requires approval, but was not approved by ACT, the ACT scores achieved using that accommodation will not be college reportable. ACT scores are college reportable only if ALL accommodations that require ACT approval are ACT-Approved for that student.
2. Under the "MME Day 2" column, ACT has indicated whether each accommodation is eligible (Yes) or not eligible (No) for WorkKeys score results. Accommodations designated "No" are also ineligible for the National Career Readiness Certificate (NCRC). HOWEVER, the student will receive valid MME scores if the accommodation is designated "MME eligible." Accommodations eligible for MME scores are also indicated by a standard accommodation in the IEP/504 or ELL column, and the accommodation is supported by the student's IEP, 504 Plan, or ELL instruction. Accommodations identified as "MME ineligible" will not count as assessed and will not receive MME scores in the affected subjects.

Spring 2011 Michigan Merit Examination (MME) Accommodations Summary Table							
Accommodation		MME Day 1 (The ACT Plus Writing)			MME Day 2	MME Day 3 Standard/Nonstandard	
		May Request	ACT Comments	College Reportable ACT Scores [3]	*WorkKeys* Eligible Scores[2]	IEP/504	ELL
54.	Student responds orally (e.g., student tells assessment administrator which answer choice they are selecting)	Yes	Only if tested individually, responses are in **English**, and responses marked on scannable document by testing staff. For college reportable ACT scores, session must be tape recorded with recording also returned to ACT.	Yes	Yes Only if tested individually, responses are in **English**, and responses marked on scannable document by testing staff. For ELL students, this accommodation is not eligible for MME purposes. A response to Local Item #2 is required	S	NS
55.	Use of a scribe for constructed response items (student must indicate punctuation, format and spell all key words) for Writing or Functional Independence Expressing Ideas	Yes	Applies only to ACT Writing Test. Only if tested individually. For college reportable ACT scores, session must be tape recorded with recording also returned to ACT.	Yes – only if recording of test session returned to ACT	NA	NA	NA

1. "Yes" in this column means ACT scores will be college reportable ONLY IF that accommodation: 1) is shown as "local decision" or 2) was approved by ACT for an individual student **with a disability.** If the accommodation requires approval, but was not approved by ACT, the ACT scores achieved using that accommodation will not be college reportable. ACT scores are college reportable only if ALL accommodations that require ACT approval are ACT-Approved for that student.
2. Under the "MME Day 2" column, ACT has indicated whether each accommodation is eligible (Yes) or not eligible (No) for WorkKeys score results. Accommodations designated "No" are also ineligible for the National Career Readiness Certificate (NCRC). HOWEVER, the student will receive valid MME scores if the accommodation is designated "MME eligible." Accommodations eligible for MME scores are also indicated by a standard accommodation in the IEP/504 or ELL column, and the accommodation is supported by the student's IEP, 504 Plan, or ELL instruction. Accommodations identified as "MME ineligible" will not count as assessed and will not receive MME scores in the affected subjects.

Spring 2011 Michigan Merit Examination (MME) Accommodations Summary Table

Accommodation	MME Day 1 (The ACT Plus Writing)			MME Day 2	MME Day 3 Standard/Nonstandard	
	May Request	**ACT Comments**	**College Reportable ACT Scores [3]**	***WorkKeys* Eligible Scores[2]**	**IEP/504**	**ELL**
56. Use of augmentative /alternative communication devices, e.g., • Picture/symbol communication boards • Speech generating devices	Submit details with request	Requests considered individually based on documentation submitted. Approval and reportable status depend on detailed information about the devices proposed for use.	Depends on details	Yes If approved for Day 1, may also be used for Day 2. For ELL students, this accommodation is not eligible for MME purposes. A response to Local Item #2 is required	S	NS
57. Use of speech to text word processor for responses to Writing, Functional Independence Expressing Ideas, and ELPA	Submit details with request	Applies only to ACT Writing Test. Requests considered individually based on documentation submitted. Approval and reportable status depend on detailed information about the proposed speech to text processor.	Depends on details	NA	NA	NA
58. Use of special adaptive writing tools such as pencil grip or larger pencil	Local decision	Provided by school or student.	Yes	Yes For ELL students, this accommodation is not eligible for MME purposes. A response to Local Item #2 is required	S	NS

1. "Yes" in this column means ACT scores will be college reportable ONLY IF that accommodation: 1) is shown as "local decision" or 2) was approved by ACT for an individual student **with a disability.** If the accommodation requires approval, but was not approved by ACT, the ACT scores achieved using that accommodation will not be college reportable. ACT scores are college reportable only if ALL accommodations that require ACT approval are ACT-Approved for that student.
2. Under the "MME Day 2" column, ACT has indicated whether each accommodation is eligible (Yes) or not eligible (No) for WorkKeys score results. Accommodations designated "No" are also ineligible for the National Career Readiness Certificate (NCRC). HOWEVER, the student will receive valid MME scores if the accommodation is designated "MME eligible." Accommodations eligible for MME scores are also indicated by a standard accommodation in the IEP/504 or ELL column, and the accommodation is supported by the student's IEP, 504 Plan, or ELL instruction. Accommodations identified as "MME ineligible" will not count as assessed and will not receive MME scores in the affected subjects.

Spring 2011 Michigan Merit Examination (MME) Accommodations Summary Table

Accommodation	MME Day 1 (The ACT Plus Writing)			MME Day 2	MME Day 3 Standard/Nonstandard	
	May Request	**ACT Comments**	**College Reportable ACT Scores [3]**	***WorkKeys* Eligible Scores[2]**	**IEP/504**	**ELL**
59. Use of adapted paper, lined or grid paper for recording answers	Yes	Provided by school. Student must test individually and responses transferred to scannable answer document by testing staff while examinee observes.	Yes	Yes For ELL students, this accommodation is not eligible for MME purposes. A response to Local Item #2 is required	S	NS
60. Use of alternative writing position (e.g., desk easel, student standing up)	Local decision	If position will disturb other examinees, must test individually.	Yes	Yes For ELL students, this accommodation is not eligible for MME purposes. A response to Local Item #2 is required	S	NS
61. Use of computer or word processor for Writing or Functional Independence Expressing Ideas with the following features disabled • spell check • thesaurus • grammar check.	Yes	Applies only to ACT Writing Test. ACT instructions for printing and returning essay must be followed precisely.	Yes	NA	NA	NA

1. "Yes" in this column means ACT scores will be college reportable ONLY IF that accommodation: 1) is shown as "local decision" or 2) was approved by ACT for an individual student **with a disability.** If the accommodation requires approval, but was not approved by ACT, the ACT scores achieved using that accommodation will not be college reportable. ACT scores are college reportable only if ALL accommodations that require ACT approval are ACT-Approved for that student.
2. Under the "MME Day 2" column, ACT has indicated whether each accommodation is eligible (Yes) or not eligible (No) for WorkKeys score results. Accommodations designated "No" are also ineligible for the National Career Readiness Certificate (NCRC). HOWEVER, the student will receive valid MME scores if the accommodation is designated "MME eligible." Accommodations eligible for MME scores are also indicated by a standard accommodation in the IEP/504 or ELL column, and the accommodation is supported by the student's IEP, 504 Plan, or ELL instruction. Accommodations identified as "MME ineligible" will not count as assessed and will not receive MME scores in the affected subjects.

Spring 2011 Michigan Merit Examination (MME) Accommodations Summary Table

Accommodation	MME Day 1 (The ACT Plus Writing)			MME Day 2	MME Day 3 Standard/Nonstandard	
	May Request	**ACT Comments**	**College Reportable ACT Scores [3]**	***WorkKeys* Eligible Scores[2]**	**IEP/504**	**ELL**
62. Use of Braillewriter or electronic Braillewriter for Writing or Functional Independence Expressing Ideas with the following features disabled • spell check • thesaurus • grammar check	Yes	Provided by school or student. Applies only to ACT Writing Test. ACT instructions for printing and returning essay must be followed precisely.	Yes	NA	NA	NA
63. Use of computers with alternative access for an alternative response mode e.g., • Switches • Alternative keyboards • Eye-gaze motion sensors • Voice recognition software • Head or mouth pointer • Specialized trackballs or mice	Submit details with request	Requests considered individually based on documentation submitted. Approval and reportable status depend on detailed information about the proposed alternative access. If approved, responses must be transferred to scannable answer document by testing staff. Required documentation of original responses to be determined on case-by-case basis for college reportable ACT scores.	Depends on details	Yes, if approved for Day 1, may also be used for Day 2. For ELL students, this accommodation is not eligible for MME purposes. A response to Local Item #2 is required	S	NS

1. "Yes" in this column means ACT scores will be college reportable ONLY IF that accommodation: 1) is shown as "local decision" or 2) was approved by ACT for an individual student **with a disability.** If the accommodation requires approval, but was not approved by ACT, the ACT scores achieved using that accommodation will not be college reportable. ACT scores are college reportable only if ALL accommodations that require ACT approval are ACT-Approved for that student.
2. Under the "MME Day 2" column, ACT has indicated whether each accommodation is eligible (Yes) or not eligible (No) for WorkKeys score results. Accommodations designated "No" are also ineligible for the National Career Readiness Certificate (NCRC). HOWEVER, the student will receive valid MME scores if the accommodation is designated "MME eligible." Accommodations eligible for MME scores are also indicated by a standard accommodation in the IEP/504 or ELL column, and the accommodation is supported by the student's IEP, 504 Plan, or ELL instruction. Accommodations identified as "MME ineligible" will not count as assessed and will not receive MME scores in the affected subjects.

State Use Questions

If the student used accommodations to take any of the three days of testing (Day 1 – ACT Plus Writing, Day 2 – WorkKeys, and Day 3 – Michigan Component), the Test Accommodations Coordinator had to indicate this on the student's answer document by gridding the answers to questions found in either the Day 1 Answer Folder Supplement, the Day 2 Answer Folder Supplement, or the Day 3 Administration Manual for Students Testing with Accommodations, These answers would be entered in the School Use Only/Local Items Boxes on the answer document for each applicable subject.

Following are copies of the Day 1 Answer Folder Supplement, the Day 2 Answer Folder Supplement, and the applicable text from the Day 3 Administration Manual showing the questions developed by Michigan to capture the information about accommodations used by the student**.**:

Must Read!

MME – Day 1

MICHIGAN DEPARTMENT OF EDUCATION
Michigan Merit Examination (MME) State Testing—March 2011
Day 1 - ACT Plus Writing Answer Folder Supplement

This document contains instructions for completing required information on your *ACT Plus Writing 2011 State Testing Answer Folder* that are **specific to the state of Michigan**. These instructions are **in addition to** the instructions provided in the *Spring 2011 ACT Supervisor's Manual State Testing (*or *State Special Testing)* that apply to all states.

If these directions are not followed, results for your school are at risk of being inaccurately reported on student, school, and accountability reports.

ACT Plus Writing vs ACT (No Writing)
All students participating in the MME must take the **ACT Plus Writing**. The *Spring 2011 ACT Supervisor's Manual State Testing* (or *Supervisor's Manual State Special Testing* for students testing with accommodations) includes directions for administering the ACT Plus Writing as well as administering the ACT without the Writing test (No Writing). **Always use the instructions for administering the ACT Plus Writing, unless different instructions are provided in this Answer Folder Supplement.**

Student who withdraws from your school before March 1, 2011
If a student withdraws from your school **before March 1, 2011**, do *not* submit an answer folder for that student. If such a student completed any pretest session portions of an answer folder, destroy that folder.

Page 1 – Blocks I and J – Grade Level and Year of Graduation
ACT has pre-gridded the **Grade Level** in Box I and the **Year of Graduation** in Box J on page 1 of the ACT Plus Writing Answer Folder. **Please disregard**. ACT will use the grade level from the student barcode label for all students testing in Michigan. (Your eligible grade 12 student will be correctly reported in grade 12 if the grade level on the student barcode label on page 10 of the Day 1 Answer Folder is correct.)

Page 1 – Block K – ACT High School Code (School Use Only)

Disregard the instructions for Block K in the "Completing the Non-Test Portions Before Test Day" section of the *Spring 2011 ACT Supervisor's Manual State Testing* and the *State Special Testing Manual)* – use the following instructions:

1. This section should **only** be completed for **home-schooled students** who participate in the MME at your school. Enter and fill in the ovals **969-999 in Block K**. These student scores will NOT be included in the scores for your school.
2. Leave blank for all other students.

Page 1 – Block M – ACT College Planning/Reporting Scores

Please inform students that their ACT scores from the MME will automatically be provided to the State of Michigan. They do not need to use code 2076 – State of Michigan Scholarship Program as one of their four score reporting choices.

Page 10 - Barcode Labels

Each Day 1 Answer Folder submitted for scoring must have a student barcode label affixed in the designated box in the lower right corner of page 10. Student barcodes are required to match the student answer documents from MME Day 1, Day 2, and Day 3 to derive MME subject scores. Your school will be charged a $50 processing fee by the Michigan Department of Education for **each** completed answer folder returned without a student barcode label.

Barcode labels are provided in your Day 1 Non-secure shipment for students who were pre-identified on the BAA Secure Site by January 11, 2011. Follow the instructions in the *Spring 2011 ACT Supervisor's Manual State Testing* and in the *State Special Testing* manual) to affix a label for each student in the space indicated on Page 10 of the Answer Folder. If you receive labels for students no longer enrolled at your school, destroy those labels.

If you did not receive labels for students now enrolled at your school, you **must** print a label:

1. Log in to the BAA Secure Site at **www.michigan.gov/baa-secure**
2. Follow directions in the Secure Site User Manual to **search** for the student.
   a. If the student is found, claim the student and assign to the Spring 2011 MME test cycle.
   b. If the student is not found, register the new student and assign to the Spring 2011 MME test cycle.
3. Print a barcode label for the student following the directions in the Secure Site User Manual. Student barcode labels printed on the BAA Secure Site will not have an ACT School Code. The ACT school code on the preprinted labels is used only for distribution of the labels from ACT to your school. The ACT school code is not used for scoring the answer document.
4. Once you have printed a label, affix it to the Day 1 Answer Folder on page 10 in the box in the lower right corner designated for the Barcode label.

Page 10 - Block S – Local ID Number (Optional)

This field is not required for MME reporting. Its use is strictly for ACT electronic reporting sent to your school.

If your school or district has assigned each student a local ID number of up to ten digits, and you wish to receive that student number on ACT electronic reporting provided to your school, you may use Block S for this purpose. Write the local ID number in the boxes and fill in the corresponding ovals. If the local ID number has fewer than ten digits, it will be reported as ten digits with leading zeros added.

NOTE: The local ID number must NOT be entered in Block C, which is reserved only for the student's Social Security number.

Block U – State Assigned Student ID Number

This block is not used in Michigan for the MME.

Block V – School Use Only– State Questions

This section must be completed for each student testing with accommodations on Day 1 – ACT Plus Writing.

The Michigan Department of Education is required to collect specific student information to ensure the results of state assessments are accurately reported for each student, school, and district. Block V questions (3-9) reflect this requirement.

Block V notes:

- Only school staff can complete Block V questions. Students are not to complete any part of Block V.
- Questions 1-2 – Do not use – Reserved for ACT use with other states
- Questions 3-9 – Must be completed for each student testing with accommodations (ACT-Approved or State-Allowed) on MME Day 1 - ACT Plus Writing. Mark only one oval for each question.
- Please leave Questions 10-15 blank

The Block V questions and response choices are provided below.

Question 1 - Do not use – Reserved for ACT use with other states.

Question 2 - Do not use – Reserved for ACT use with other states.

Block V - Questions 3–9 - Please respond with information about the accommodations actually used for testing on the ACT Plus Writing. Mark **only one** oval for each question.

Question 3 – Which of the following most accurately describes this student's **reason for using test accommodations**?

A. IEP
B. Section 504 Plan
C. ELL instruction
D. Rapid Medical Onset

Question 4 – Under what **testing time** did this student take the ACT Plus Writing tests?

A. Standard time in single session with standard breaks
B. Standard time with stop-the-clock breaks or testing over multiple days
C. Extended time - 50% extended time in self-paced single session
D. Extended time - 50% extended time over multiple days
E. Extended time - More than 50% extended time
F. Extended time on Writing test only (standard time for multiple-choice)

Question 5 – Which **test booklet format** did this student use for the ACT Plus Writing tests?

A. Regular type (10-point)
B. Large type (18-point)
C. Large type (larger than 18-point)
D. Braille/raised line drawings

Question 6 – Which audio, video, or sign language **presentation of test items**, if any, did this student use for the ACT Plus Writing tests?

A. Audio cassette or DVD in English
B. Reader's script in English
C. Video or audio in English **for ELL**
D. Video or audio in Arabic **for ELL**
E. Video or audio in Spanish **for ELL**
F. Translation of reader's script into student's native language **for ELL**
G. Exact English Signing (EES) of test items
H. Signing of test items in any sign language other than EES
I. Closed circuit TV
J. None of the above

Question 7 – Which **accommodated presentation of test directions**, if any, did this student use for the ACT Plus Writing tests?

A. Student restatement or clarification of test booklet and administration directions
B. Test booklet and administration directions in sign language
C. All directions read in student's native language **for ELL**
D. Read/repeat directions exactly as worded in test booklet with emphasis on key words in directions
E. Auditory amplification devices/sound systems
F. Closed circuit TV
G. None of the above

Question 8 – Which **accommodated response mode**, if any, did this student use for the ACT Plus Writing tests?

A. Oral responses, dictation into tape recorder, or pointing to answers
B. Scribe for ACT Writing Test **only**
C. Respond in sign language
D. Augmentative communication devices
E. Computer for Writing Test with certain functions disabled
F. Computer for Writing Test with functions not disabled
G. Braillewriter
H. Computer with alternative access for alternative response mode
I. Speech to text word processor
J. Mark answers in test booklet
K. None of the above

Question 9 – Which **specialized tools**, if any, did this student use for the ACT Plus Writing tests?

A. Concentration tools or noise buffers
B. Bilingual glossary/dictionary
C. Dictionary, thesaurus, spelling, grammar book
D. Visual, auditory, or physical cues to stay on task
E. Magnification devices
F. None of the above

Must Read!

MME – Day 2

MICHIGAN DEPARTMENT OF EDUCATION
Michigan Merit Examination (MME) State Testing—March 2011
WorkKeys Answer Folder Supplement

In this Document
This document contains instructions you and your students will use to complete Michigan-specific information on the *WorkKeys* State Testing Answer Folder. These instructions are in addition to the instructions provided in the *WorkKeys* supervisor manuals. If you do not follow these instructions, results for your school are at risk of being inaccurately reported on student and school reports.

Students Who Withdraw from Your School before March 2, 2011
If a student withdraws from your school before March 2, 2011, do not submit an answer folder for him or her. If such a student partially completed an answer folder, destroy that folder.

Barcode Labels
Requirement: Each MME Day 2 *WorkKeys* State Testing Answer Folder submitted for scoring must have a student barcode label affixed. Student barcode labels are used to match the student answer documents from MME Day 1, Day 2, and Day 3 to derive each student's MME subject scores. Your school will be charged a $50 processing fee by the Michigan Department of Education for each answer folder returned for processing without a student barcode label.

Receiving and Attaching the Labels: Barcode labels are provided in your Day 2 non-secure shipment for students who were pre-identified on the BAA Secure Site by January 11, 2011. When you receive the labels, affix the correct student barcode label for each student in the designated space in the upper-right corner of page 4 of the *WorkKeys* State Testing Answer Folder. If you receive labels for students no longer enrolled at your school, destroy those labels.

Printing Labels: If you did **not** receive labels for students now enrolled at your school, you **must** print labels for them:

1. Log in to BAA Secure Site at **www.michigan.gov/baa-secure**
2. Follow directions in the Secure Site User Manual to search for the student.
   a. If the student is found, claim the student and assign to the Spring 2011 MME test cycle.
   b. If the student is not found, register the new student and assign to the Spring 2011 MME test cycle.
3. Print a barcode label for the student following the directions in the Secure Site User Manual. (Student barcode labels printed on the BAA Secure Site will not have an ACT School Code. The ACT school code on the preprinted labels was used only for distribution of the labels from ACT to your school. It is not used for scoring the answer document.)
4. Once you have printed a label, affix it to the Day 2 Answer Folder.

Demographic Information to be Completed During the Pre-Test Session
For the MME, students MUST complete demographic information on the answer folder in a pre-test session completed before test day. The following fields must be completed in the pre-test session:

- Block 1—School

- Block 2—Name*
- Block 4—Date of Birth*
- Block 5—Gender
- Block 6—Racial or Ethnic Group

The following blocks must be left blank:
- Block 3—ID Number: This number is populated via the barcode label.
- Blocks 7 – 15

* These fields along with the Student-UIC number provided on the student barcode label constitute the matching criteria for National Career Readiness Certificate (NCRC) eligibility. Exact matches are required when creating an online *WorkKeys* account and/or submitting certificate requests. These fields will be displayed on your *WorkKeys Memo to Examinee* reports.

Other Demographic Information

The following blocks are not to be completed until the day of the test. (The verbal instructions in the *WorkKeys* supervisor manuals include instructions to have students grid these blocks on test day before testing begins):
- Block 16—Type of Testing
- Block 25—Student Signature and Date

After Testing: Information Completed by Your School

Overview: The Michigan Department of Education is required to collect specific student information to ensure the results of state assessments are accurately reported for each Michigan student, school, and district. The following blocks are to be completed by the school. Students are not to complete any part of these blocks.

Block 17—Report Codes: You must fill in all applicable ovals for each of your students. (For example, if a student is home-schooled, you would grid the appropriate oval in Block 17 on that student's answer folder.)

Block 26—Local Items (Michigan-Specific Questions): For each student testing with accommodations on any part of MME Day 2—*WorkKeys*, room supervisors must complete items 1 – 17 in Block 26. Use the questions that start on page 3 of this supplement to complete these items. Respond with information about the accommodations **actually used** for testing.
- Questions 1 – 2 gather information about the student's accommodations.
- Questions 3 – 7 gather information about the accommodations provided on the Reading for Information test.
- Questions 8 – 12 gather information about the accommodations provided on the Applied Mathematics test.
- Questions 12 – 17 gather information about the accommodations provided on the Locating Information test.
- Questions 18 – 20 are to be left blank.

Additional Notes for Test Accommodations Coordinators: Enhancements for 2011

Test form numbers: Students testing with accommodations must have test form numbers beginning with "R." You must confirm this before distributing materials. The three test form numbers are listed near the bottom of the test book cover. Do not use standard-time test materials for students testing with accommodations. Standard-time materials have test form numbers beginning with "M."

Administration codes: It is your responsibility to record the proper 3-digit administration code for **every test** you administer with accommodations. Grid all the administration codes on page 3 of the

answer folder. The codes are found on page 5 in the *WorkKeys Supervisor's Manual for State Testing—Special Testing*.
Spanish or Arabic video DVDs (translations): As shown in the supervisor's manual:

- For DVD translations of Applied Mathematics or Locating Information, you must record administration codes beginning with a "7." These tests are eligible for *WorkKeys* scores but are **not** eligible for the NCRC.
- For a DVD translation of Reading for Information, you must record an administration code beginning with a "5" or "6." These tests are eligible for *WorkKeys* scores and the NCRC.

***WorkKeys*-ineligible accommodations:** If you administer a test with these accommodations, you must:

1. Grid an administration code beginning with "9" for that test.
2. Describe the specific accommodation you provided on the "*WorkKeys* Testing Time Verification Form and Administration Report" (page 89 in the *WorkKeys Supervisor's Manual for State Testing—Special Testing*).
3. Complete question #2 in this supplement using Box 26 on the answer folder.

WorkKeys-ineligible accommodations are neither eligible for the NCRC nor eligible for *WorkKeys* scores.

Accommodations Summary Table: Under the "*WorkKeys* Eligible Scores" column, ACT has indicated whether each accommodation is eligible for *WorkKeys* score results. An accommodation designated "No" is ineligible for *WorkKeys* scores and for the NCRC. However, if it is designated "MME eligible," the student will receive MME scores. These accommodations are also noted as standard in the "IEP/504" or "ELL" column, and they are supported by the student's IEP, 504 Plan, or ELL instruction. Accommodations identified as "MME ineligible" will not count as assessed, will not receive MME scores in the affected subjects, and will not be eligible for the Michigan Promise scholarship.

Block 26 Questions

Q1– Which of the following most accurately describes this student's reason for testing with accommodations?

1. IEP
2. Section 504 Plan
3. ELL instruction
4. Rapid Medical Onset

Q2 – If the student tested using nonstandard accommodations on any section of the *WorkKeys* (MME Day 2), mark all that apply (Refer to the MME Accommodations Summary Table to identify nonstandard accommodations—extended time is NOT a nonstandard accommodation).

1. Non-Standard Accommodations for Reading for Information
2. Non-Standard Accommodations for Applied Mathematics
3. Non-Standard Accommodations for Locating Information

Q3 – Which audio presentation of test items, if any, did this student use for the *WorkKeys* Reading for Information test?

1. Reader Script or audio in English
2. Video in English for ELL
3. Video in Arabic for ELL (admin code must begin with a "5" or "6")
4. Video in Spanish for ELL (admin code must begin with a "5" or "6")
5. Translation of reader's script into student's native language for ELL (admin code must begin with "9")

Q4 – Which sign language presentation of test items, if any, did this student use for the *WorkKeys* Reading for Information test?

1. Exact English Signing (EES) of test items
2. Signing of test items in any sign language other than EES (admin code must begin with "9")

Q5 – Which accommodated presentation of test directions, if any, did this student use for the *WorkKeys* Reading for Information test?

1. Student restatement or clarification of test booklet and administration directions
2. Test booklet and administration directions in sign language
3. All directions read in student's native language for ELL
4. Read/repeat directions exactly as worded in test booklet with emphasis on key words in directions
5. Auditory amplification devices/sound systems

Q6 – Which accommodated response mode, if any, did this student use for the *WorkKeys* Reading for Information test?

1. Oral responses, dictation into tape recorder, or pointing to answers
2. Respond in sign language
3. Augmentative communication devices or computer with alternative access for alternative response mode
4. Mark answers in test booklet

Q7 – Which specialized tools, if any, did this student use for the *WorkKeys* Reading for Information test?

1. Concentration tools or noise buffers
2. Bilingual glossary
3. Dictionary, thesaurus, spelling, grammar book (admin code must begin with "9")
4. Visual, auditory, or physical cues to stay on task
5. Magnification devices

Q8 – Which audio presentation of test items, if any, did this student use for the *WorkKeys* Applied Mathematics test?

1. Reader Script or audio in English
2. Video in English for ELL
3. Video in Arabic for ELL (admin code must begin with "7")
4. Video in Spanish for ELL (admin code must begin with "7")
5. Translation of reader's script into student's native language for ELL (admin code must begin with "9")

Q9 – Which sign language presentation of test items, if any, did this student use for the *WorkKeys* Applied Mathematics test?

1. Exact English Signing (EES) of test items
2. Signing of test items in any sign language other than EES (admin code must begin with "9")

Q10 – Which accommodated presentation of test directions, if any, did this student use for the *WorkKeys* Applied Mathematics test?

1. Student restatement or clarification of test booklet and administration directions
2. Test booklet and administration directions in sign language
3. All directions read in student's native language for ELL
4. Read/repeat directions exactly as worded in test booklet with emphasis on key words in directions
5. Auditory amplification devices/sound systems

Q11 – Which accommodated response mode, if any, did this student use for the *WorkKeys* Applied Mathematics test?

1. Oral responses, dictation into tape recorder, or pointing to answers
2. Respond in sign language
3. Augmentative communication devices or computer with alternative access for alternative response mode
4. Mark answers in test booklet

Q12 – Which specialized tools, if any, did this student use for the *WorkKeys* Applied Mathematics test?

1. Concentration tools or noise buffers
2. Bilingual glossary
3. Dictionary, thesaurus, spelling, grammar book (admin code must begin with "9")
4. Visual, auditory, or physical cues to stay on task
5. Magnification devices

Q13 – Which audio presentation of test items, if any, did this student use for the *WorkKeys* Locating Information test?

1. Reader Script or audio in English
2. Video in English for ELL
3. Video in Arabic for ELL (admin code must begin with "7")
4. Video in Spanish for ELL (admin code must begin with "7")
5. Translation of reader's script into student's native language for ELL (admin code must begin with "9")

Q14 – Which sign language presentation of test items, if any, did this student use for the *WorkKeys* Locating Information test?

1. Exact English Signing (EES) of test items
2. Signing of test items in any sign language other than EES (admin code must begin with "9")

Q15 – Which accommodated presentation of test directions, if any, did this student use for the *WorkKeys* Locating Information test?

1. Student restatement or clarification of test booklet and administration directions
2. Test booklet and administration directions in sign language
3. All directions read in student's native language for ELL
4. Read/repeat directions exactly as worded in test booklet with emphasis on key words in directions
5. Auditory amplification devices/sound systems

Q16 – Which accommodated response mode, if any, did this student use for the *WorkKeys* Locating Information test?

1. Oral responses, dictation into tape recorder, or pointing to answers
2. Respond in sign language
3. Augmentative communication devices or computer with alternative access for alternative response mode
4. Mark answers in test booklet

Q17 – Which specialized tools, if any, did this student use for the *WorkKeys* Locating Information test?

1. Concentration tools or noise buffers
2. Bilingual glossary
3. Dictionary, thesaurus, spelling, grammar book (admin code must begin with "9")
4. Visual, auditory, or physical cues to stay on task
5. Magnification devices

The following text is used to capture information regarding accommodations testing for **Day 3** (excerpted from the *Spring 2011 MME Day 3 Administration Manual for Students Testing with Accommodations*, p 79).

Accommodations Documentation

If the student took any part of the MME with an accommodation, Room Supervisors must grid the answers to questions 1-5 below in Box 16 for Mathematics, Box 18 for Science, and Box 20 for Social Studies. The questions are the same for each subject, but it is necessary to grid the answers separately for each subject, based on the actual accommodations **used** for each subject. For example, if the student used a Reader for Social Studies but not for Mathematics or Science, do not fill in any circle under number 2 in Box 16 (Mathematics), do not fill in any circle under number 2 in Box 18 (Science), but fill in circle B under number 2 in Box 20 (Social Studies).

Questions to Answer to Complete Boxes 16, 18, and 20

Question 1 – Which of the following most accurately describes this student's reason for using test accommodations?

A. IEP
B. Section 504 Plan
C. ELL
D. Rapid Medical Onset

Question 2 – Which audio or sign language presentation of test items, if any, did this student use for the test?

A. Audio cassette or DVD in English
B. Reader Script in English
C. Video with English audio for ELL
D. Video with Arabic audio for ELL
E. Video with Spanish audio for ELL
F. Translation of Reader Script into student's native language for ELL
G. Exact English Signing (EES) of test items
H. Signing of test items in any sign language other than EES

Question 3 – Which accommodated presentation of test directions, if any, did this student use for the test?

A. Student restatement or clarification of Test Booklet and administration directions
B. Test Booklet and administration directions in sign language
C. All directions read in student's native language for ELL
D. _Read/repeat directions exactly as worded in Test Booklet with emphasis on key words in directions
E. Auditory amplification devices/sound systems

Question 4 – Which accommodated response mode, if any, did this student use for the test?

A. Oral responses, dictation into tape recorder, or pointing to answers
B. Respond in sign language
C. Augmentative communication devices
D. Computer with alternative access for alternative response mode
E. Mark answers in Test Booklet

Question 5 – Which specialized tools, if any, did this student use for the test?

A. Concentration tools or noise buffers
B. Bilingual glossary/dictionary
C. Dictionary, thesaurus, spelling, or grammar book
D. Visual, auditory, or physical cues to stay on task
E. Magnification devices

Chapter 5: Test Development Analyses

MME Components

The MME is composed of the following components for each subject. This structure is based on the results of alignment analyses between the ACT Plus Writing, *WorkKeys* assessments, and Michigan High School Content Standards, as well as mandates from Michigan legislation. See Table 5.1.

Table 5.1. Components of MME Test Scores for Spring 2011 Administration

			Components Contributing to MME Scores*				
Day	**Test**	**Subject Session**	**Reading**	**Writing**	**Mathematics**	**Science**	**Social Studies**
Day 1	**ACT Plus Writing**	English		X			
		Mathematics			X		
		Reading	X				
		Science				X	
		Writing		X			
Day 2	***WorkKeys***	*Reading for Information*	X				
		Applied Mathematics			X		
		Locating Information			3 items		6 items
Day 3	**Michigan Mathematics, Science and Social Studies**	Mathematics			X		
		Science				X	
		Social Studies					X

Note that the ACT Plus Writing was given on Day 1 of the assessment, the *WorkKeys* tests were given on Day 2, and the Michigan components were given on Day 3. For each subject (column), students needed to complete a certain portion to meet the attemptedness criteria for each section shown with an "X" to obtain a valid score on an MME subject.

Three of the *WorkKeys Locating Information* items count towards MME Mathematics and six of the *WorkKeys Locating Information* items count towards MME Social Studies. This occurs because these items align well with Michigan's high school mathematics or social studies content expectations.

Test Specifications and Alignment Between Contributing Components

Because intact ACT Plus Writing and *WorkKeys* (*Reading for Information*, *Locating Information* and *Applied Mathematics*) assessments must be included in the Michigan Merit Examination (MME), the MME test specifications must start with 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, a yearly alignment process is undertaken. This process is described in detail below. In addition, several in-depth alignment analyses were conducted during the development of the Michigan Merit Examination. These are detailed below in the "Historical Alignment Analyses" section, and is adapted from the materials submitted to the United States Department of Education for peer review of the MME prior to the first implementation in 2007. The evidence referenced in this section is provided as addenda to this technical report.

Alignment of the 2011 MME with HSCEs: Item Selection for Day 1 and Day 2 Scoring

In May of 2010, specialists in English language arts, mathematics, science, and social studies at the BAA conducted an alignment study of the ACT and *WorkKeys* test forms to be used on the Spring 2011 MME. The BAA staff reviewed secure copies of the test booklets and coded each test item to the High School Content Expectation (HSCE) the item most clearly measured. (Specialists in both mathematics and social studies reviewed the *WorkKeys Locating Information* forms.) If an item did not appear to measure an HSCE, the item was left uncoded. A tally of the assigned codes was made for each test form, by standard, to determine the breadth of standards coverage for each form. For both the ACT and *WorkKeys* tests, these tallies differed across the three administration types (initial, makeup, and accommodated). Upon reviewing the tallies the BAA determined how many ACT and *WorkKeys* items would be selected from each form and in each subject area to count toward students' MME scores. These numbers of items were chosen so that the three ACT forms (i.e., initial, makeup and accommodated) would have a common distribution of MME-scored items by standard, and likewise for the three forms of each *WorkKeys* test. This ensured that the MME–scored items taken by each student covered the standards identically, regardless of the combination of Day 1 and Day 2 forms the student took.

Once the number of items for each standard was decided, the BAA content specialists conducted a second review of the Day 1 and Day 2 forms to select these items. Table 5.2 shows the number of items on each Day 1 and Day 2, and how many were selected for MME scoring. Note that at least half of the items on each ACT English, Mathematics, and Science Test, and each *WorkKeys Reading for Information* Test, were selected. Nearly all (32 out of 40) ACT Reading Test items were selected. For each *WorkKeys Locating Information* form, three items were selected to count toward MME Mathematics scores, and six were selected to count toward MME Social Studies scores.

As mentioned above, the purpose of selecting items for MME scoring was to ensure that the MME-scored items taken by every student would cover the standards identically. Table 5.3 shows how the selected items from each Day 1 or Day 2 test were distributed across the standards for that content area.

Table 5.2. Number of items and number of selected items on each MME Day 1 or Day 2 test

Test	Number of scored items	Number of items scored for MME
ACT English	75	45
ACT Writing	one prompt	one
ACT Reading	40	32
ACT Mathematics	60	36
ACT Science	40	20
WorkKeys Reading for Information	30	19
WorkKeys Applied Mathematics	30	7
WorkKeys Locating Information (scored as Mathematics)	32	3
WorkKeys Locating Information (scored as Social Studies)	32	6

Table 5.3. Number of selected items by Day 1 and Day 2 test and by standard

Test	Standard	Number of selected items
ACT English	W1.1	15
	W1.3	15
	W1.4	2
	LAN4.1	13
ACT Writing	W1.3	one prompt (12 points)
ACT Reading	R2.1	13
	R2.2	13
	L3.1	6
WorkKeys Reading for Information	R2.1	5
	R2.2	2
	R2.3	12
ACT Mathematics	L1	3
	L2	1
	A1	11
	A2	2
	A3	1
	G1	15
	G2	1
	S1	1
	S4	1
WorkKeys Applied Mathematics	L2	4
	G1	2
	S1	1
WorkKeys Locating Information (scored as Mathematics)	L1	2
	S1	1
ACT Science	R1	20
WorkKeys Locating Information (scored as Social Studies)	P2	6

Test Development for Michigan Components

In developing the augmentation of ACT Plus Writing and *WorkKeys* to produce the overall MME, it was not feasible to apply 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 IRT-based 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). 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 third 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 that 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.

Historical Alignment Analyses Prior to 2011 Administration

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 an 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.* (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.* (Webb, 2005). The weaknesses in science are summarized on pages 15-16 of *Alignment Analysis of Science Standards and Assessments: Michigan Grades 9–12.* (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 High School Content Standards. 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*. (Dossey, 2005). Although science was well covered, identified weaknesses in life, physical, and earth science are summarized on page 20 of the same document (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, Shanahan clearly states on page 7 of *Review of ACT Coverage of Michigan Language Arts Standards* (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, the BAA chose to augment the Writing portion. Specifically, in order to resolve an issue with balance of representation, a score for Social Studies Decision Making (constructed response item) was added 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 Examination

Norman L. 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) High School Content Expectations.

Below, findings are presented only from the alignment with the 2004 Michigan Curriculum Framework 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 (2004) standards were 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* (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 were six (2004) standards, all of which were addressed in an on-demand assessment. As described in *Alignment Analysis of Mathematics Standards and Michigan Merit Exam: Michigan High School* (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* (Webb, 2006). Of the five 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 the ACT Plus Writing and three *WorkKeys* assessments (*Reading for Information, Applied Mathematics* and *Locating Information*), 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.

Post-Hoc Alignment Studies of the 2010 Michigan Merit Examination

Norman L. Webb conducted an alignment study of the 2010 Michigan Merit Examination in May, 2010. Six reviewers in each of three content areas—English Language Arts (ELA), mathematics, and science—reviewed the Michigan High School Content Expectations and standards with two forms (initial and makeup) of the 2010 MME. For either form, only the selected Day 1 and Day 2 items and the operational Day 3 items were reviewed. The reviewers came from within and outside the state, and included content experts, assessment experts, and state curriculum consultants. As in the previous study, 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 following summarizes the findings and conclusions from this alignment study.

English Language Arts

The alignment between Strand 2 (reading, listening, and viewing) of the Michigan English language arts standards and the ACT/WorkKeys reading assessments administered in 2010 with selected items for Michigan needed slight improvement. The most critical alignment issue was with Range-of-Knowledge Consistency. On the average, reviewers found 45 items on the two forms (R1 and R2) of the 51 item reading assessment that mapped to content expectations under Strand 2. The assessment items were weak with regards to the DOK levels compared to the level of complexity expected by the content expectations. Reviewers fount that the two forms of the assessments had items that only matched about one-fourth of the content expectations under Strand 2. This was too low of a percentage to be considered to have an acceptable level for range. Reviewers found some differences between the two forms in how the items were distributed among the six or seven content expectations. The assessment items on Form R1 were judged to primarily target two content expectations and were not considered to be evenly balanced among the content expectations. Reviewers judged that the items on Form R2 were more evenly distributed among four or five content expectations. Overall, to attain full alignment about six items would need to be replaced or added to target additional content expectations.

The alignment between the two forms of the English ACT assessment and Strand 1 (writing, speaking, and visual expressions) for the Michigan English language arts standards for high school needed slight improvement. The ACT assessment with selected items for the Michigan analysis included one writing prompt worth 12 points and 45 other multiple-choice items, each assigned one point. Nearly all of the item/points, 75, mapped to about seven of the content expectations under Strand I. The DOK levels of the items were sufficiently complex in relationship to the complexity required by the corresponding content expectations. The assessment items, however, were judged to only map to seven (23%) of the content expectations underlying Strand 1. This was an unacceptable level for the range criterion. Balance was acceptable for both assessment forms. Overall, about 10 items would need to be added or replaced to attain full alignment by targeting more of the content expectations under Strand 1. Reviewers were limited to only mapping the writing prompt to three content expectations. They indicated that the writing prompt measured more expectations under Content Expectation CE 1.3 than they were allowed to code. Thus, the range between the Michigan standards and the ACT English assessments E1 and E2 probably was better than indicated here, but still would need improvement.

Mathematics

The alignment between the Michigan high school standards and expectations and the two combined assessment forms (M1 and M2) of the ACT, WorkKeys, and Michigan augmented form depended on the level of analysis performed. If the analysis was performed at the content expectation level, as normally done, then the alignment between the two assessment forms and the Michigan standards needed major improvement as indicated in the summary table below. If the analysis was done at the benchmark level—the level below the standard level (e.g. L.2.1 or A.2.1)—then the alignment only needed slight improvement for each form. Assessment Forms M1 and M2 had nearly the same number of items that corresponded to each of Strands 1, 2, and 3, over 20 items for most strands. This was sufficient to have an acceptable level for the Categorical Concurrence criterion (six or more items) for these three strands.

Neither Form M1 nor M2 had more than two items that reviewers mapped to content expectations under Strand 4, an insufficient number of items for Categorical Concurrence for this strand. The Depth-of-Knowledge Consistency criterion was acceptable for two or three strands for each form that had a sufficient number of items to be analyzed. The DOK levels of items from Form M1 targeting Strand 3 and from Form M2 targeting Strand 2 were too low.

Overall, the alignment for mathematics and the full assessment needed improvement—major improvement if the analysis was done at the content expectation level and slight improvement if the analysis was done at the benchmark level. The augmented item set targeted some content expectations that were not addressed by forms M1 and M2 and had a DOK level that was acceptable.

Science

The science reviewers concluded that the alignment between the Michigan science expectations and standards and the 2010 MME science assessments needed slight improvement. Both forms of the assessment demonstrated Categorical Concurrence and Range of Knowledge in each of the five major areas: biology, chemistry, physics, Earth science, and Inquiry, Reflection, and Social Implications (R1). In addition, both forms demonstrated Depth-of-Knowledge Consistency and Balance of Representation in biology, chemistry, physics, and Earth science. For R1, neither form had a sufficient number of items with a complexity level comparable to the expectations in that area to satisfy the Depth-of-Knowledge Consistency criterion. In addition, most of the R1 items on either form targeted the inquiry content statement (R1.1), while only one or two items mapped to the reflection content statement (R1.2), so that Balance of Representation was not met. The analysis concluded that about six to eight items would need to be replaced on either form in order to attain full alignment.

More details and full documentation of the 2010 MME alignment studies can be found from the following BAA documents: (1.) Alignment Analysis of Mathematics Standards and Two Forms of ACT, WorkKeys, and Augmented Assessments (Forms M1 and M2), One Form of Augmented Assessment with Sixteen Common and Unique Items (Form 1011), and Eight Forms of Augmented Assessment with Six Unique Items (Forms 1003-1010) (Webb, 2011); (2.) Alignment Analysis of English Language Arts Standards and Two Forms of ACT/WorkKeys Reading Assessment and Two Forms of ACT English Assessment (Webb, 2011); and (3.) Alignment Analysis of Science Standards and Two Assessment Forms of ACT and Michigan Items (Forms S1.1 and S2); Nine Assessments Each with Sixteen Unique Items (Forms 1002-1010) (Webb, 2011).

Chapter 6: Erasure Analyses

Description and Purpose

Erasure analysis (also known as mark darkness analysis) is the study 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 (Number 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 5th percentile)
3. 3SDhigh (3 standard deviations above the mean)
4. Prcnthigh (The 95th percentile)

The following flags are applied in the summary data files, based on the thresholds above:

- 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 the BAA 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.

Day 1 and 2 Analysis

Overview

The Day 1 and Day 2 systems employed by ACT each generated an item response file for BAA's use in erasure analysis. The files include one record, representing one answer document, for each Select record provided.

Input

Three types of files are input to the process:

Mark Intensity Files

These contain data from the original scanning of the answer documents, including mark intensities for the test items. There is a mark intensity value for each possible test item response (oval), using a 16 intensity level scan. The lowest mark intensity values (0–3) are not used.

Mark Intensity Files do not identify erasures per se; they identify scan intensity values. Scan intensity is based on a number of factors such as area completed and darkness (light reflection). There is no automated way to distinguish a light mark from an erasure or dark erasure from a true mark. The scanner can only measure that one item response has more intensity (or the same intensity) as another. For building of the erasure patterns, the process assumes the item response with the least intensity is an erasure within an item.

Select File(s)

These contain student records, including the item responses (one response for each item) that were used for scoring. For each item, the scored response is the one for which the darkest mark was scanned (unless there is a double grid, defined below).

Scoring Keys

To identify the correct response to each item.

Output

Match to mark intensity files

For each Select record, the matching mark intensity record is found using the batch and PAS or UIN (a unique identifier for the answer document) for Day 1 and the batch and UIN for Day 2.

Create item response record

An item response record, containing response analysis values, is created for each Select record.

Response analysis values and item response file layout are shown in Tables 6.1 and 6.2 below**.**

Table 6.1. Day 1 and Day 2 Response Analysis Values for Erasure Analysis

Categories	Proposed Response Analysis Value	How determined
"No item response or no erasure detected (normal mark)"	0	There is no more than one mark intensity value for the item.
"Incorrect response changed to correct response"	1	The scored response is correct, and there are one or more marks with a lesser intensity value for the item.
"Correct response changed to an incorrect response"	2	The scored response is incorrect, and there is a lesser mark intensity value for the correct response (but not a double grid).
"Incorrect response changed to another incorrect response"	3	The scored response is incorrect (and is not a double grid), and there are one or more marks with a lesser intensity value for the item that are also incorrect and the correct response has not been marked (i.e., erased).
"Double grid"	4	The two highest mark intensity values for the item are equal intensity or side-by-side on the intensity scale.

Table 6.2. Day 1 and Day 2 Erasure Data Item response file layout

Pipe-Delimited Field #	Fixed Field Start	Fixed Field End	Field Name	ACT (Day 1)		*WorkKeys* (Day 2)	
				Length	Content	Length	Content
1	1	1	Program	1	"1"	1	"2"
2	2	11	Student Barcode Number	10	Select file pos. 412–421	10	Select file pos. 458–467
3	12	19	Student Batch/Process Number	4	Select file pos. 1016–1019	8	Select file pos. 475–482
4	20	25	Student PAS	6	Select file pos. 1029–1034	0	n/a for Day 2 as a separate field (Day 2 PAS is included in field #5 below.)
5	26	45	Student UIN	20	Select file pos. 686–705	6	Select file pos. 469–474
6	46	120	Test 1 Response Analysis	75	Calculated from Select file pos. 436–510	33	Calculated from Select file pos.192–224
7	126	185	Test 2 Response Analysis	60	Calculated from Select file pos. 511–570	33	Calculated from Select file pos. 279–311
8	186	225	Test 3 Response Analysis	40	Calculated from Select file pos. 571–610	38	Calculated from Select file pos. 366–403
9	226	265	Test 4 Response Analysis	40	Calculated from Select file pos. 611–650	0	n/a for Day 2
10	266	271	Test Site ACT Code	6	Select file pos. 204-209	6	Select file pos. 136-141

Day 3 Erasure Analysis

Erasure analysis was performed on all operational multiple-choice responses once all scanning, data correction, and multiple-choice scoring was completed. Data for each student multiple-choice response was programmatically analyzed to determine if the response contained a mark that exceeded the mark threshold and if the lighter marks were potential erasures. Statistics were captured and aggregated on a school and district level to determine whether the school/district data was outside the state norm. Final results were provided to BAA for review and analysis.

Specifically, a program processed a JPEG grayscale image and assigned a Hex value for each multiple-choice bubble. The Hex range was 0 – 15; where Hex 0 was the lightest and represented no shading contained in the bubble, and Hex 15 was the darkest and represented a dark, filled bubble. A student selected response was captured when the Hex value for the bubble was Hex 12 (definite mark threshold) or above. A bubble detected in the range of 9 – 11 was captured as the student response, if no other

bubble for the multiple-choice question was above an 8. A bubble was considered an erasure if the Hex value for the bubble was greater than 5 and less than 12 and not identified as the student response. The following diagram demonstrates the student response and erasure identification process.

	Minimum Hex value	Maximum Hex Value	Examples						
Valid Student Selected Response	9	15		**A**	**B**	**C**	**D**	**Mark**	**Erasure**
			Hex	15	2	3	4	A	No
			Hex	3	11	3	3	B	No
			Hex	9	2	2	4	A	No
			Hex	3	3	1	13	D	No
Multiple Student Selected Response				**A**	**B**	**C**	**D**	**Mark**	**Erasure**
			Hex	2	9	9	3	BC/Multiple	No
			Hex	1	13	2	15	BD/Multiple	No
			Hex	9	1	10	3	AC/Multiple	No
			Hex	2	11	13	3	BC/Multiple	No
Student Selected Response with Erasures				**A**	**B**	**C**	**D**	**Mark**	**Erasure**
			Hex	9	13	3	3	B	A
			Hex	3	6	3	9	D	B
			Hex	7	11	13	3	BC/Multiple	A
			Hex	1	13	7	6	B	CD
			Hex	1	1	14	6	C	D

Figure 6.1. Mark identification examples for Day 3 erasure analysis.

The answer key for each test was used to compare the student selected response, the correct answer, and the erased bubble to determine multiple-choice erasure results. There were three possible results for an erased multiple-choice question: wrong answer to correct answer; correct answer to wrong answer; or wrong answer to wrong answer. A result flag was set for each erased multiple-choice case.

Using the image processed Hex value for each bubble in a multiple-choice question, each Hex value was analyzed to determine if an erasure was present. A flag was set for each bubble that was detected as an

erasure. The iErasureA flag was set if the A bubble was erased; iErasureB was set if the B bubble was erased; and so on.

Tabulated Data Format

All multiple-choice erasure information was tabulated for aggregation into various result sets. The tabulated data was stored in the following format:

Table 6.3. Tabulated Data Format

Data Field	Type	Description
District Number	5 Byte Text	Unique numeric district identifier; (e.g. 73903)
School Number	5 Byte Text	Unique numeric school identifier; (e.g. 08294)
Grade	2 Byte Text	Numeric grade value; padded with 0; (e.g.,11)
Subject	1 Byte Text	A=Mathematics, B=Science, and C=Social Studies
Class Group Number	4 Byte Text	Captured from the answer document
Student Litho	8 Byte Text	Unique student document identifier
Item Position	2 Byte Text	Item position within form (e.g.01, 02)
Erasure A	Bit	Indicates bubble contains an erasure
Erasure B	Bit	Indicates bubble contains an erasure
Erasure C	Bit	Indicates bubble contains an erasure
Erasure D	Bit	Indicates bubble contains an erasure
Wrong to Right	Bit	Indicates response changed from wrong to right
Right to Wrong	Bit	Indicates response changed from right to wrong
Wrong to Wrong	Bit	Indicates response changed from wrong to wrong

Determining the erasure results was a two-step process. The first step was to analyze the tabulated wrong-to-right data and calculate the state average and standard deviation for each subject at each grade level. The second step was to identify student tests containing wrong-to-right erasures that exceeded the state average by more than four standard deviations.

Student Data File

A data file was generated that contained only students exceeding the four standard deviations criterion. The file was formatted in the following layout:

Table 6.4. Day 3 Student Data File Layout for Erasure Analysis

Data Field	Type	Description
District Number	5 Byte Text	Unique numeric district identifier; (e.g. 73903)
School Number	5 Byte Text	Unique numeric school identifier; (e.g. 08294)
Grade	2 Byte Text	NA – grades 11 and 12 reported together. File value = "?"
Subject	1 Byte Text	A=Mathematics, B=Science, and C=Social Studies
Class Group Number	4 Byte Text	Captured from the answer document
Student Barcode	10 Byte Text	Captured from the answer document
Student Lithocode	8 Byte Text	Unique student document identifier
Total Erasures	Integer	Number of erasures on the student test
Total Wrong-to-Right	Integer	Number of responses that had a correct response and one or more erasures
Standard Deviation for Wrong-to-Right answer for specific test	Real	Standard deviation for erasure resulting in the answer changing from Wrong-to-Right answer for the specific test
Average number of erasures for Wrong-to-Right answer for specific test	Real	Average number of erasures in a Wrong-to-Right answer for the specific test

Psychometric Data File

A data file was generated containing a list of **all** students. The file was formatted in the following layout:

Table 6.5. Day 3 Psychometric Data File Layout for Erasure Analysis

Data Field	Type	Description
District Number	5 Byte Text	Unique numeric district identifier; (e.g. 73903)
School Number	5 Byte Text	Unique numeric school identifier; (e.g. 08294)
Grade	2 Byte Text	NA – grades 11 and 12 reported together. File value = "?"
Subject	1 Byte Text	A=Mathematics, B=Science, and C=Social Studies
Class Group Number	4 Byte Text	Captured from the answer document
Student Barcode	10 Byte Text	Captured from the answer document
Student Lithocode	8 Byte Text	Unique student document identifier
Total Erasures	Integer	Number of erasures on the student test
Total Wrong-to-Right	Integer	Number of responses that had a correct response and one or more erasures
Standard Deviation for Wrong-to-Right answer for specific test	Real	Standard deviation for erasure resulting in the answer changing from Wrong-to-Right answer for the specific test
Average number of erasures for Wrong-to-Right answer for specific test	Real	Average number of erasures in a Wrong-to-Right answer for the specific test

Chapter 7: ACT Writing Training and Scoring

Results of Constructed Response Scoring Procedures

The MME assessment includes the written essay component of the ACT Writing Test. The procedure for scoring ACT Writing responses is outlined below. This is the scoring process that Pearson Educational Measurement Performance Scoring Center follows.

Rangefinding

The goal of the rangefinding sessions is to identify a sufficient pool of student responses which illustrate the full range of student performance in response to the prompt, and for which consensus scores can be resolved. This pool of responses includes 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.

All contracted scorers are trained and qualify to score ACT Writing Test responses using the Baseline Prompt training. The Baseline prompt is chosen from a retired operational prompt that performed well in operational scoring. The Baseline prompt and training materials are selected to represent the range and types of responses scorers will see during prompt-specific operational scoring.

Papers are chosen for the Baseline Anchor from operational student responses. The Baseline Anchor Set consists of three papers at each score point, for a total of eighteen papers. Baseline training also consists of four Practice sets, each with ten papers. Contract scorers must then pass two of three ten-paper Qualification sets. Rangefinding sessions for the Baseline Anchor Set are held biennially.

In addition to training and qualifying on the Baseline Anchor set, contracted scorers undergo prompt-specific training on each operational prompt. Prompt-specific training consists of an Anchor Set of nine papers and a Practice Set of four papers.

Rangefinding sessions for prompt-specific training sets are held annually in a separate session from Baseline Training rangefinding.

Prior to all Baseline and Prompt-Specific rangefinding sessions, the Contractor compiles rangefinding papers for prompts into proposed training sets, including writing annotations for all Anchor and Practice papers. ACT staff attends all rangefinding sessions and has final approval of scores assigned to all rangefinding papers.

Rater Training

Thorough training is vital to the consistent application of the scoring rubric and, therefore, accurate scores. The primary goal of training is to convey to the contract 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. Trainers draw scorers' attention to the score point description from the rubric,

as well as the illustrative anchor papers encouraging scorers to immediately connect the language of the rubric with actual student performance. Each anchor paper is also annotated with a scoring explanation that describes why the paper earned the given score. Annotations are meant to further illustrate the connection between the rubric descriptors and the elements present in a given essay.

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.

Scoring

All responses are blind-scored by two scorers using a 6-point holistic scale. If the scores between the two scorers differ by more than 1, the paper is routed for resolution scoring. The resolution scorer will assign a holistic score using a scale of 1-6 inclusive of 0.5, representing adjacent scores. Resolution scoring is non-blind.

Comment Codes

Essay comments, derived from the scoring rubric, are selected by contract scorers to help student writers understand the strengths and weaknesses of their essays. ACT has developed five comment codes per each whole- and half-point score points. During operational scoring, one of the two contract scorers and the resolution scorer, if resolution is required, must assign at least one and not more than four comment codes to each response. Comment code training occurs on the Baseline prompt.

Contractor will identify validity responses for ACT approval.

Rater Monitoring

Pearson Educational Measurement (the Contractor) is responsible for the management and overall monitoring of the operational rangefinding and scoring, but ACT has ongoing access to performance reports.

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. ACT approves all validity responses and has access to ongoing calibration responses and annotations.

Readers score a validity response approximately every 30 responses for ACT 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. The Contractor

provides scorers who perform below the standard validity percentage with constructive feedback, close monitoring, and/or recalibration in the form of calibration papers. Calibration papers are used to correct scorer drift or to illustrate differences between problematic score points for struggling scorers. Appropriate intervention measures are initiated as needed, including the retraining or releasing of scorers who continue to perform below project standards.

Inter-Rater Reliability

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). The Contractor must obtain a cumulative inter-rater reliability ("IRR") level of 0.60 at the conclusion of each scoring window.

In addition, the Contractor must obtain a perfect plus adjacent agreement of 0.95 at the conclusion of each scoring window—that is, 5% or less of resolution scoring.

Contractor staff monitors the accuracy of scoring to maintain the agreed upon inter-rater reliability through back reading, validity, and calibration papers. The validity percentage is 3%.

Chapter 8: Model Fit

This chapter presents the item-model fit and person-model fit analyses for MME.

Item-Model Fit

Item-Model Fit Indexes

The MME Writing, Mathematics, Reading, and Science assessments were scaled and equated using PARSCALE (Muraki & Bock, 1997) and a three parameter logistic IRT/generalized partial credit model for item calibration. (The methods used for estimating examinee scores are discussed later in the chapter on scaling and equating.) The MME Social Studies assessment was scaled with the Rasch model using WINSTEPS.

The MME calibration runs for Writing, Mathematics, Reading and Science were conducted using PARSCALE 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, 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 8.1 to 8.4 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. Based on the S-X^2 statistics, approximately 32%, 25%, 47% and 34% of the scored items for MME Mathematics, Science, Reading and Writing, respectively, were found to be significant. 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 it is more difficult to obtain assessment results that load heavily on the first principal component. Given more complete control over test design and development, it is possible to construct a more unidimensional test that would have better goodness of fit indices for each item.

However, this does not invalidate these measures. 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 10 where all internal consistency and empirical IRT reliabilities are reported.

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. Therefore, we are confident that these are not contributing factors in the fit statistics presented above.

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 of items employed for MME scoring are presented in Figures 8.1 to 8.4. In these plots, there are some item characteristic curves (ICCs) that have flat ICCs.

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 8.5. Based on the MNSQ INFIT and MNSQ OUTFIT statistics, zero percent of items was flagged as moderately or highly misfit.

This chapter examines individual item fit. It also examines person fit.

Table 8.1. Item Fit Statistics – Writing for Spring 2011

ITEM	SX^2	df_SX^2	p_SX^2	χ^2	df	p
AE03	194.14	58	0.00	40.66	10	0.00
AE07	74.72	58	0.07	23.06	10	0.01
AE08	78.37	55	0.02	8.46	9	0.49
AE09	51.11	55	0.62	94.02	9	0.00
AE10	34.10	52	0.97	26.79	8	0.00
AE11	199.51	60	0.00	215.78	10	0.00
AE12	84.92	58	0.01	14.48	10	0.15
AE14	132.00	57	0.00	30.81	9	0.00
AE15	91.64	57	0.00	40.84	9	0.00
AE17	82.73	59	0.02	78.40	10	0.00
AE22	70.79	58	0.12	29.81	10	0.00
AE23	103.44	58	0.00	11.39	10	0.33
AE24	59.93	60	0.48	27.37	10	0.00
AE26	78.99	61	0.06	29.12	10	0.00
AE27	72.29	60	0.13	36.31	10	0.00
AE28	43.04	52	0.81	82.30	8	0.00
AE29	77.67	58	0.04	31.38	9	0.00
AE30	51.46	53	0.53	56.89	8	0.00
AE31	74.06	57	0.06	49.15	9	0.00
AE33	61.68	55	0.25	20.68	8	0.01
AE34	50.06	56	0.70	72.87	9	0.00
AE36	61.46	54	0.23	78.86	8	0.00
AE38	42.84	58	0.93	18.55	10	0.05
AE39	47.37	59	0.86	40.88	10	0.00
AE41	71.28	53	0.05	12.31	8	0.14
AE43	53.71	55	0.52	69.50	9	0.00
AE44	56.65	54	0.38	28.24	8	0.00
AE46	44.08	56	0.88	58.83	9	0.00
AE47	47.89	59	0.85	52.79	10	0.00
AE50	76.05	55	0.03	66.72	9	0.00
AE51	66.70	56	0.16	48.29	9	0.00
AE53	69.66	55	0.09	87.83	9	0.00
AE56	111.44	57	0.00	15.62	10	0.11
AE57	49.99	57	0.73	69.00	9	0.00
AE59	56.89	55	0.40	149.31	9	0.00
AE60	74.01	55	0.04	111.36	9	0.00
AE61	106.15	58	0.00	187.20	10	0.00
AE64	57.49	57	0.46	33.26	10	0.00
AE65	71.89	59	0.12	68.52	10	0.00
AE67	50.58	58	0.74	48.00	9	0.00
AE68	51.21	57	0.69	100.27	9	0.00
AE69	85.62	58	0.01	137.47	10	0.00
AE70	102.92	57	0.00	158.03	9	0.00
AE71	81.74	59	0.03	102.18	10	0.00
AE75	52.16	57	0.66	74.42	9	0.00
AW01	123.34	142	0.87	120.73	38	0.00
AW02	113.96	141	0.95	102.03	38	0.00

Table 8.2. Item Fit Statistics – Reading for Spring 2011

ITEM	SX^2	df_SX^2	p_SX^2	χ^2	df	p
AR01	51.21	38	0.07	55.67	9	0.00
AR02	54.72	38	0.04	49.88	9	0.00
AR03	44.77	37	0.18	22.53	9	0.01
AR04	43.41	42	0.41	17.78	10	0.06
AR06	64.85	38	0.00	38.18	9	0.00
AR07	40.51	37	0.32	31.55	9	0.00
AR08	45.39	39	0.22	14.55	9	0.10
AR10	114.48	35	0.00	25.95	9	0.00
AR11	58.71	35	0.01	38.16	9	0.00
AR12	41.49	31	0.10	80.57	8	0.00
AR13	65.26	32	0.00	44.11	8	0.00
AR15	73.14	33	0.00	100.65	8	0.00
AR16	26.40	35	0.85	34.34	9	0.00
AR17	75.84	28	0.00	97.89	7	0.00
AR19	46.21	37	0.14	14.01	9	0.12
AR20	63.93	36	0.00	31.43	9	0.00
AR21	53.34	40	0.08	56.26	10	0.00
AR22	44.41	41	0.33	9.76	9	0.37
AR23	52.03	40	0.10	10.37	9	0.32
AR25	79.12	34	0.00	159.33	9	0.00
AR26	75.87	39	0.00	46.32	9	0.00
AR27	51.84	35	0.03	133.90	8	0.00
AR28	48.74	34	0.05	109.46	8	0.00
AR30	50.81	32	0.02	156.79	8	0.00
AR31	61.55	40	0.02	49.51	9	0.00
AR32	30.20	43	0.93	28.95	10	0.00
AR33	77.83	38	0.00	116.08	9	0.00
AR34	95.78	40	0.00	124.09	9	0.00
AR35	53.42	42	0.11	25.75	10	0.00
AR38	94.08	39	0.00	136.63	9	0.00
AR39	138.19	36	0.00	217.45	9	0.00
AR40	116.54	37	0.00	187.63	8	0.00
WK01	12.77	16	0.69	28.66	7	0.00
WK03	42.70	32	0.10	22.12	8	0.01
WK05	18.17	13	0.15	13.64	7	0.06
WK06	9.53	5	0.09	9.13	6	0.17
WK09	64.06	32	0.00	2.88	8	0.94
WK10	58.37	37	0.01	11.84	9	0.22
WK12	51.44	37	0.06	37.87	9	0.00
WK13	39.47	41	0.54	25.96	9	0.00
WK14	36.71	38	0.53	49.91	9	0.00
WK15	49.65	34	0.04	17.79	8	0.02
WK18	34.89	38	0.61	28.88	9	0.00
WK19	43.88	40	0.31	29.35	9	0.00
WK20	50.11	39	0.11	44.45	9	0.00
WK23	87.96	36	0.00	8.91	9	0.45
WK28	43.55	41	0.36	29.20	10	0.00
WK29	44.04	38	0.23	31.74	10	0.00
WK31	56.63	43	0.08	20.46	10	0.03

WK32	63.43	43	0.02	59.07	10	0.00
WK33	68.70	39	0.00	22.32	10	0.01

Table 8.3. Item Fit Statistics – Mathematics for Spring 2011

ITEM	SX^2	df_SX^2	p_SX^2	χ^2	df	p
AM01	43.69	45	0.53	32.80	8	0.00
AM03	71.82	44	0.01	60.77	8	0.00
AM05	102.50	42	0.00	121.25	8	0.00
AM07	39.54	40	0.49	32.35	7	0.00
AM08	78.98	52	0.01	85.83	8	0.00
AM12	42.00	45	0.60	143.68	8	0.00
AM13	50.66	42	0.17	83.96	7	0.00
AM14	65.06	52	0.11	57.25	8	0.00
AM17	88.15	51	0.00	129.78	8	0.00
AM20	82.02	64	0.06	54.76	9	0.00
AM21	62.61	52	0.15	157.24	8	0.00
AM22	78.82	66	0.13	131.94	9	0.00
AM23	103.90	50	0.00	215.81	8	0.00
AM27	80.87	54	0.01	202.88	8	0.00
AM28	59.57	56	0.35	49.16	9	0.00
AM30	72.81	64	0.21	108.06	9	0.00
AM31	116.69	58	0.00	179.45	9	0.00
AM32	58.01	59	0.51	200.78	9	0.00
AM33	81.16	64	0.07	133.56	9	0.00
AM34	81.42	64	0.07	169.55	9	0.00
AM35	122.77	62	0.00	111.65	9	0.00
AM39	59.89	59	0.44	232.03	9	0.00
AM40	72.83	64	0.21	263.54	9	0.00
AM41	50.92	58	0.73	149.68	9	0.00
AM42	78.12	68	0.19	285.23	9	0.00
AM43	76.28	65	0.16	182.43	9	0.00
AM44	58.46	67	0.76	184.88	9	0.00
AM46	118.29	63	0.00	45.26	9	0.00
AM47	86.16	68	0.07	143.44	9	0.00
AM48	82.21	61	0.04	120.95	9	0.00
AM49	96.96	64	0.00	61.76	9	0.00
AM52	55.46	57	0.53	199.46	9	0.00
AM56	90.10	68	0.04	37.82	10	0.00
AM57	89.02	67	0.04	240.89	9	0.00
AM58	68.40	70	0.53	89.96	9	0.00
AM59	101.32	71	0.01	70.24	10	0.00
WM07	58.69	24	0.00	19.22	7	0.01
WM17	61.93	59	0.37	58.33	9	0.00
WM20	58.24	48	0.15	94.72	8	0.00
WM23	173.95	54	0.00	468.37	8	0.00
WM25	105.07	59	0.00	252.25	9	0.00
WM28	97.05	61	0.00	148.38	9	0.00
WM29	62.39	70	0.73	62.00	9	0.00
WL17	74.07	67	0.26	23.76	9	0.01
WL24	67.07	65	0.41	71.44	9	0.00
WL25	80.39	74	0.29	23.13	10	0.01
MI01	56.71	48	0.18	85.52	8	0.00
MI02	69.28	58	0.15	154.44	8	0.00
MI03	73.42	58	0.08	105.01	9	0.00

MI04	82.37	66	0.08	88.09	9	0.00
MI05	63.79	60	0.34	21.59	9	0.01
MI06	91.70	55	0.00	43.16	8	0.00
MI07	62.39	66	0.60	190.23	9	0.00
MI08	80.23	67	0.13	64.61	9	0.00
MI09	74.21	68	0.28	22.68	10	0.01
MI10	92.96	65	0.01	115.53	9	0.00
MI11	96.32	57	0.00	101.23	9	0.00
MI12	77.88	64	0.11	233.04	9	0.00
MI13	62.76	60	0.38	109.87	8	0.00
MI14	74.96	57	0.06	132.88	8	0.00
MI15	78.87	60	0.05	132.92	8	0.00
MI16	73.43	54	0.04	18.29	8	0.02
MI17	92.86	51	0.00	61.69	8	0.00
MI18	69.25	60	0.19	342.14	9	0.00
MI19	87.92	68	0.05	17.53	10	0.06
MI20	57.66	61	0.60	92.44	8	0.00
MI21	80.98	63	0.06	146.18	9	0.00
MI22	143.42	66	0.00	11.73	9	0.23
MI23	72.81	67	0.29	114.42	10	0.00
MI24	85.58	61	0.02	156.85	9	0.00
MI25	132.44	62	0.00	348.72	9	0.00
MI26	91.40	67	0.03	158.23	10	0.00
MI27	92.66	58	0.00	179.76	9	0.00
MI28	67.44	63	0.33	31.22	9	0.00
MI29	89.69	68	0.04	265.41	9	0.00
MI30	81.17	58	0.02	115.38	9	0.00
MI31	73.74	51	0.02	23.62	8	0.00
MI32	62.33	64	0.54	82.05	9	0.00
MI33	69.49	59	0.17	68.47	9	0.00
MI34	79.48	60	0.05	172.39	9	0.00
MI35	68.27	58	0.17	41.54	9	0.00
MI36	78.37	68	0.18	121.43	9	0.00
MI37	64.28	56	0.21	10.44	9	0.32
MI38	796.06	66	0.00	404.78	9	0.00
MI39	66.28	70	0.60	68.96	10	0.00
MI40	92.88	66	0.02	240.62	9	0.00
MI41	45.86	61	0.93	234.83	9	0.00
MI42	78.04	68	0.19	73.09	9	0.00
MI43	70.42	54	0.07	162.28	8	0.00
MI44	47.24	50	0.58	100.22	8	0.00
MI45	71.57	59	0.13	210.65	9	0.00
MI46	64.63	56	0.20	282.77	9	0.00
MI47	70.51	65	0.30	117.01	9	0.00
MI48	88.48	68	0.05	69.83	10	0.00
MI49	75.66	65	0.17	133.15	9	0.00
MI50	77.25	64	0.12	71.83	9	0.00
MI51	65.16	63	0.40	116.26	9	0.00
MI52	80.70	66	0.11	198.55	9	0.00
MI53	72.11	64	0.23	171.73	9	0.00
MI54	52.20	57	0.66	70.94	9	0.00

MI55	48.13	54	0.70	36.72	8	0.00
MI56	33.68	31	0.34	7.16	7	0.41
MI57	62.91	63	0.48	206.15	9	0.00
MI58	62.99	65	0.55	92.67	9	0.00
MI59	74.42	70	0.34	71.48	10	0.00
MI60	66.15	66	0.47	134.51	10	0.00
MI61	71.81	52	0.04	65.80	8	0.00
MI62	73.99	58	0.08	367.27	9	0.00
MI63	67.55	65	0.39	219.33	9	0.00
MI64	59.05	55	0.33	174.50	8	0.00
MI65	105.52	69	0.00	34.56	9	0.00

Table 8.4. Item Fit Statistics – Science for Spring 2011

ITEM	SX^2	df_SX^2	p_SX^2	χ^2	df	p
AS01	44.38	31	0.06	171.00	8	0.00
AS03	60.74	35	0.00	152.73	8	0.00
AS06	34.66	41	0.75	39.00	9	0.00
AS07	56.80	37	0.02	116.14	9	0.00
AS09	60.24	40	0.02	81.19	9	0.00
AS10	57.88	43	0.06	52.03	10	0.00
AS11	60.63	34	0.00	95.73	8	0.00
AS15	62.29	39	0.01	51.64	9	0.00
AS17	48.62	29	0.01	161.00	7	0.00
AS22	46.04	41	0.27	28.84	9	0.00
AS23	55.36	38	0.03	169.68	9	0.00
AS25	59.03	42	0.04	92.22	9	0.00
AS29	87.13	44	0.00	93.39	9	0.00
AS31	58.15	41	0.04	154.66	9	0.00
AS32	37.08	44	0.76	29.47	10	0.00
AS33	75.03	46	0.00	59.38	10	0.00
AS35	62.61	46	0.05	66.86	9	0.00
AS36	114.12	45	0.00	140.35	9	0.00
AS38	55.86	47	0.18	42.19	10	0.00
AS39	43.46	47	0.62	45.01	10	0.00
MI01	39.74	38	0.39	38.96	8	0.00
MI02	47.58	39	0.16	118.34	8	0.00
MI03	85.26	46	0.00	15.59	10	0.11
MI04	59.85	40	0.02	27.97	9	0.00
MI05	60.40	37	0.01	72.80	8	0.00
MI06	35.00	44	0.83	24.08	9	0.00
MI07	43.32	45	0.54	48.31	10	0.00
MI08	44.60	39	0.25	52.31	9	0.00
MI09	83.24	47	0.00	29.81	10	0.00
MI10	51.39	45	0.24	32.57	10	0.00
MI11	65.74	48	0.05	79.24	10	0.00
MI12	35.25	39	0.64	55.07	9	0.00
MI13	40.27	42	0.55	53.94	9	0.00
MI14	55.02	39	0.05	69.68	9	0.00
MI15	50.81	47	0.33	19.46	10	0.04
MI16	60.33	36	0.01	115.77	8	0.00
MI17	42.91	39	0.31	56.66	9	0.00
MI18	60.77	41	0.02	16.78	9	0.05
MI19	71.93	41	0.00	20.36	9	0.02
MI20	66.57	30	0.00	36.57	8	0.00
MI21	77.07	42	0.00	13.32	9	0.15
MI22	91.55	44	0.00	17.42	10	0.07
MI23	40.47	42	0.54	68.77	9	0.00
MI24	47.69	39	0.16	75.00	9	0.00
MI25	50.85	48	0.36	29.56	10	0.00
MI26	39.61	46	0.74	22.52	10	0.01
MI27	42.43	45	0.58	22.38	10	0.01
MI28	42.88	42	0.43	95.19	9	0.00

MI29	56.09	41	0.06	67.16	9	0.00
MI30	32.94	46	0.93	21.76	10	0.02
MI31	78.72	43	0.00	59.49	9	0.00
MI32	46.75	37	0.13	117.95	8	0.00
MI33	42.83	46	0.61	32.75	10	0.00
MI34	51.04	46	0.28	31.99	10	0.00
MI35	46.23	46	0.46	25.71	10	0.00
MI36	39.63	44	0.66	37.50	10	0.00
MI37	93.62	42	0.00	16.52	10	0.09
MI38	39.36	42	0.59	68.39	10	0.00
MI39	49.04	46	0.35	34.78	10	0.00
MI40	144.86	44	0.00	21.72	10	0.02
MI41	59.84	47	0.10	41.62	10	0.00
MI42	48.91	39	0.13	52.88	9	0.00
MI43	37.76	47	0.83	24.42	10	0.01
MI44	60.49	47	0.09	26.05	10	0.00
MI45	54.66	40	0.06	58.85	9	0.00
MI46	40.18	40	0.46	90.37	9	0.00
MI47	131.16	35	0.00	27.31	9	0.00
MI48	35.74	39	0.62	103.07	9	0.00
MI49	37.41	38	0.50	86.14	9	0.00
MI50	66.40	46	0.03	27.59	10	0.00
MI51	27.70	33	0.73	51.28	8	0.00
MI52	36.42	44	0.78	73.83	9	0.00
MI53	46.89	37	0.13	79.10	9	0.00
MI54	35.03	32	0.33	78.95	8	0.00
MI55	56.25	45	0.12	108.18	10	0.00
MI56	35.35	42	0.76	102.68	9	0.00
MI57	45.51	33	0.07	138.39	8	0.00
MI58	54.67	47	0.21	23.36	10	0.01
MI59	38.52	42	0.62	110.46	9	0.00
MI60	62.90	42	0.02	60.18	9	0.00
MI61	54.40	46	0.19	64.29	10	0.00
MI62	53.64	41	0.09	103.67	9	0.00
MI63	42.23	40	0.37	120.24	9	0.00
MI64	31.54	39	0.80	139.59	9	0.00
MI65	166.19	43	0.00	26.24	10	0.00
MI66	62.92	45	0.04	78.54	10	0.00
MI67	38.57	45	0.74	48.62	10	0.00
MI68	51.52	40	0.10	92.81	9	0.00
MI69	34.77	38	0.62	94.03	9	0.00
MI70	58.00	49	0.18	13.85	10	0.18
MI71	116.63	44	0.00	44.26	9	0.00
MI72	40.36	47	0.74	49.23	10	0.00
MI73	49.94	46	0.32	36.25	10	0.00
MI74	39.79	33	0.19	84.63	8	0.00
MI75	58.95	45	0.08	71.13	10	0.00
MI76	49.21	45	0.31	68.28	10	0.00
MI77	72.76	46	0.01	47.83	10	0.00
MI78	35.45	38	0.59	157.87	9	0.00
MI79	45.32	46	0.50	43.86	10	0.00

MI80	26.24	39	0.94	96.07	8	0.00
MI81	49.85	47	0.36	24.81	10	0.01
MI82	51.68	45	0.23	44.66	10	0.00
MI83	49.62	46	0.33	43.12	10	0.00
MI84	36.78	43	0.74	82.10	10	0.00
MI85	44.94	44	0.43	70.93	10	0.00
MI86	78.77	41	0.00	40.60	9	0.00
MI87	70.22	40	0.00	25.27	9	0.00
MI88	47.82	42	0.25	98.52	10	0.00
MI89	40.17	39	0.42	123.57	9	0.00
MI90	42.86	49	0.72	12.54	10	0.25
MI91	41.04	42	0.51	78.49	9	0.00
MI92	45.62	37	0.16	144.10	9	0.00
MI93	114.84	44	0.00	276.29	10	0.00
MI94	108.75	42	0.00	21.61	9	0.01
MI95	513.58	31	0.00	133.70	8	0.00
MI96	63.44	42	0.02	80.96	9	0.00
MI97	48.07	45	0.35	44.43	9	0.00
MI98	42.72	46	0.61	29.98	10	0.00
MI99	40.04	43	0.60	41.00	9	0.00
MI100	98.82	41	0.00	7.81	9	0.56
MI101	68.42	42	0.01	7.98	9	0.54
MI102	64.72	36	0.00	27.35	8	0.00
MI103	78.53	45	0.00	116.68	9	0.00
MI104	50.97	40	0.11	74.40	9	0.00
MI105	379.32	44	0.00	163.54	10	0.00
MI106	53.41	48	0.27	25.92	10	0.00
MI107	45.03	45	0.47	44.92	10	0.00
MI108	28.63	43	0.95	56.21	9	0.00
MI109	65.07	44	0.02	30.32	10	0.00
MI110	91.13	43	0.00	12.68	9	0.18
MI111	45.78	37	0.15	61.70	8	0.00
MI112	86.27	34	0.00	32.44	8	0.00
MI113	102.54	42	0.00	10.42	10	0.40
MI114	35.82	43	0.77	59.85	10	0.00
MI115	40.31	40	0.46	83.05	9	0.00
MI116	40.41	46	0.70	44.92	10	0.00
MI117	51.10	44	0.21	72.47	10	0.00
MI118	40.50	45	0.66	34.09	10	0.00
MI119	51.03	44	0.22	38.34	10	0.00
MI120	41.66	44	0.57	63.53	10	0.00
MI121	218.15	41	0.00	12.08	9	0.21
MI122	35.38	44	0.82	56.73	10	0.00
MI123	60.65	45	0.06	64.53	10	0.00
MI124	47.77	42	0.25	43.47	9	0.00
MI125	37.03	38	0.51	113.30	9	0.00
MI126	36.84	41	0.66	75.82	9	0.00
MI127	41.19	40	0.42	80.82	9	0.00
MI128	35.70	43	0.78	48.14	9	0.00
MI129	52.67	41	0.10	80.48	9	0.00
MI130	31.70	38	0.75	62.51	9	0.00

MI131	40.87	43	0.56	68.74	10	0.00
MI132	42.04	45	0.60	23.26	9	0.01
MI133	51.57	38	0.07	107.45	9	0.00
MI134	49.10	41	0.18	62.24	9	0.00
MI135	50.58	35	0.04	62.35	8	0.00
MI136	60.98	44	0.05	91.99	10	0.00
MI137	50.65	47	0.33	23.95	10	0.01
MI138	73.02	45	0.01	66.90	10	0.00
MI139	70.17	43	0.01	38.84	9	0.00
MI140	74.60	32	0.00	24.34	8	0.00
MI141	60.80	46	0.07	25.49	10	0.01
MI142	65.05	46	0.03	30.49	10	0.00
MI143	56.85	43	0.08	56.47	9	0.00
MI144	44.95	29	0.03	27.80	8	0.00
MI145	48.03	47	0.43	10.29	10	0.42
MI146	41.22	40	0.42	53.64	9	0.00
MI147	109.69	43	0.00	18.80	9	0.03
MI148	40.03	34	0.22	114.36	8	0.00
MI149	47.50	45	0.37	45.91	9	0.00
MI150	41.59	47	0.70	51.79	10	0.00
MI151	50.61	43	0.20	57.67	9	0.00
MI152	30.54	41	0.88	117.16	9	0.00
MI153	304.05	44	0.00	139.86	9	0.00
MI154	44.32	42	0.37	50.77	9	0.00
MI155	48.51	41	0.20	71.58	9	0.00
MI156	49.25	44	0.27	27.58	9	0.00
MI157	50.46	41	0.15	57.56	9	0.00
MI158	90.51	43	0.00	11.65	10	0.31
MI159	28.84	44	0.96	60.75	10	0.00
MI160	40.48	43	0.58	53.19	10	0.00
MI161	101.63	35	0.00	13.87	9	0.13
MI162	51.80	46	0.26	51.53	10	0.00
MI163	52.99	39	0.07	93.69	9	0.00
MI164	45.22	43	0.38	55.56	9	0.00
MI165	57.77	47	0.13	18.28	10	0.05
MI166	51.61	44	0.20	52.05	10	0.00
MI167	59.59	44	0.06	29.61	10	0.00
MI168	37.63	45	0.77	43.16	10	0.00
MI169	212.31	45	0.00	42.43	10	0.00
MI170	132.03	40	0.00	10.44	9	0.32
MI171	463.74	26	0.00	80.55	8	0.00
MI172	71.84	42	0.00	20.15	9	0.02
MI173	42.86	43	0.48	50.06	10	0.00
MI174	42.06	39	0.34	72.15	9	0.00

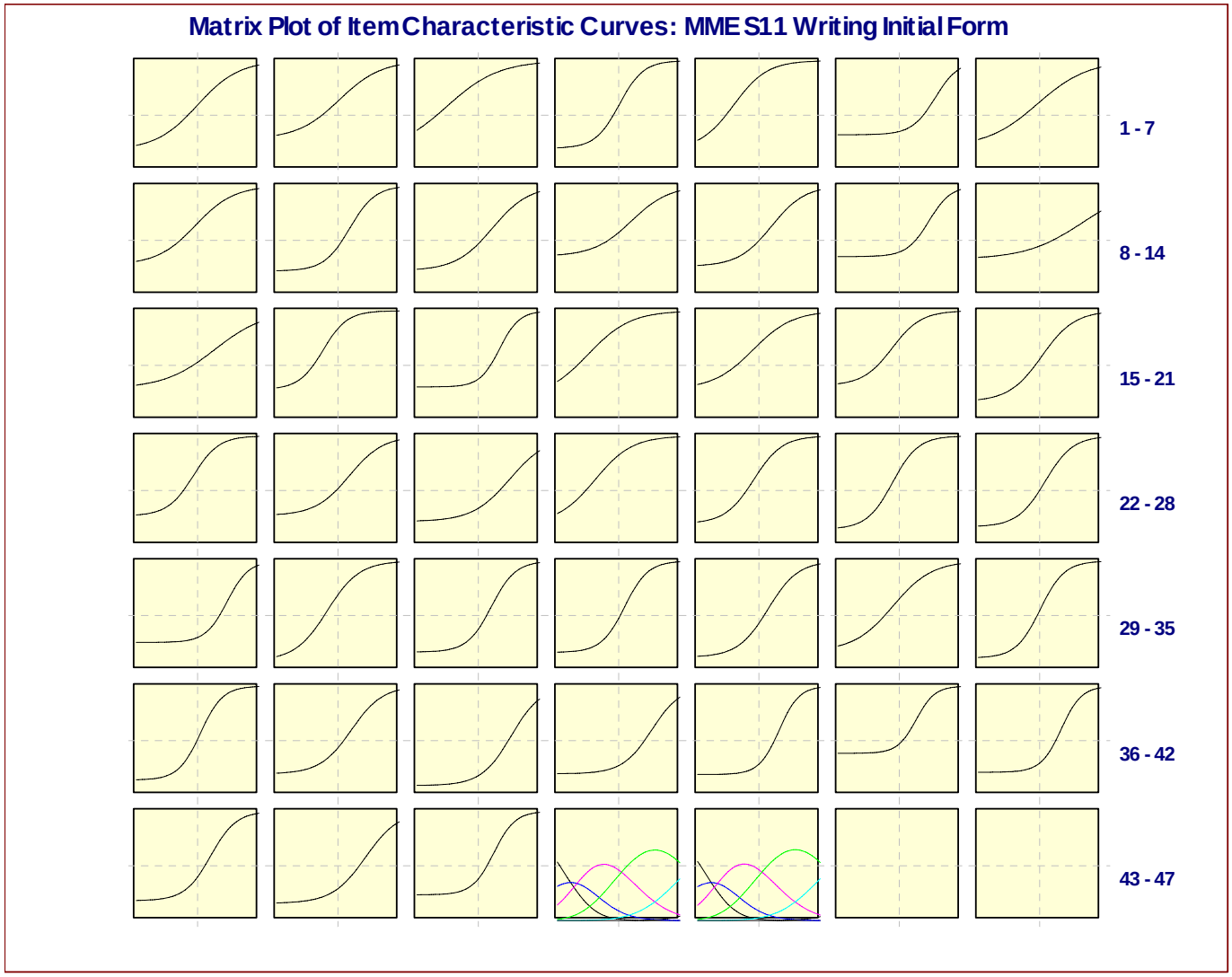


Figure 8.1. Item Characteristic Curves – Writing Spring 2011: 45 selected ACT English items plus one ACT CR item.

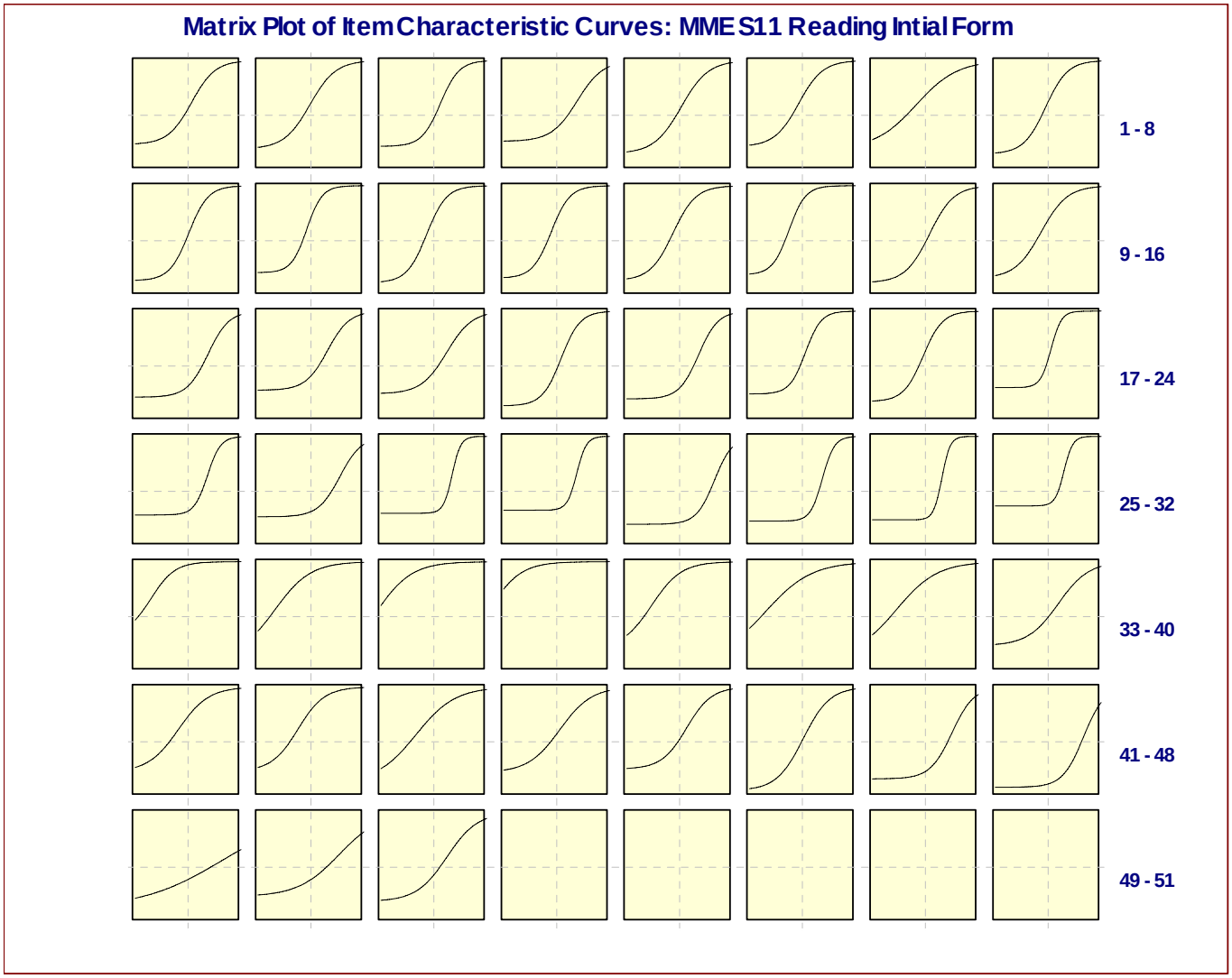
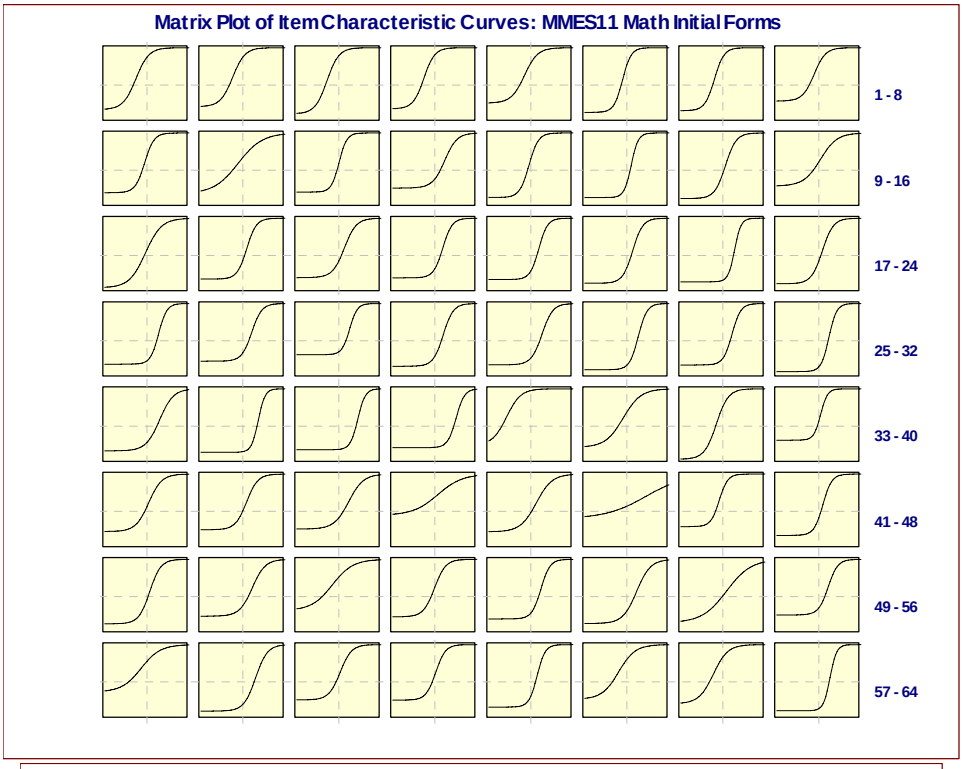


Figure 8.2. Item Characteristic Curves – Reading Spring 2011: 32 selected ACT Reading items plus 19 selected *WorkKeys Reading for Information* items.



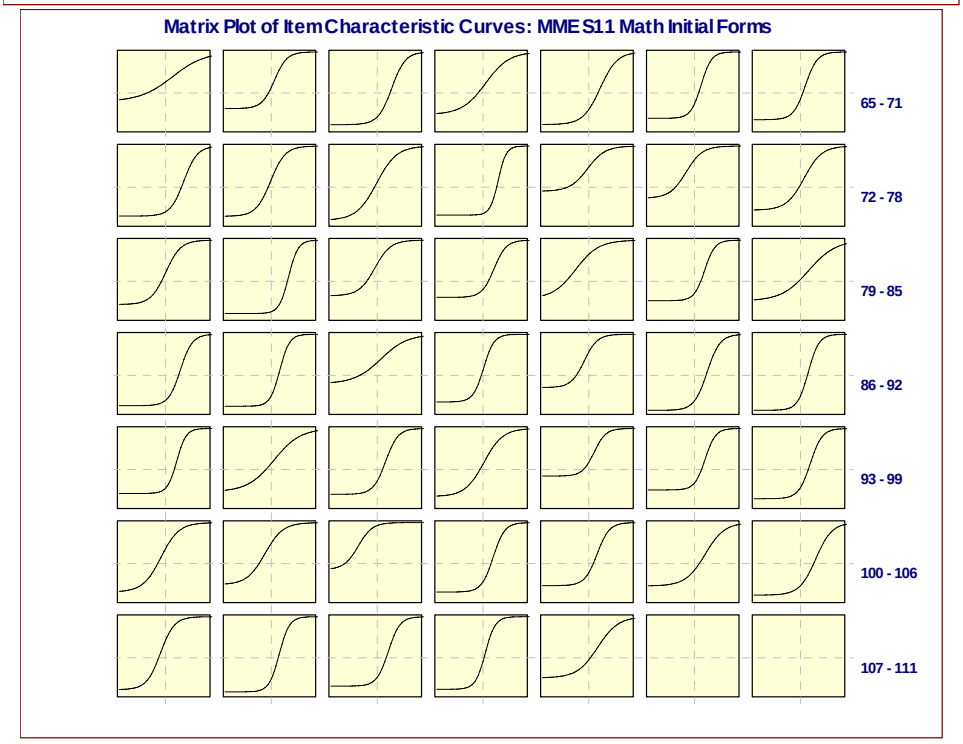


Figure 8.3. Item Characteristic Curves – Mathematics Spring 2011: 36 selected ACT Mathematics items plus 7 selected *WorkKeys Applied Mathematics* items plus 3 selected *WorkKeys Locating Information* items plus 65 unique Michigan-developed Mathematics items.

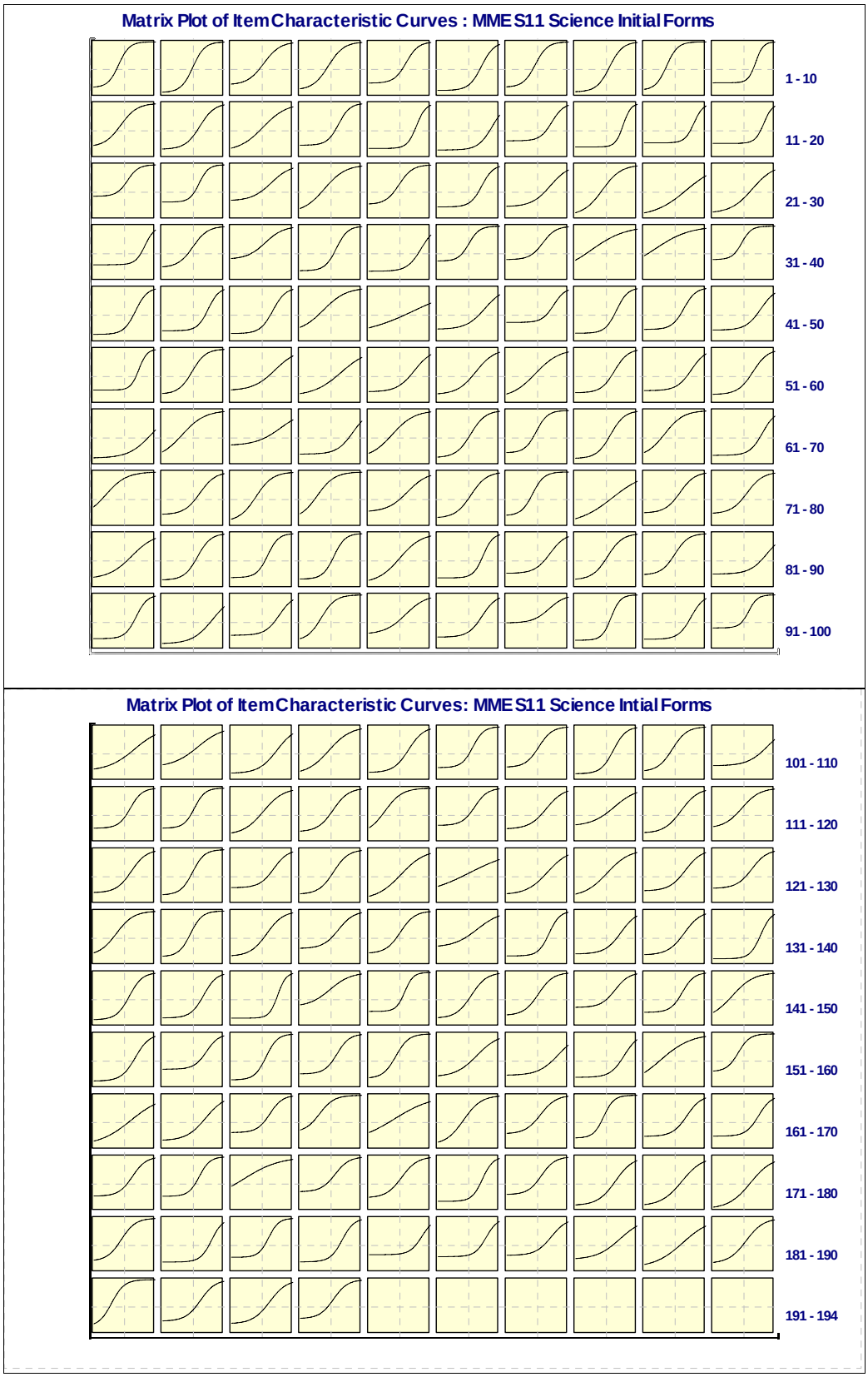


Figure 8.4. Item Characteristic Curves – Science Spring 2011: 20 selected ACT Science items plus 174 unique Michigan-developed Science items.

Table 8.5. Item Fit Statistics – Social Studies for Spring 2011

Item	INFIT MNSQ	OUTFIT MNSQ
WKLI10	1.05	1.34
WKLI12	1.03	1.05
WKLI21	1.11	1.17
WKLI23	1.18	1.23
WKLI35	1.13	1.17
WKLI36	1.18	1.33
SocS01	1.15	1.19
SocS02	1.06	1.08
SocS03	1.07	1.09
SocS04	1.12	1.17
SocS08	0.82	0.75
SocS09	0.81	0.74
SocS10	0.92	0.86
SocS13	1.08	1.15
SocS14	0.92	0.85
SocS15	1.15	1.2
SocS16	0.98	0.91
SocS17	1.05	1.05
SocS18	0.85	0.76
SocS21	0.91	0.85
SocS22	0.9	0.82
SocS23	0.94	0.91
SocS24	0.91	0.85
SocS28	1.37	1.54
SocS29	1.07	1.11
SocS30	1.15	1.19
SocS31	0.91	0.87
SocS32	1.04	1.11
SocS33	0.94	0.9
SocS37	0.91	0.81
SocS38	0.91	0.88
SocS39	0.87	0.79
SocS40	0.94	0.89
SocS41	1	1.01
SocS42	1.03	1.04
SocS46	0.94	0.91
SocS47	0.89	0.86
SocS48	0.99	0.99

Person-Model Fit

The *Lz* Person Fit Index

Person fit measures have been developed to assess the appropriateness of a person's item response pattern consistent with the model employed for characterizing his or her performance. *Lz* is an IRT-based person fit index proposed by Drasgow, Levine & Williams (1985). This index is based on a standardization of the person log likelihood and can be computed as follows:

$$L_z = \frac{\ln L - E(\ln L)}{\sqrt{Var(\ln L)}}$$

where *L*, *E(ln L)* and *Var(ln L)* denote, respectively, the log likelihood, the conditional expected value and conditional variance of the log likelihood for a given theta estimate under the IRT model of interest. For multiple choice and constructed response items, *E(ln L)* and *Var(ln L)* were presented in Drasgow, Levine & Williams (1985). Theoretically, *Lz* follows a standard normal distribution. For individual response patterns, an *Lz* value around 0 indicates good fit and a high negative value reflects a relatively unlikely response pattern than expected based on the model, or misfit (de Ayala, 2009). Previous studies have found that the *Lz* statistics performed as well as, and sometimes better than, other alternative person-fit indices under the 2-parameter logistic and 3-parameter logistic IRT models (Li & Olejnik, 1997, Nering & Meijer, 1998).

Data and computation programs

The MME 2011 match dataset was employed for the *Lz* person-fit analyses. If any component of a test subject is accommodated, the student was included in the accommodated group for the analyses (e.g., a student using standard testing on Day 1 and accommodated testing on Days 2 and 3 is considered "accommodated" for these analyses). Students who did not meet the MME attemptedness criteria for a test subject were excluded from the analyses for that subject. The following table presents the total N-counts for the four test subjects and a breakdown by "accommodated" versus "regular" (i.e., non-accommodated.) It was found that each test subject had approximately 8% of accommodated-test students.

Table 8.6. Number of Examinees by Subjects

test subject	total	regular	accommodated
Math	116,693	107,342	9,351
Science	117,136	107,786	9,350
Reading	117,536	108,092	9,444
Writing	118,267	109,265	9,002
Social Studies	117,248	107,947	9,301

ACT-developed programs for computing the *Lz* person fit index under the 3-parameter logistic model and the mixed-model were applied to the aforementioned datasets. Only the MME-selected items were included in the *Lz* person fit analyses. The following table exhibits numbers of MME-selected items of each test component for an MME form by test subject.

Table 8.7. Number of Selected Items by Subjects and Test Components

test subject	ACT	*WorkKeys* (WK)	Michigan
Math	36	7 WK math and 3 WK locating information	16
Science	20		32
Reading	32	19	
Writing	45 ACT English and 1 ACT writing		
Social Studies		6 WK locating information	32

Results

For each test subject and each group of students (regular and accommodated), ten sub-groups (namely levels) of students were created based on the theta scores. The width of a theta interval for each level is set to be one with the exception that the last interval (level 10) includes theta scores ranging from +3.0 through +6.0. For each level, a common *Lz* critical value of -1.65 was employed for flagging aberrant response patterns. This critical value was chosen because approximately the lowest 5% region is below -1.65 under the standard normal curve. If an observed $Lz \leq -1.65$, the corresponding response pattern is considered to be aberrant/misfit. For each test subject, the following presents histograms of *Lz* for the regular and accommodated groups and a table of breakdowns of the percentages in classification categories by theta levels.

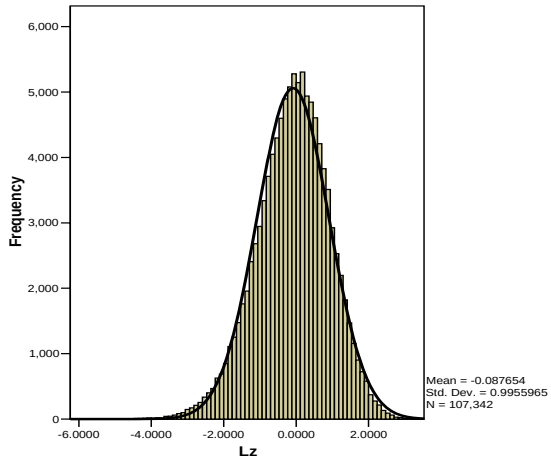


Figure 8.5a. Histogram of *Lz* – Regular Mathematics.

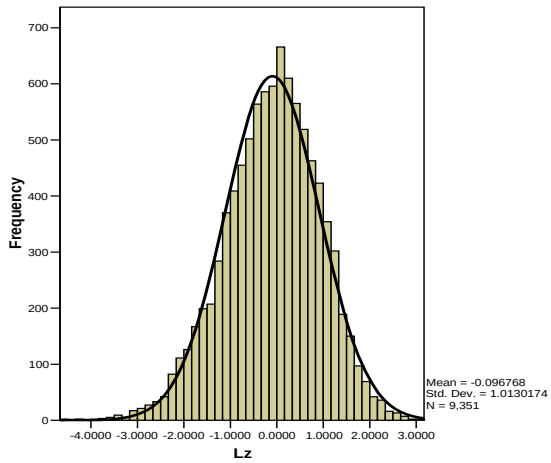


Figure 8.5b. Histogram of *Lz* – Accommodated Mathematics.

Table 8.8. Percentages in Classification Categories: Mathematics

		Regular group			Accommodated group		
Level	theta	N	%misfit	%fit	N	%misfit	%fit
1	[-6.00,-5.00)	2,893	9.19	90.81	995	13.57	86.43
2	[-5.00,-4.00)	549	5.65	94.35			
3	[-4.00,-3.00)	835	7.9	92.1	725	4.0	96.0
4	[-3.00,-2.00)	2,056	5.59	94.41	1,211	3.55	96.45
5	[-2.00,-1.00)	13,074	5.1	94.9	3,051	5.87	94.13
6	[-1.00,0.00)	44,691	6.95	93.05	2,675	8.93	91.07
7	[0.00,1.00)	37,775	6.86	93.14	640	6.25	93.75
8	[1.00,2.00)	5,288	1.46	98.54	53		100
9	[2.00,3.00)	146		100	1		100
10	[3.00,6.00]	35		100			
Total		107,342	6.44	93.56	9,351	7.11	92.89

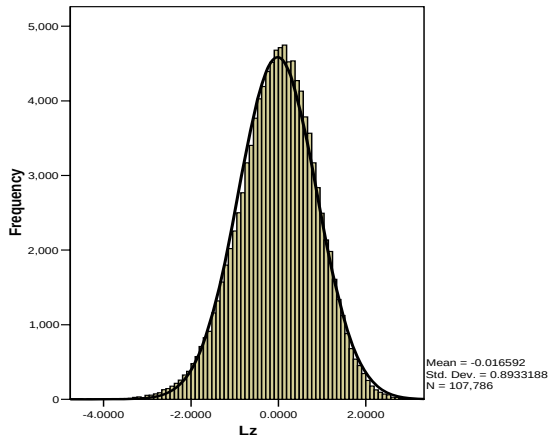


Figure 8.6a. Histogram of *Lz* – Regular Science.

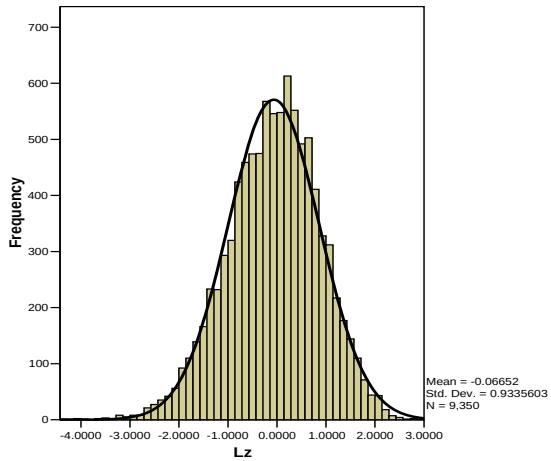


Figure 8.6b. Histogram of *Lz* – Accommodated Science.

Table 8.9. Percentages in Classification Categories: Science

		Regular group			Accommodated group		
Level	theta	N	%misfit	%fit	N	%misfit	%fit
1	[-6.00,-5.00)	2,064	5.81	94.19	853	5.98	94.02
2	[-5.00,-4.00)	210	2.38	97.62	46		100
3	[-4.00,-3.00)	572	2.27	97.73	256	3.13	96.88
4	[-3.00,-2.00)	2,017	2.33	97.67	1,034	3.58	96.42
5	[-2.00,-1.00)	10,233	2.77	97.23	2,711	4.35	95.65
6	[-1.00,0.00)	29,623	5.39	94.61	2,916	5.83	94.17
7	[0.00,1.00)	38,648	4.54	95.46	1,148	7.14	92.86
8	[1.00,2.00)	20,498	2.42	97.58	308	5.19	94.81
9	[2.00,3.00)	3,518	0.63	99.37	69	1.45	98.55
10	[3.00,6.00]	403		100	9		100
Total		107,786	4.02	95.98	9,350	5.17	94.83

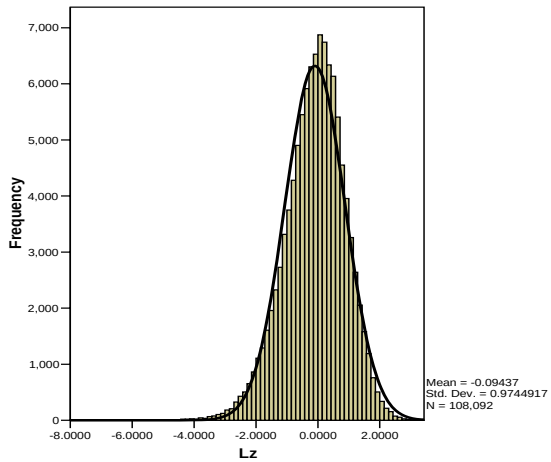


Figure 8.7a. Histogram of *Lz* – Regular Reading.

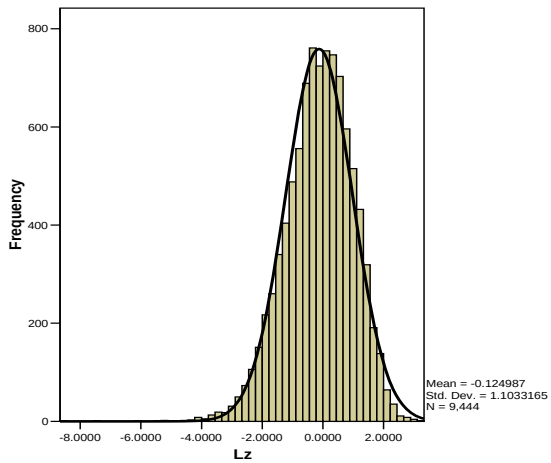


Figure 8.7b. Histogram of *Lz* – Accommodated Reading.

Table 8.10. Percentages in Classification Categories: Reading

		Regular group			Accommodated group		
Level	theta	N	%misfit	%fit	N	%misfit	%fit
1	[-6.00,-5.00)	452	25.66	74.34	215	17.21	82.79
2	[-5.00,-4.00)	365	14.79	85.21	217	14.75	85.25
3	[-4.00,-3.00)	818	15.4	84.6	575	14.78	85.22
4	[-3.00,-2.00)	2,525	7.25	92.75	1,237	8.49	91.51
5	[-2.00,-1.00)	11,336	4.87	95.13	2,548	6.08	93.92
6	[-1.00,0.00)	32,300	7.72	92.28	2,840	7.92	92.08
7	[0.00,1.00)	37,947	6.9	93.1	1,305	11.72	88.28
8	[1.00,2.00)	19,185	3.52	96.48	444	9.01	90.99
9	[2.00,3.00)	2,639	0.27	99.73	60	3.33	96.67
10	[3.00,6.00]	525	0.19	99.81	3		100
Total		108,092	6.32	93.68	9,444	8.83	91.17

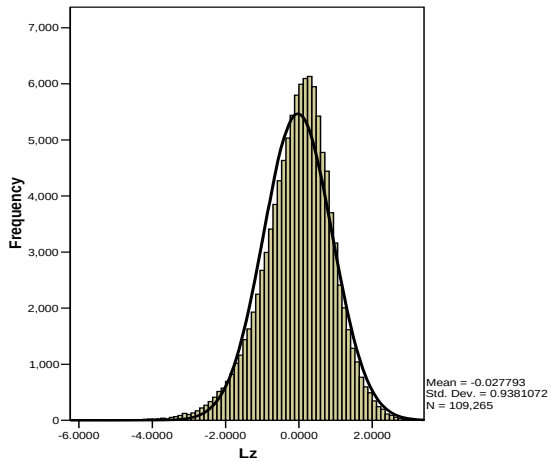


Figure 8.8a. Histogram of *Lz* – Regular Writing.

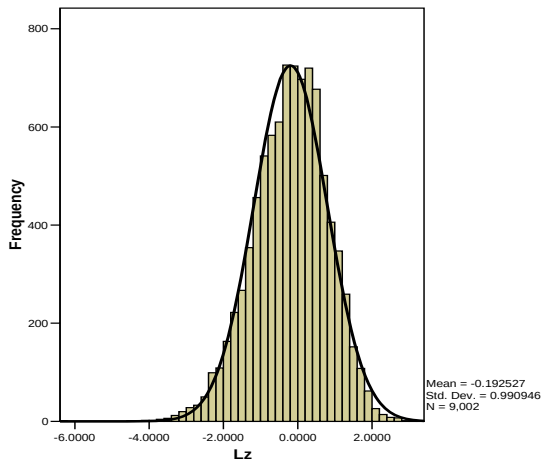


Figure 8.8b. Histogram of *Lz* – Accommodated Writing.

Table 8.11. Percentages in Classification Categories: Writing

		Regular group			Accommodated group		
Level	theta	N	%misfit	%fit	N	%misfit	%fit
1	[-6.00,-5.00)	1,454	21.87	78.13	1195	19.92	80.08
2	[-5.00,-4.00)	214	15.42	84.58	14	28.57	71.43
3	[-4.00,-3.00)	649	12.48	87.52	139	5.76	94.24
4	[-3.00,-2.00)	2,655	7.46	92.54	672	1.64	98.36
5	[-2.00,-1.00)	10,271	6.23	93.77	2,352	4.68	95.32
6	[-1.00,0.00)	29,887	6.89	93.11	3,036	7.94	92.06
7	[0.00,1.00)	38,309	5.05	94.95	1,267	5.37	94.63
8	[1.00,2.00)	19,799	1.73	98.27	269	3.35	96.65
9	[2.00,3.00)	5,025	1.45	98.55	47	2.13	97.87
10	[3.00,6.00]	1,002	4.29	95.71	11		100
Total		109,265	5.24	94.76	9,002	7.66	92.34

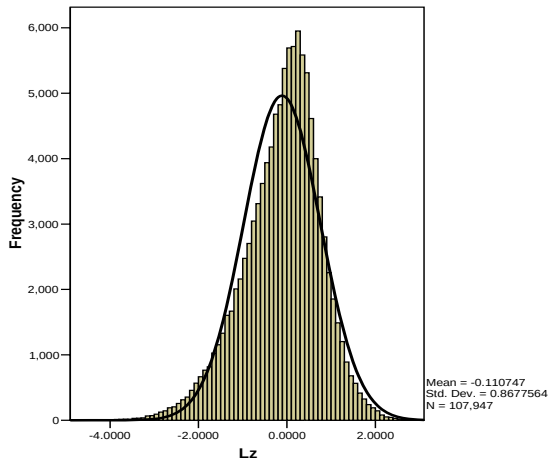


Figure 8.9a. Histogram of *Lz* – Regular Social Studies.

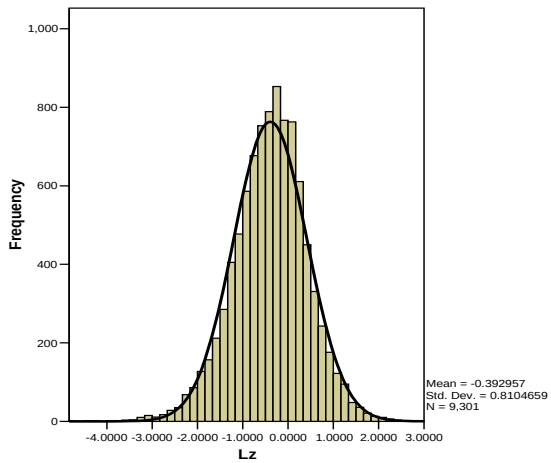


Figure 8.9b. Histogram of *Lz* – Accommodated Social Studies.

Table 8.12. Percentages in Classification Categories: Social Studies

		Regular group			Accommodated group		
Level	theta	N	%misfit	%fit	N	%misfit	%fit
1	[-6.00,-5.00)	2		100			
2	[-5.00,-4.00)						
3	[-4.00,-3.00)	6		100			
4	[-3.00,-2.00)	118	0.85	99.15	60		100
5	[-2.00,-1.00)	6,878	4.62	95.38	2,227	1.12	98.88
6	[-1.00,0.00)	27,462	11.00	89.00	4,524	9.75	90.25
7	[0.00,1.00)	40,490	5.83	94.17	1,753	7.02	92.98
8	[1.00,2.00)	25,871	0.21	99.79	592		100
9	[2.00,3.00)	6,180		100	117		100
10	[3.00,6.00]	940		100	28		100
Total		107,947	5.33	94.67	9,301	6.33	93.67

For Braille-test students

Among the MME state questions given during the MME administration, question #5 of Block V was about the test booklet format students used for the ACT Plus Writing tests. The response to this item was employed to identify Braille-test students. For the MME 2011 administration, four Braille-test students were identified. Table 8.13 presents the *Lz* values for the four Braille-test students by subjects. Based on the item responses of these students who met the MME attemptedness criteria, zero were identified as misfit for mathematics, two for science, zero for reading, two for writing and zero for social studies.

Table 8.13. *Lz* Values for Braille-Test Students

Student	Math	Science	Reading	Writing	Social studies
1	-0.8435	-2.2933*	-1.5508	-0.3636	-1.4561
2	-0.7493	-2.5551*	-1.1535	-1.9869*	0.5589
3			1.1030	0.1601	
4			0.3816	-2.2715*	

* indicates misfit ($Lz \leq -1.65$).

Conclusion

For each test subject, the histograms of *Lz* for the regular and accommodated groups present similar patterns, though the overall accommodated group percentage of misfit is slightly larger than the regular group. As shown in tables of percentages in classification categories, the overall percentages of misfit for each group by test are 6.44 for math regular versus 7.11 for math accommodated, 4.02 for science regular versus 5.17 for science accommodated, 6.32 for reading regular versus 8.83 reading accommodated, 5.24 for writing regular versus 7.66 writing accommodated, and 5.33 for social studies regular versus 6.33 for social studies accommodated. For social studies, the percentages of misfit in theta level 6 ($-1.00 \leq \theta < 0.00$) were found to be relatively high for both the regular and accommodated groups. It is also noted for all test subjects except social studies that in general the percentages of misfit for the lower level groups (e.g., levels 1 – 3) were higher than those for the higher level groups. A plausible explanation for this could be that the test performance (e.g., guessing) of the lower level groups yielded relatively unlikely response patterns on the basis of the model. However, the breakdown of percentages in classification categories exhibit comparable patterns for the regular and accommodated groups of students.

Chapter 9: Scaling and Equating

Quality Control Protocols for MME Calibrations

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 three parameter logistic model (3PLM) for dichotomously scored multiple choice (MC) items and the generalized partial credit model (GPCM) for constructed response (CR) items. For calibrating MME Social Studies, the Rasch model was employed.

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 Measurement Inc. /ACT ftp site was reviewed.

The linking items were also reviewed and verified. Specifically, based on 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.

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.

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.

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.

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 the discrimination parameter estimates. 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.
- 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. Because the raters are randomly assigned, the difficulty, discrimination, and step parameters should be reasonably close across raters.

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 six 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* (AERA et al., 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. For example, a Level Score of 3 or a Scale Score of 73 in *Reading for Information* does not have the same meaning as a Level Score of 3 or a Scale Score of 73 on any other *WorkKeys* test (e.g., *Applied Mathematics*). Two common equating designs are used with the *WorkKeys* tests (Kolen & Brennan, 2004).

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 placed to the score scale using various equating methodologies including linear and equipercentile procedures (e.g., see Kolen & Brennan, 2004). 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 of performances between groups 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 group and form differences.

The various equating methods under the common-item nonequivalent groups design are distinguished in terms of their statistical assumptions (Kolen & Brennan, 2004). 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 is the only MME subject using the Rasch model to derive MME scale scores. The model provides a one-to-one relationship between the derived (i.e., scale) and the raw scores. The item calibration and proficiency estimates are obtained using the Rasch model and procedures implemented in WINSTEPS

version 3.63. The statistical elements of the calibration/scaling process are referred to as Rasch Calibration/Scaling as described in the WINSTEPS manual.

Starting from spring 2009, the MME Social Studies included selected *WorkKeys Locating Information* items. These items were calibrated concurrently with other MME Social Studies items. The item scores for selected *WorkKeys Locating Information* items and MME Social Studies items were summed to obtain a MME Social Studies raw score. The MME Social Studies raw scores were then converted to MME Social Studies scale scores.

Following calibration, operational items are "fixed" when the field test items are calibrated. Each year, new test forms are built based on the test blueprint and available statistical information obtained from previous field testing. New field test items are embedded in test forms for building and replenishing the item pool. These forms are spiraled in the administration. This procedure puts field test item parameters on the scale of the operational items.

Specific Steps for Equating of Social Studies are as follows:

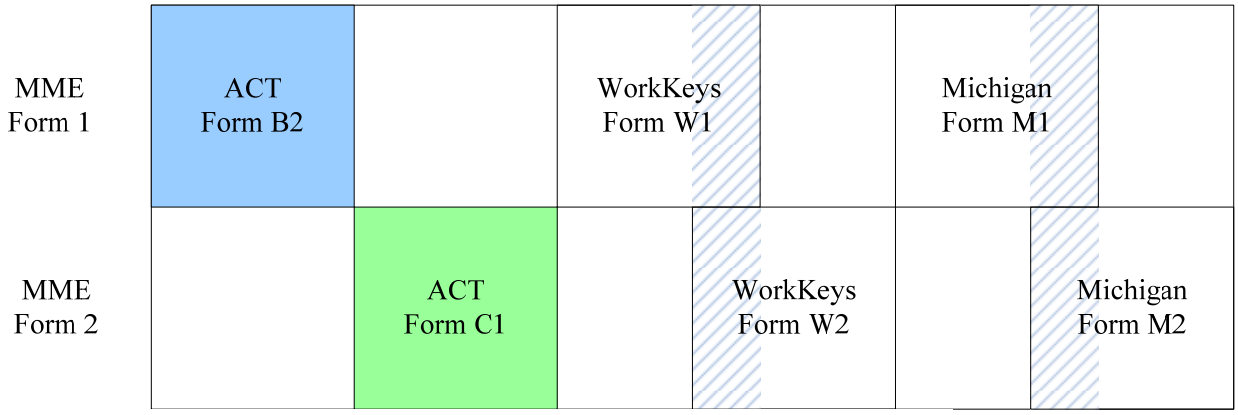
- Review test maps and obtain item parameters from the MME item pool for anchored items.
- Create data sets for item calibration and equating.
- Check the parameter stability of anchored items.
- Run operational item calibration with fixed anchored items using WINSTEPS (version 3.63).
- Review calibration results.
- Create a raw-to-scale score conversion table for scoring.
- Run field test item calibration using WINSTEPS.
- Review field test 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 four components across three days of testing: items from the ACT tests (from Day 1), one or more of the three *WorkKeys* tests (*Reading for Information*, *Applied Mathematics* or *Locating Information*, from Day 2), and Michigan-developed tests (Mathematics, Science, or Social Studies, from Day 3). To develop the MME scale, an MME base form was administered in the spring 2006 Baseline Study. A fixed-parameter calibration approach is employed for equating MME forms, and putting new form scores on the base form scale.

The MME equating plan is exhibited in Figure 9.1. The shaded areas in Figure 9.1 indicate *WorkKeys* and Michigan-developed common items that link between forms (e.g., *WorkKeys* form W1 and W2). The common items have parameter estimates from previous MME administrations. These item parameter estimates are placed on the MME scale and fixed for equating new MME forms. For instance, as illustrated in Figure 9.1, for equating MME form 2, items that are fixed in MME calibration runs include, depending on the MME testing subject, *WorkKeys* common items with item parameters existing from MME form 1, Michigan-developed common items with item parameters existing from MME form 1, and all ACT items on form C1 which have been placed on the MME scale. The equating for MME ACT forms is discussed in the following section.

Figure for MME Linking/Equating Plan



The shaded areas indicate common items between forms.

Figure 9.1. MME linking/equating plan.

The item parameter estimates for all ACT forms administered in MME are separately calibrated under the 3PLM using the ACT national samples discussed previously, and then placed on the MME scale using the Stocking-Lord characteristic curve method (Stocking & Lord, 1983). Figure 9.2 below exhibits the ACT linking studies. Within the same ACT linking study, the randomly equivalent groups design is employed to ensure that form groups are equivalent. For instance, in study 2 as shown in Figure 9.2, form B1 and forms A1 through A4 are administered to randomly equivalent groups. Across ACT equating studies, the Stocking-Lord transformation is employed. For example, study 1 and study 2 are linked through form B1, and forms A1 through A4 can then be placed on the study 1 scale accordingly.

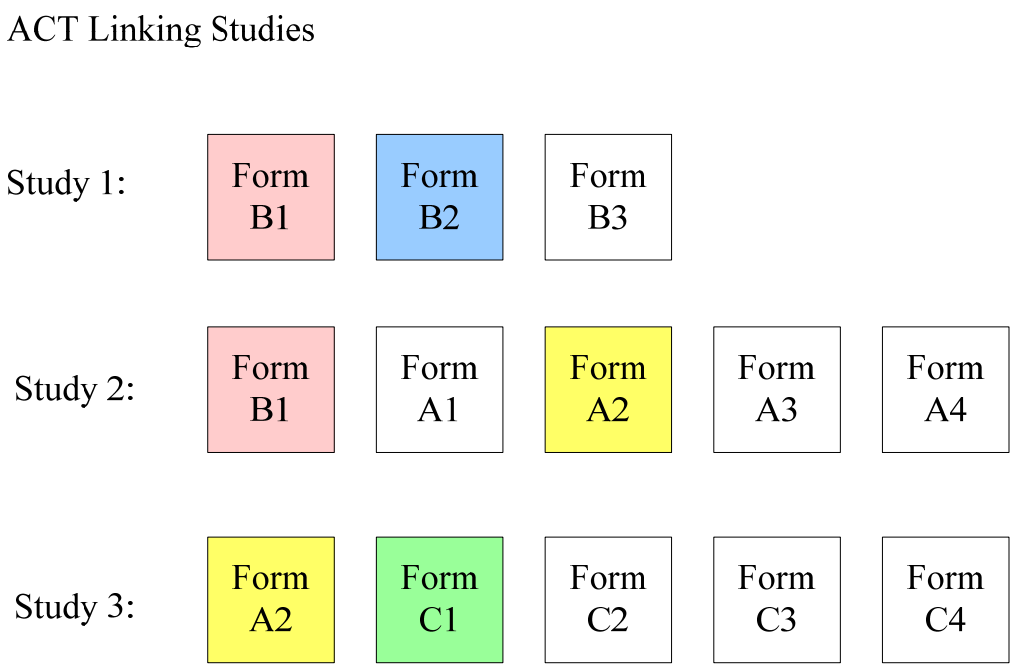


Figure 9.2. ACT linking studies.

In light of the ACT linking studies, any ACT form can be placed on the MME scale. Figure 9.3 depicts the linkage for MME ACT forms. For example, as shown in Figure 9.3, forms A3, B2 and C1 exist in the MME pool. In the ACT linking study comprising forms B1 and B2 that are administered to randomly equivalent groups, form B2 can be equated to the MME pool via the Stocking-Lord procedure. Form B1 can then be equated to the MME pool.

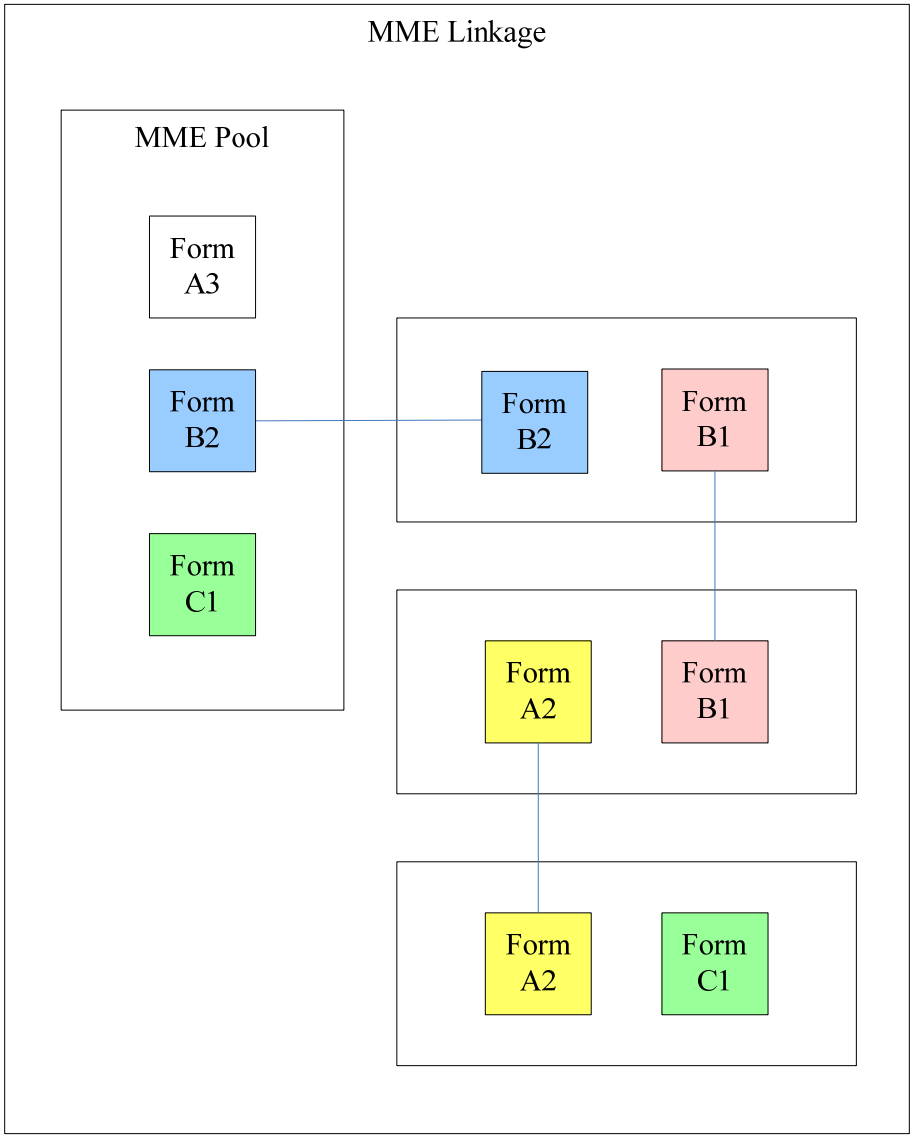


Figure 9.3. Diagram for MME linkage.

To link the *WorkKeys* and Michigan-developed test forms, respectively, to the MME base form, a set of anchor items is employed and calibrated using MME sample data as shown in Figure 9.1.

For MME calibrations, the 0/1 scores of 11th grade students who meet MME attemptedness criteria were used. The MME calibration runs were conducted using PARSCALE version 4.1 (Muraki & Bock, 1997) under the GPCM for CR items and the 3PLM for dichotomous items. These models are given as follows:

Under the GPCM, the probability that an examinee j scores z with $z = 0, 1, \ldots, Z_i$ on item i with $Z_i + 1$ response categories is modeled by

$$P(z \mid \theta_j, \alpha_i, \beta_i, \tau_{ci}) = \frac{exp \sum_{c=0}^{z} \alpha_i \left[\theta_j - (\beta_i - \tau_{ci})\right]}{\sum_{y=0}^{Z_i} exp \sum_{c=0}^{y} \alpha_i \left[\theta_j - (\beta_i - \tau_{ci})\right]},$$

where α_i is the discrimination of item *i*, β_i denotes the difficulty of item *i*, and τ_{ci} represents the location parameter for a category on item *i*. For model identification, it needs to set $\tau_{0i} = 0$, $\sum_{c=1}^{Z_i} \tau_{ci} = 0$ and $exp \sum_{c=0}^{0} \alpha_i \left[\theta_j - (\beta_i - \tau_{ci})\right] = 1$ in the model above (Muraki, 1992, 1996).

Under the 3PLM, the probability that an examinee *j* scores *z* with *z* = 0 or 1 on item *i* is modeled by

$$P(z \mid \theta_j, \alpha_i, \beta_i, c_i) = c_i + (1 - c_i) \frac{1}{1 + \exp[-1.7\alpha_i(\theta_j - \beta_i)]},$$

where α_i, β_i and c_i denote the discrimination, difficulty and pseudo-guessing parameter estimates, respectively.

For the MME administration, a concurrent calibration run for the various components is implemented with fixed item parameter estimates for all ACT items, fixed item parameter estimates for the *WorkKeys* anchor items, and fixed item parameter estimates for the Michigan-developed items; with all other items being placed on the MME scale by the calibration run. Michigan-developed operational items that were administered as field test items in previous MME administrations were recalibrated.

For scrambled versions of the Michigan-developed forms that are used in different testing situations, (i.e., initial, makeup and accommodated), the item parameter estimates for Michigan-developed anchor items are obtained from a master initial calibration run using the data for the initial forms for all of the various MME components. These calibration analyses are 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 property of group invariance, these item parameters were fixed for the calibration runs for other form combinations. Also, for calibrating Michigan-developed field test items, item parameters of ACT items, *WorkKeys* items and Michigan-developed operational items were fixed. Field test items with point biserials less than .10 were excluded from the field test item calibrations as per BAA's direction.

For MME scoring, only the selected ACT items, selected *WorkKeys* items and Michigan-developed items were employed.

Specific steps for equating MME Writing, Mathematics, Reading and Science are as follows:

1. Review test maps.
2. Obtain item parameter estimates from the pool for anchor items.
   - For forms with small N-counts (e.g., Braille or makeup), item parameter estimates obtained from master initial calibration runs are employed if available.
   - For forms that are a scrambled version of the initial form, item parameter estimates of the initial form are used.
3. Create data sets for calibration and equating.
4. Check anchor item parameter stability.
5. Conduct fixed-parameter calibration runs using PARSCALE without field test items.

6. Evaluate calibration results of operational items and pass item parameter estimates for MME scoring. (Only the selected ACT items, selected *WorkKeys* items and Michigan-developed items are employed for MME scoring.)
7. Run PARSCALE to calibrate field test items with item parameter estimates of all operational items fixed. Field test items with point biserials less than .10 are excluded from the field test item calibrations as per BAA's direction.
8. Review calibration results of field test items for future form construction considerations and linking.

Calibration Summary Reports

Calibration summary reports that discussed N-counts, calibration convergence, and ACT's suggestions for MME scoring were presented to the BAA for their review.

IRT Model Fit and Plots

Matrix plots of item characteristic curves resulting from PARSCALE calibration runs were presented to the BAA for their review. The plots of SE/information curves produced by PARSCALE with the MME cut scores imposed for the testing subjects of Writing, Reading, Mathematics and Science, respectively, were also created and presented to the BAA for their review.

For MME Social Studies, the mean square fit (MNSQ) statistics obtained from WINSTEPS are 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.

The MME calibration runs for Writing, Mathematics, Reading and Science are conducted using PARSCALE (Muraki & Bock, 1997) under the GPCM for CR items and the 3PLM for dichotomous items. Two model fit indices are used for the dichotomous and polytomous items. They are the Chi-square (χ^2) statistics provided in PARSCALE phase 2 output, and Orlando and Thissen's (2000) S-X^2 statistics. To compute the Chi-square index, ten ability groups are used. To test the goodness of fit for each item, a significance level (α) of .05 is used. If the observed p-value associated with a fit index for an item is lower than .05, the item is considered to be a "poorly" fitting item. The χ^2 tests of item fit are, however, extremely sensitive to sample size, which is very large for MME. The item fit statistics are reported in Tables 8.1 through 8.5.

Item Analysis

After the MME administration, the Measurement Research Department (MRD) at ACT receives matched data files. MRD computes classical item statistics as specified by the BAA for the Michigan-developed operational and field-test MC items and creates a SAS dataset containing these item statistics that it sends to the BAA for their review.

MRD computes IRT based item statistics as specified by the BAA for the Michigan-developed operational and field-test MC items and adds these statistics to the classical item statistics SAS dataset. MRD also adds the item parameters to this file. It then sends this combined item analysis file to the BAA for their review.

Theta Generation

ACT developed an ACT-written program, SCOREST, to compute MME thetas (θs), and uses independent checks on the thetas. The SCOREST program is written in C++ and developed by the Computer Based Testing Research (CBTR) team in conjunction with IT. For the purpose of independent checks on MME theta scores, MRD staff developed two FORTRAN programs for estimating θs: MULTEST for multiple choice data and MIXEDEST for mixed format data. These programs produce theta estimates and standard errors of (SE) theta.

IRT Models

Two IRT models are employed in the scoring programs: the 3PLM and the GPCM. The 3PLM for dichotomous MC items with $z = 0$ or 1 is given as follows:

$$P(z \mid \theta_j, \alpha_i, \beta_i, c_i) = c_i + (1 - c_i)\frac{1}{1 + exp\left[-D\alpha_i\left(\theta_j - \beta_i\right)\right]},$$

where $D = 1.7$, α_i is the discrimination of item i, β_i denotes the difficulty of item i, and c_i is the pseudo-guessing parameter of item i. Under the GPCM, the probability that an examinee j scores z with $z = 0, 1, \ldots, Z_i$ on item i with $Z_i + 1$ response categories is modeled by the following:

$$P(z \mid \theta_j, \alpha_i, \beta_i, \tau_{ci}) = \frac{exp\sum_{c=0}^{z} D\alpha_i\left[\theta_j - (\beta_i - \tau_{ci})\right]}{\sum_{y=0}^{Z_i} exp\sum_{c=0}^{y} D\alpha_i\left[\theta_j - (\beta_i - \tau_{ci})\right]},$$

where α_i is the discrimination of item i, β_i denotes the difficulty of item i, and τ_{ci} represents the location parameter for a category on item i. For model identification, set $\tau_{0i} = 0$, $\sum_{c=1}^{Z_i} \tau_{ci} = 0$ and $exp\sum_{c=0}^{0} \alpha_i\left[\theta_j - (\beta_i - \tau_{ci})\right] = 1$. For both 3PLM and GPCM, item parameter estimates are computed using PARSCALE.

Algorithms for the Scoring Programs

The MULTEST and MIXEDEST programs use the grid search algorithm to estimate the maximum likelihood estimate (MLE) of θ values. Under the grid search algorithm, the θ space ranging from -6.0 to +6.0 is divided into grids with a width of .001, and magnitudes of the log-likelihood are computed for all grid points under the appropriate IRT model(s). The theta score with the highest log-likelihood value is selected and denoted θ^*. A finer search with a grid width of .0001 is then conducted in the neighborhood of θ^*. The MLE theta score $\hat{\theta}$ is then given by the theta score that yields the highest log-likelihood value in the finer search, and SE for $\hat{\theta}$ is computed accordingly. One advantage of the grid search algorithm is that the non-convergence for cases with irregular log-likelihood curves (e.g., flat, monotonically increasing, monotonically decreasing, or multi-modal) under the Newton-Raphson algorithm is avoided.

The algorithm employed by SCOREST for computing MLE $\hat{\theta}$ is a modified grid search, using the appropriate psychometric model(s) for each item. First, the theta score denoted θ_1 which maximizes log-likelihood over 121 equally spaced thetas between -6 and 6 (spaced by .1) is selected. Then the lower bound

is set to be $\theta_1 - .1$ and the upper bound is set to be $\theta_1 + .1$, and the theta score denoted θ_2 that yields the maximum log-likelihood over 121 theta values spaced by .01 is computed. Similarly, the lower bound is set to be $\theta_2 - .01$ and the upper bound is set to be $\theta_2 + .01$, and the theta score denoted θ_3 that yields the maximum likelihood over 121 theta values spaced by .001 is computed. This procedure is repeated until the spacing between theta values is less than .00001.

After the MLE $\hat{\theta}$ is computed, the SE for $\hat{\theta}$ is computed using the following algorithm. The test information at $\hat{\theta}$ is evaluated by summing item information functions over operational items administered. The calculation of item information depends on whether the item is a multiple choice item (and satisfies a 3PLM) or is a constructed response item (and satisfies a GPCM). The SE for $\hat{\theta}$ is then computed as the square root of the inverse of the test information.

Results of Test Runs

Comparisons between Pearson's ISE, PARSCALE, MULTEST (or MIXEDEST), and SCOREST results were conducted to produce some initial information on how well results from these programs match on writing, reading, mathematics and science thetas for the 2007 spring administration. For this comparison study, ISE, MULTEST, MIXEDEST and SCOREST used the 2007 spring initial form test samples from the final match file, and PARSCALE used the 2007 initial test calibration sample datasets. N-counts for these samples were over 100,000. Note that student records with missing item scores or $\hat{\theta}$ in the match file were excluded from this study. Also, student records for which PARSCALE did not produce a theta estimate (i.e., $\hat{\theta}$ = 999 reported by PARSCALE) were excluded from the analysis. The study results demonstrated that SCOREST, MULTEST and MIXEDEST yielded acceptable thetas in comparison to PARSCALE and PEM's ISE. All the absolute differences among the methods were within .001.

Scoring Procedures for the Spring 2011 MME Administration

Upon BAA's approval of item parameter estimates for MME forms (i.e., the initial, makeup or accommodation forms) for MME writing, mathematics, reading and science and raw-to-scale conversions for MME social studies, a score file using the matched file record layout was produced by ACT. This score file contained IDs, MME attemptedness flag, MLE theta estimates computed using SCOREST, MME scale scores and other scores reported for MME. This file was passed to ACT's Measurement Research Department (MRD) for QC checks.

Using the score file, for students who did not meet MME attemptedness criteria by MME subjects, MRD checked that no MME scores (e.g., MLE thetas and MME scale scores) were computed. For students who met MME attemptedness criteria by MME subjects, MRD independently computed MLE thetas, SE of thetas, MME scale scores and SE of MME scale scores using MULTEST or MIXEDEST, and these scores were checked against those in the score file. Other scores in the score file that were checked by MRD include MME raw scores, MME performance levels, all raw scores for all MME standards along with their percent correct and possible points. Also, the high and low MME scale score values for each student were checked. Upon all the scores passing MRD's QC checks, a file was created and passed to ACT's IT team. For the spring 2011 MME administration, all score files delivered to the BAA passed MRD's QC checks.

Chapter 10: Score Precision

Reliability of a test can be regarded as a measure of consistency of the observed test scores. Since consistency entails repeated realization of an entity, reliability of a test is hard to estimate, particularly with a single administration of a test. If observed test scores are not consistent, or unreliable, the inconsistency is due to measurement error, more specifically, random error of measurement. Statistically, reliability is thus defined as the ratio of the true score (i.e., score without random error) variance to the observed score variance. Several methods of estimating reliability coefficients have been presented. Among them, coefficient alpha is popular because it is based on internal consistency from a single administration. When the measurement is used for classification decisions, classification consistency may function as a reliability measure.

Internal Consistency Reliability

For the spring 2011 administration, over 80,000 examinees of the initial test samples were included in the reliability analysis dataset, depending on the content area. Table 10.1 exhibits the alpha coefficients (Cronbach's alpha) for the 2011 spring MME administration.

Table 10.1. Cronbach's alpha for Spring 2011

Assessment	Cronbach's alpha
Writing	0.89
Reading	0.88
Mathematics	0.91
Science	0.88
Social Studies	0.86

Table 10.2 presents the percentage of agreement between two raters on the constructed response items.

Table 10.2. Rater Validity Percent of Agreement for Spring 2011

Absolute Score Difference Between Two Raters	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	86093	72.8	86093	72.8
1	31644	26.76	117737	99.55
2	510	0.43	118247	99.99
3	16	0.01	118263	100
4	1	0	118264	100

Further Evidence of Reliability on ACT Writing

Data from a special study (ACT, 2009) was used to estimate alternate forms reliability of the ACT writing test, where approximately 6,000 high school students took two forms of the essay. Counter-balancing was used so that each form was administered in both the morning/first session and the afternoon/second session.

Approximately 30 different essay forms were used in this study and essays were assigned randomly to both students and a pool of raters. A "test-retest" correlation was computed for each pair of essays by comparing the scores an examinee received on each of his/her two prompts, and the average of these reliability coefficients yielded the value of 0.67. This value includes variability due to different essay forms and different essay raters.

Reliability analyses were also conducted using data from a field test study in which new prompts were administered to students prior to operational use, to examine how well they worked. The BAA requested that ACT provide additional reliability analyses for constructed response items, so ACT conducted the following study to comply with that request. Each examinee responded to two prompts on successive days. The prompts were spiraled to control for sampling error, and administered in counterbalanced order to control for order effects. To carry out these reliability analyses, several prompts were scored in a students x prompts x raters facet model, utilizing a completely crossed design. There were six prompts, each administered to 20 examinees, and scored by two raters on a 1-6 scale. The prompts and examinees were chosen randomly from those in the field test study. Generalizability Theory analyses produced G-coefficients (internal consistency indices of score consistency) for each prompt pair. The median G-coefficient for the writing test was .70 over the six prompt pairs. Prompts and raters contributed negligible amounts to the total variance, which means the level of student achievement, not the particular prompt asked or the particular raters doing the scoring, is what most strongly determines the scores. Lastly, it was found that the median inter-rater reliability was .94 over the 12 prompts in the Generalizability study.

Empirical IRT Reliability

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. Once the mean squared CSEM over examinees is computed, the equation below can be used to compute the reliability given in Table 10.3. In reference to this equation, $\overline{\sigma}^2(E)$ is the mean squared CSEM and $\sigma^2(S)$ is the observed variance of scale scores for the test taken over examinees.

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

Table 10.3. Empirical IRT reliability for Spring 2011

Assessment	Empirical IRT reliability
Writing	0.88
Reading	0.89
Mathematics	0.88
Science	0.86
Social Studies	0.84

MME Scale Score Reliability

Because the MME scale is a linear function of theta, MME scale score reliabilities are the same as the theta reliabilities. Therefore, the reliabilities in Table 10.3 are also the reliabilities of the MME scale scores.

SEM/Information Curves with Cut 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 SEM/information curve for MME Social Studies based on the Rasch model is also exhibited in appendix A. note that all the SEM/information curves are based on MME selected items. The vertical lines represent the performance level cut scores. For MME, the four performance levels are Not Proficient, Partially Proficient, Proficient, and Advanced.

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 $c_1, c_2, \ldots, c_{(H-1)}$ 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 Table 10.4 were obtained from the MME tests administered in spring 2011. The three parameter logistic model, the generalized partial credit model and the Rasch model are used to estimate the 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 10.4 shows the 4 × 4 contingency tables and indices of classification consistency for the MME assessments.

Table 10.4. The 4 × 4 contingency table and classification consistency for the MME assessments for Spring 2011

MME Writing

	Not Proficient	Partially Proficient	Proficient	Advanced
Not Proficient	0.02993	0.03111	0.00007	0.00000
Partially Proficient	0.03111	0.33805	0.06811	0.00000
Proficient	0.00007	0.06811	0.36506	0.01949
Advanced	0.00000	0.00000	0.01949	0.02940

MME Reading

	Not Proficient	Partially Proficient	Proficient	Advanced
Not Proficient	0.04659	0.03587	0.00387	0.00000
Partially Proficient	0.03587	0.11295	0.06085	0.00000
Proficient	0.00387	0.06085	0.57992	0.01813
Advanced	0.00000	0.00000	0.01813	0.02310

MME Mathematics

	Not Proficient	Partially Proficient	Proficient	Advanced
Not Proficient	0.19765	0.04867	0.01269	0.00000
Partially Proficient	0.04867	0.05782	0.04695	0.00001
Proficient	0.01269	0.04695	0.31774	0.03193
Advanced	0.00000	0.00001	0.03193	0.14631

MME Science

	Not Proficient	Partially Proficient	Proficient	Advanced
Not Proficient	0.11057	0.04275	0.01957	0.00000
Partially Proficient	0.04275	0.04774	0.05444	0.00000
Proficient	0.01957	0.05444	0.46834	0.03143
Advanced	0.00000	0.00000	0.03143	0.07696

MME Social Studies

	Not Proficient	Partially Proficient	Proficient	Advanced
Not Proficient	0.08649	0.02447	0.00581	0.00001
Partially Proficient	0.02447	0.03849	0.03084	0.00030
Proficient	0.00581	0.03084	0.16442	0.04432
Advanced	0.00001	0.00030	0.04432	0.49912

Table 10.5 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 10.5. Classification accuracy for the MME assessments using four classification categories Spring 2011

Assessment	Index Value
Writing	0.85
Reading	0.84
Mathematics	0.79
Science	0.79
Social Studies	0.79

Chapter 11: Validity

Validity refers to the extent to which scores reflect what the test is intended to measure. As stated in the *Standards for Educational and Psychological Testing* (AERA, et al.,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 validation is an ongoing process, which begins the moment that work on a test begins and continues throughout the life of the test. Validation is the process of continually accumulating and reviewing evidence from various sources to refine the utility of a test for making recommended interpretations consistent with the intended uses of the test scores.

Construct Validity Evidence from Content and Curricular Validity

Content validity involves the systematic examination of test content to determine whether it covers the curricular standards to be measured. As stated in Chapter 3, the MME augmentation is used to measure content which Michigan educators believe all students should know and be able to achieve in the content areas that are not measured by the ACT and *WorkKeys* assessments. Assessment results quantify how Michigan students and schools perform 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

Prior to the development and implementation of the MME, a committee of educators, item development experts, assessment experts, and BAA staff met annually to review new and field-tested items for use on the MEAP (the old high school assessment). These stakeholders now meet to review new and field-tested items for use in augmenting the MME. The BAA has established a sequential review process, as illustrated in Figure 11.1. This process continues to provide 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 necessary 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 with MME results regarding knowledge of this content. 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.

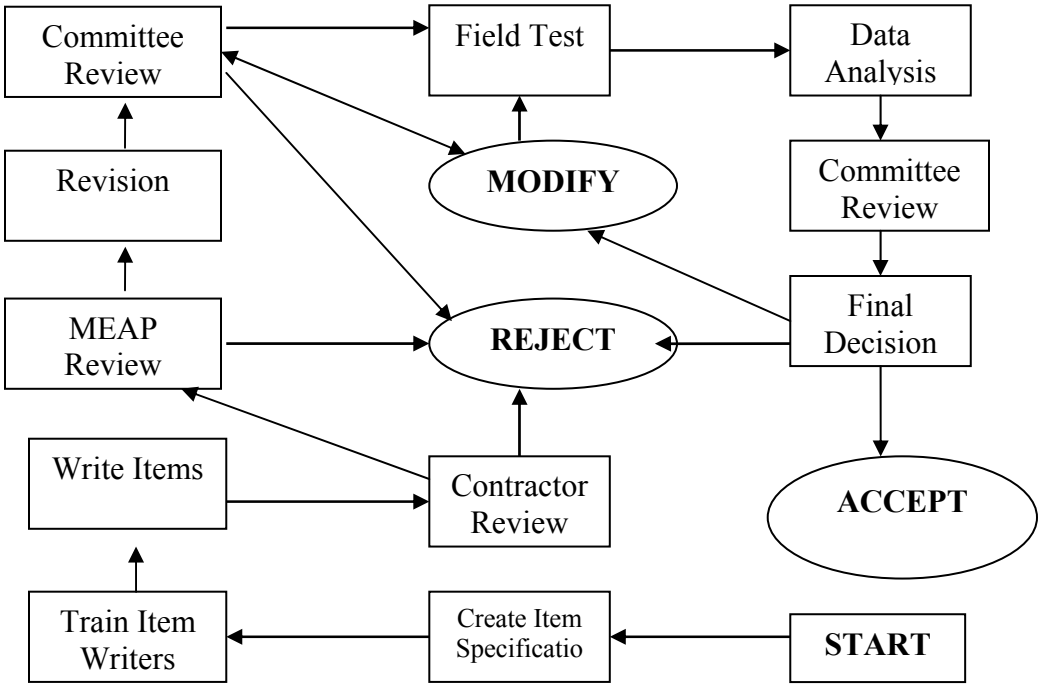


Figure 11.1. Item development/review cycle.

MME Alignment Studies

As detailed in Chapter 5, "Test Development Analyses," 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. Additional alignment studies will be performed in future years.

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 supply 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-related Validity Evidence from Criterion Validity Analyses

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. 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 substantial 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 that the ACT successfully predicts success in college, and that *WorkKeys* successfully predicts workplace success. As a criterion, a strong correlation of MME with *WorkKeys* and the ACT would indicate 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 11.1. 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, 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.

For the MME Spring 2011 administration, the correlations among the MME and the ACT and *WorkKeys* scale scores were as follows. The sample sizes employed for computing these correlations were over 100,000.

- MME Writing and ACT English: .87.
- MME Reading and ACT Reading: .87.
- MME Reading and *WorkKeys Reading for Information*: .84.
- MME Mathematics and ACT Mathematics: .75.
- MME Mathematics and *WorkKeys Applied Mathematics*: .76.
- MME Science and ACT Science: .78.

Table 11.1. 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

Criterion-related Validity Evidence for MME Science

Science standards underwent significant revisions prior to the 2009 MME administration. In order to compile additional evidence for criterion-related validity of the MME science scale scores, additional analyses were conducted. These analyses examine the criterion related validity of MME science scale scores. Using spring 2009 data, three external criterion variables were selected: 1) science course grades, 2) number of semesters students have taken science courses, and 3) whether students have taken advanced science courses.

Average MME science scale scores, grouped by each of the criterion variables are presented in Table 11.2, 11.3, and 11.4 respectively. As shown, the average MME science score increases as the course grade increases for the subjects of General Science, Biology, Chemistry and Physics. Students tend to have higher MME scores if they have taken science courses for a longer period of time, and students who have taken advanced science courses score higher than students who haven't. The criterion related validity of MME science is supported by this evidence.

Table 11.2. Average MME Science Scale Scores, by Course Grade of Science Courses

General Science	MME	Biology	MME	Chemistry	MME	Physics	MME
F	1070	F	1071	F	1079	F	1082
D	1077	D	1080	D	1088	D	1087
C	1085	C	1089	C	1098	C	1096
B	1098	B	1103	B	1112	B	1113
A	1118	A	1122	A	1128	A	1130

Table 11.3. Average MME Science Scale Scores, by Semesters of Science

Number of Semesters of Science	Mean MME Science Score
1	1061
2	1073
3	1079
4	1090
5	1090
6	1101
7	1101
8	1119

Table 11.4. Average MME Science Scale Scores by Students with or without Advanced Courses in Natural Sciences

AP, Accelerated, or Honors Courses in Natural Sciences	Mean MME Science Score
Yes	1118
No	1110

DIF Analyses of the Spring 2011 MME

For the DIF analyses, only the item responses of students who took all three days of the MME in the same administration mode—all initial, all makeup, or all accommodated—were considered. Of those students, only those who had valid flags of "Y" in the match file for a subject area (i.e., students who met attemptedness in the subject area, did not have nonstandard accommodations, did not have prohibited behavior, and were not involved in a misadministration) were included in the data sets on which the DIF analyses were conducted. Even with these restrictions on the data, at least 97% of all students were included in the DIF analyses in each subject area.

Two focal/reference group comparisons were conducted for each item: females compared with males, and African Americans (Blacks) compared with Whites. For each multiple-choice item and each comparison, several statistics were computed: the Mantel-Haenszel delta statistic (MH-D), the value of the associated chi-square statistic (MH-CHISQ), the probability (P) of this chi-square value under the null hypothesis of no DIF, and the ETS A, B, or C category for the item based on the values of MH-D and P. Table 11.5 presents the criteria for the A, B, and C categories. For a further description of these statistics and the categories see, for example, Holland and Wainer (1993). A positive MH-D denotes an item that favors the focal group, while a negative value indicates an item that favors the reference group.

Table 11.5. Criteria for the A, B, and C DIF Categories

Category	Description	Criterion
A	Negligible DIF	Nonsignificant MH-CHISQ ($P > 0.05$) or $\lvert MH\text{-}D\rvert < 1.0$
B	Moderate DIF	Significant MH-CHISQ ($P \leq 0.05$) and $1.0 \leq \lvert MH\text{-}D\rvert < 1.5$
C	Large DIF	Significant MH-CHISQ ($P \leq 0.05$) and $\lvert MH\text{-}D\rvert \geq 1.5$

For the polytomously-scored ACT Writing Test, in place of the MH-D statistic, the standardized mean difference (SMD) index, the standard deviation in Writing Test scores (SD) for the focal and reference groups combined, and the resulting effect size (ES = |SMD/SD|) are computed, as are the AA, BB, and CC classifications resulting from the values of ES and P. Table 11.6 presents the criterion for those classifications. For a further description of these statistics and the AA, BB, and CC categories see, for example, Dorans and Schmitt, (1991). A positive SMD index denotes an item that favors the focal group, while a negative value indicates an item that favors the reference group.

Table 11.6. Criteria for the AA, BB, and CC DIF Categories

Category	Description	Criterion
AA	Negligible DIF	Nonsignificant MH-CHISQ ($P > 0.05$) or Significant MH-CHISQ ($P \leq 0.05$) and $ES \leq 0.17$
BB	Moderate DIF	Significant MH-CHISQ ($P \leq 0.05$) and $0.17 < ES \leq 0.25$
CC	Large DIF	Significant MH-CHISQ ($P \leq 0.05$) and $ES > 0.25$

Matching Criterion

The matching criterion for each comparison was the total raw score for the subject area to which the item belonged. These raw scores are described below:

1. Writing: This was the sum of the ACT English Test raw score and the ACT Writing Test score. This sum ranged from 2 to 87.

2. Reading: This was the sum of the ACT Reading Test raw score and the *WorkKeys Reading for Information* Test raw score. This sum ranged from 0 to 70.

3. Math: This was the sum of the ACT Mathematics Test raw score, the *WorkKeys Applied Mathematics* Test raw score, the raw score on the three *WorkKeys Locating Information Test* items that counted toward a student's MME Mathematics score, and the Day 3 Mathematics Test raw score (operational items only). This sum ranged from 0 to 109.

4. Science: This was the sum of the ACT Science Test raw score and the Day 3 Science Test raw score (operational items only). This sum ranged from 0 to 72.

5. Social Studies: This was the sum of the Day 3 Social Studies Test raw score (operational items only) and the raw score on the six *WorkKeys Locating Information* Test items that counted toward a student's MME SS score. This sum ranged from 0 to 38.

For the *WorkKeys Locating Information* items not selected to count either toward MME Math or Social Studies scores, the matching criterion for each comparison was the *WorkKeys Locating Information* raw score, which ranged from 0 to 32.

For Days 1 and 2, there was just one operational initial form of each test. All students who took the initial ACT form, for example, responded to the same operational items. For Day 3, there were 10 versions of each initial form: Forms 1101-1110. For both the Day 3 Mathematics and Science forms, the set of operational items varied from one version to the next, with a number of items appearing in more than one version. (The operational items for Social Studies were the same across all 10 versions; the versions differed only in field-test items. The Mathematics and Science forms differed in field-test items, as well.)

Because of this, Mathematics or Science raw scores attained on any version of the Day 3 initial form were not directly comparable to those attained on any other version, and therefore the matching criteria described in (3) and (4) above were not comparable across the 10 versions. It was therefore necessary to perform a separate DIF analysis - on the item responses of students who took each of the 10 Day 3 Mathematics initial forms, and on those of students who took each of the 10 Day 3 Science initial forms. The total numbers of Mathematics and Science items appearing in Table 11.7, below, reflect this; while each student took 109 operational Mathematics items and 72 operational Science items, 10 times as many DIF analyses were required for each comparison because of the multiple versions of the Day 3 initial forms.

Tables 11.7 through 11.9 present the number of "A", "B," and "C" operational items, by subject area, for the initial, makeup and accommodated testing. Tables 11.10 through 11.12 break down the "B" and "C" items by the favored group (i.e., males or females, blacks or whites) for the initial, makeup and accommodated testing. Table 11.13 gives the SMD results for the ACT Writing Test.

Table 11.7. Summary of Mantel-Haenszel Results by Focal/Reference Groups and Subject, Operational Items Only, Initial Testing

	Number of items in Category			
	A	B	C	Total
Females/Males				
Writing	74	1		75
Reading	68	2		70
Mathematics	1005	70	15	1090
Science	694	25	1	720
Social Studies	38			38
Locating Information	23			23
Total	1902	98	16	2016
Blacks/Whites				
Writing	69	6		75
Reading	68	2		70
Mathematics	1068	22		1090
Science	705	15		720
Social Studies	37	1		38
Locating Information	22	1		23
Total	1969	47		2016

Table 11.8. Summary of Mantel-Haenszel Results by Focal/Reference Groups and Subject, Operational Items Only, Makeup Testing

	Number of items in Category			
	A	B	C	Total
Females/Males				
Writing	73	1	1	75
Reading	68	1	1	70
Mathematics	96	7	6	109
Science	67	5		72
Social Studies	36	1	1	38
Locating Information	7	5	11	23
Total	347	20	20	387
Blacks/Whites				
Writing	71	4		75
Reading	65	5		70
Mathematics	98	6	5	109
Science	68	4		72
Social Studies	37	1		38
Locating Information	20	3		23
Total	359	23	5	387

Table 11.9. Summary of Mantel-Haenszel Results by Focal/Reference Groups and Subject, Operational Items Only, Accommodated Testing

	Number of items in Category			
	A	B	C	Total
Females/Males				
Writing	75			75
Reading	69	1		70
Mathematics	104	5		109
Science	72			72
Social Studies	37	1		38
Locating Information	24			24
Total	381	7		388
Blacks/Whites				
Writing	75			75
Reading	70			70
Mathematics	106	3		109
Science	71		1	72
Social Studies	38			38
Locating Information	21	3		24
Total	381	6	1	388

Table 11.10. Numbers of Category "B" and "C" items, by favored group, Initial testing

	B		C		
	Females	Males	Females	Males	Total
Writing		1			1
Reading		2			2
Mathematics	48	22	4	11	85
Science	6	19		1	26
Social Studies					
Locating Information					
Total	54	44	4	12	114
	Blacks	Whites	Blacks	Whites	Total
Writing	3	3			6
Reading		2			2
Mathematics	9	13			22
Science	3	12			15
Social Studies		1			1
Locating Information		1			1
Total	15	32			47

Table 11.11. Numbers of Category "B" and "C" items, by favored group, Makeup testing

	B		C		
	Females	Males	Females	Males	Total
Writing		1		1	2
Reading		1		1	2
Mathematics	2	5	6		13
Science	3	2			5
Social Studies	1			1	2
Locating Information	2	3	5	6	16
Total	8	12	11	9	40
	Blacks	Whites	Blacks	Whites	Total
Writing	2	2			4
Reading	3	2			5
Mathematics	2	4	4	1	11
Science	2	2			4
Social Studies		1			1
Locating Information	2	1			3
Total	11	12	4	1	28

Table 11.12. Numbers of Category "B" and "C" items, by favored group, Accommodated testing

	B		C		
	Females	Males	Females	Males	Total
Writing					
Reading	1				1
Mathematics	3	2			5
Science					
Social Studies	1				1
Locating Information					
Total	5	2			7
	Blacks	Whites	Blacks	Whites	Total
Writing					
Reading					
Mathematics		3			3
Science				1	1
Social Studies					
Locating Information		3			3
Total		6		1	7

Table 11.13. Summary of SMD results, by focal/reference groups, all testings

	Females/Males		Blacks/Whites	
Testing	Category	Group Favored	Category	Group Favored
Initial	BB	Females	AA	
Makeup	CC	Females	AA	
Accommodated	CC	Females	AA	

Validity Evidence for the Day 1 Stand Alone Component: ACT Assessment

Validity is often categorized into several types such as content validity, construct validity, and criterion-related validity. More fundamentally, validity can be defined as "the appropriateness, meaningfulness, and usefulness of the specific inferences made from test scores" (AERA et al., 1999, p.9). Since ACT scores can be used for diverse purposes, ACT scores have been thoroughly studied for common interpretations and uses, such as measuring educational achievement, making admissions decisions, making course placement decisions, evaluating the effectiveness of high school college-preparatory programs, and evaluating students' probable success in the first year of college and beyond (ACT, 2007b). The following is a brief summary of the validity evidence of the ACT for its various uses. For the technical details such as descriptive and inferential statistics, see *The ACT Technical Manual* (ACT, 2007b).

Measuring Educational Achievement

Among the validity types, content validity is particularly important for the use of ACT to measure educational achievement. The ACT tests are designed to measure students' problem-solving skills and knowledge in particular subject domains and are closely reviewed to ensure that the test content represents current high school and university curricula. This content validation process is standardized so that the ACT test scores can have the same meaning for all students, test forms, and test dates. Statistical analyses were also conducted. For example, ACT test results were compared with high school grades, and a strong relationship was found between them. Also, longitudinal growth was investigated using ACT, PLAN, and EXPLORE: the three testing programs of ACT's Educational Planning and Assessment System (EPAS). The large intertest correlations and the increases in the average scores indicate EPAS is measuring educational achievement as students progress through the grades.

Making Admissions Decisions

Appropriate admissions decisions are important for students, parents, and postsecondary institutions alike. For this use of ACT tests, validity can be measured in relationship with first-year college grades and GPAs. Validity of ACT test scores and high school grades are also important since they can serve as multiple measures for making college admission decisions. Research studies conclude that the ACT Composite scores provide greater differentiation across levels of academic achievement during the first year of college than do high school GPAs, in terms of probable success during the first year of college.

Course Placement Decisions

The ACT tests were also designed to facilitate placement of first-year college students to appropriate-level courses such as "standard," "remedial," or "advanced." Helping with placement decisions is accomplished by the close connection of the ACT test battery with subject matter content. The content specifications of the ACT tests are based on the recommendations of nationally representative panels of secondary and postsecondary educators. Statistical relationships of ACT scores with course grades and high school GPAs have also been investigated by subjects and for subgroups. It was found that a typical institution using the ACT optimal cutoff score from their data could expect that at least 64% of the placement decisions would be correct decisions.

Indicators of Educational Effectiveness

The ACT tests can be used to evaluate college-preparatory programs since they have been developed to measure academic skills and knowledge that are obtained in high school and are necessary for academic success in the first year of college. However, a content review should be conducted to determine the extent to which the tests represent important outcomes the college-preparatory programs wish to measure.

Evaluating Probable College Success

The use of the ACT tests to evaluate probable college success is closely connected with the other uses of the ACT. According to recent studies, the ACT College Readiness Benchmarks show that students who are college-ready are more likely to immediately enroll in college, and once they enroll, tend to be more successful during their first year. Also, the ACT College Readiness Benchmarks can assist in determining who will succeed in college, even into the second year.

Validity Evidence for the Day 2 Stand Alone Component: the *WorkKeys* Assessments

WorkKeys assessments are designed for use in both business and educational settings. To support these uses, ACT has adopted a multi-faceted approach to validation of *WorkKeys* Assessments: *Reading for Information* (RFI), *Applied Mathematics* (AM), and *Locating Information* (LI). Three types of validity evidence have been collected to justify the use of *WorkKeys* assessment scores, including content-related evidence, criterion-related evidence, and construct related evidence. To accumulate such evidence, ACT has conducted validity studies or worked with organizations to collect data on students and employees. The results are reported in *WorkKeys* Assessment Technical Manuals (ACT, 2008a, 2008b, and 2008c).

Content-Related Evidence

To support the content-related validity of the three test scores, ACT uses two job analysis procedures—*WorkKeys* Job Profiling and the SkillMap Job Inventory—to link the *Reading for Information*, *Applied Mathematics*, and *Locating Information* Skill Levels, to relevant job behaviors. *WorkKeys* Job Profiling and SkillMap are both designed to meet federal standards and other industry guidelines for content validation of employment tests used for high-stakes decisions such as hiring and promotion. Both procedures can be used to define critical job tasks, determine which *WorkKeys* skills are relevant to performing the tasks, and identify the level of skill required for performing them.

Criterion-Related Evidence

To support the criterion-related validity of the three test scores, ACT has gathered data from various organizations on the correlation between the test scores used to select job applicants and their subsequent job performance ratings. While sample sizes and correlations vary from study to study, all of the correlations have been positive, ranging from 0.12 to 0.86, which compares favorably with the correlations typically found in the general research literature on criterion-related validity of employment tests. ACT has also conducted classification consistency studies, comparing the employees' job performance classification to their classification by *WorkKeys* Assessment Skill Levels. In these studies, the percentage of employees classified the same way by both measures ranged from 30 percent to 100 percent depending on sample size, skill level, and participants.

Construct-Related Evidence

To support the construct-related validity of the three test scores, ACT examined the relationship between *WorkKeys* assessments and other tests measuring somewhat similar skills and found a moderate correlation between the test scores. In addition, ACT examined the relationship among the three fundamental skill assessments. Initial results suggest that (1) each assessment measures unique job-related skills as well as some general skills, and (2) each assessment has a strong unidimensional structure.

Gender and Race/Ethnicity Analyses

Because *WorkKeys* assessments can be used for high-stakes employment decisions, ACT has analyzed Skill Level scores for evidence of adverse impact by gender and racial/ethnic groups. Evidence of adverse impact has been found to be consistent with existing research on the validity of employment test scores used for high-stakes selection decisions. In this context, such findings reinforce the need to clearly link use of *WorkKeys* test scores to the critical tasks and skills required for the job.

Fairness Review

Fairness Review for the ACT

According to the *Code of Fair Testing Practices in Education* (Joint Committee on Testing Practices, 2004), test developers should provide "tests that are fair to all test takers regardless of age, gender, disability, race, ethnicity, national origin, religion, sexual orientation, linguistic background, or other personal characteristics" (p.1). As a testing organization, ACT endorses the *Code* and makes every effort to see that all ACT tests are fair to the populations for which the tests are intended. The work of ensuring test fairness takes place during every stage of the test development process, including item (test question) writing and review, item pretesting, forms construction, and forms review. ACT is committed to ensuring that each of its testing programs upholds the *Code*'s standards for appropriate test development practice.

Item Writing, Review, and Pretesting

Most of the individuals who write items for ACT's tests are actively engaged in teaching at the high school or the university level. ACT makes every attempt to include item writers who represent the diversity of the population of the United States with respect to ethnic background, gender, and geographic location. Item writers work closely with ACT test development associates in producing items and passages of high quality that are designed to meet the test specifications, represent diversity, and be fair to all examinees. Item writers are provided with detailed guidelines to assist them in developing test materials, including specific information on fairness concerns. Among these is the way in which various groups of the population are portrayed, and the degree to which representatives of various groups are depicted in active versus passive circumstances, as exhibiting stereotypic mental or physical characteristics or tendencies, or as engaged in particular occupations or roles. In addition, the construction of fair test items requires sensitivity to the changing circumstances of our society: increased variation in family structures; the multiethnic composition of the population; and a wide range of socioeconomic and urban, suburban, and rural lifestyles.

Every new test item and passage is comprehensively reviewed for fairness, interest, and appropriateness to the grade level for which the test is designed; adherence to specifications; soundness and defensibility;

grammatical accuracy; and sound measurement characteristics. The items are first reviewed by ACT's test development associates and editorial staff. Any problems found in this review are corrected immediately. The items are then sent to two groups of external reviewers: content experts (including classroom teachers, college faculty, and curriculum specialists, representing diversity as to geographic region, ethnicity, and gender), who focus on content accuracy, item classifications, skill levels, and grade-level appropriateness; and fairness reviewers, who are of diverse ethnicity, gender, and geographic background and are sensitive to issues of test and item fairness. The fairness reviewers carefully examine all items and stimulus materials to make sure that they do not contain any language, roles, situations, or contexts that could be considered offensive or demeaning to any population group.

ACT selects fairness reviewers from among African American, Asian American, Latino, American Indian, and female consultants. ACT communicated with prominent, nationally recognized advocacy groups to obtain nominees who could review the ACT test materials. (See the ACT publication, *Fairness Report for the ACT Tests 2009–2010*, for a list of these organizations.) From the recommendations of these groups, ACT selects fairness reviewers who have a history of active participation in promoting the concerns of the group within educational settings and beyond.

Items that pass the content and fairness reviews are pretested on a representative sample of the ACT examinee population in a national administration of the ACT. The purpose of pretesting items is to determine whether the items are technically sound and at the appropriate level of difficulty for the ACT examinee population. Statistical indices of item difficulty and discrimination, among other statistics, are compiled on the basis of pretest results. Items are evaluated according to their performance in the pretest. Those that perform acceptably on all criteria are included in the item pool from which preliminary forms of the ACT tests are constructed.

Operational Forms Construction

Preliminary forms of the ACT tests are constructed using the items that survive the pretest. Items are selected to match the requirements of both the content and statistical specifications for the tests. The distributions of the items in each test form are also examined for fairness, variety, diversity, and balance. Each test form is balanced with regard to multicultural and gender representation. While it is impossible, given the constraint of the limited amount of material in each test form, to represent every group in every form, a good-faith effort to represent diversity should be discernible in every final form. Two strategies ACT uses to attain this diversity are ensuring the inclusion of culturally diverse passages within each form and ensuring that all passages depict universal themes applicable to all groups.

Preliminary versions of ACT test forms are subjected to the same comprehensive reviews for content and fairness issues that items undergo. ACT's test development associates, editorial staff, and measurement staff conduct the initial forms reviews. The forms are then sent to external content experts and fairness reviewers (not the same individuals who conducted the reviews prior to pretest). The comments made by all reviewers are collated, and the items/passages identified as problematic are replaced as necessary. ACT's test development associates work with the fairness and content reviewers to prepare the final forms of the test for printing. In all, at least sixteen independent reviews are made of each test item before its appearance on a national form of the ACT, primarily to ensure that each student's level of achievement is accurately and fairly evaluated.

Differential Item Functioning (DIF)

Fairness in the content of the items does not necessarily prevent items from functioning statistically in different ways for different population subgroups. Differential item functioning (DIF) can be described as a statistical difference between the probability of a specific population group (the "focal" group) getting the

item right and a comparison population group (the "base" group) getting the item right *given that both groups have the same level of expertise* with respect to the content being tested.

To detect the existence of differential item functioning (DIF) for items in each test form, ACT analyzes the response data from actual national and state administrations of each of the forms. After each national or state administration of a test form, large random samples representing the examinee groups of interest are selected from the total number of examinees taking the test. The groups compared are African Americans/Caucasians, Mexican Americans/Caucasians, Hispanics (other than Mexican origin)/Caucasians, Asian Americans or Pacific Islanders/ Caucasians, American Indians or Alaska Natives/Caucasians, and females/males. The statistics ACT uses for detecting DIF are the standardized difference in proportion-correct (STD) and the Mantel-Haenszel common odds-ratio (MH). The samples of examinees' responses to each item are analyzed using the STD and MH procedures. All items with MH and/or STD values exceeding a preestablished statistical tolerance level are flagged for further review. The flagged items are reviewed by ACT's test development associates for possible explanations of the unusual STD or MH results. In the event that a problem is found with an item, actions are taken as necessary to eliminate or minimize the influence of the problem.

Fairness review for *WorkKeys*

Fairness review is an important step in developing and pretesting news items for the three *WorkKeys* assessments. Participants in fairness review include ACT test development staff and external business and education experts from diverse cultural and ethnic backgrounds. Fairness reviewers representing gender, cultural, and ethnic/racial subgroups work to ensure that no item was unfair to any minority group members. ACT gives the reviewers written guidelines and requires them to write an evaluation of each item. ACT reviews the evaluations and responds to any concerns the reviewers raise. Any item rejected by the reviewers is removed from the operational pool. Items that pass reviews and meet specifications are left intact to preserve the accuracy of the pretest item data. Such items constitute the pool from which subsequent operational forms are drawn. Please see more details in *WorkKeys* Assessment Technical Manuals (ACT, 2008a, 2008b, and 2008c).

Conclusion

The evidence from the methods used for item development, item review, augmentation, alignment, and correlation with related measures provide validity-related evidences for the interpretation of MME scores. Given the desired interpretation of scores as described in this chapter, the validity-related evidence strongly support the interpretability of the MME scores.

Chapter 12: 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 2011 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 Chapter 3, 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 versus female comparison and white versus 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, 5th, 25th, 50th, 75th, 95th 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 reviewed by Pearson and the BAA for creating item labels for the 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. Table 12.1 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 versus female and white versus black;
6. Review decision.

All statistics prepared for the review committee are explained in Appendix C. When the p-value for an MC item, or Rasch difficulty was out of the desired range, a difficulty flag was shown. When a point-biserial correlation for an MC 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 item are presented in Table 12.2.

For further psychometric reference, conditional mean plots and box & whisker plot for two student group comparison, male versus female and white versus black were prepared for the flagged items for the two post-field-test reviews. See Figure 12.1for conditional mean plots and Figure 12.2 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 12.1. Item Label for an 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

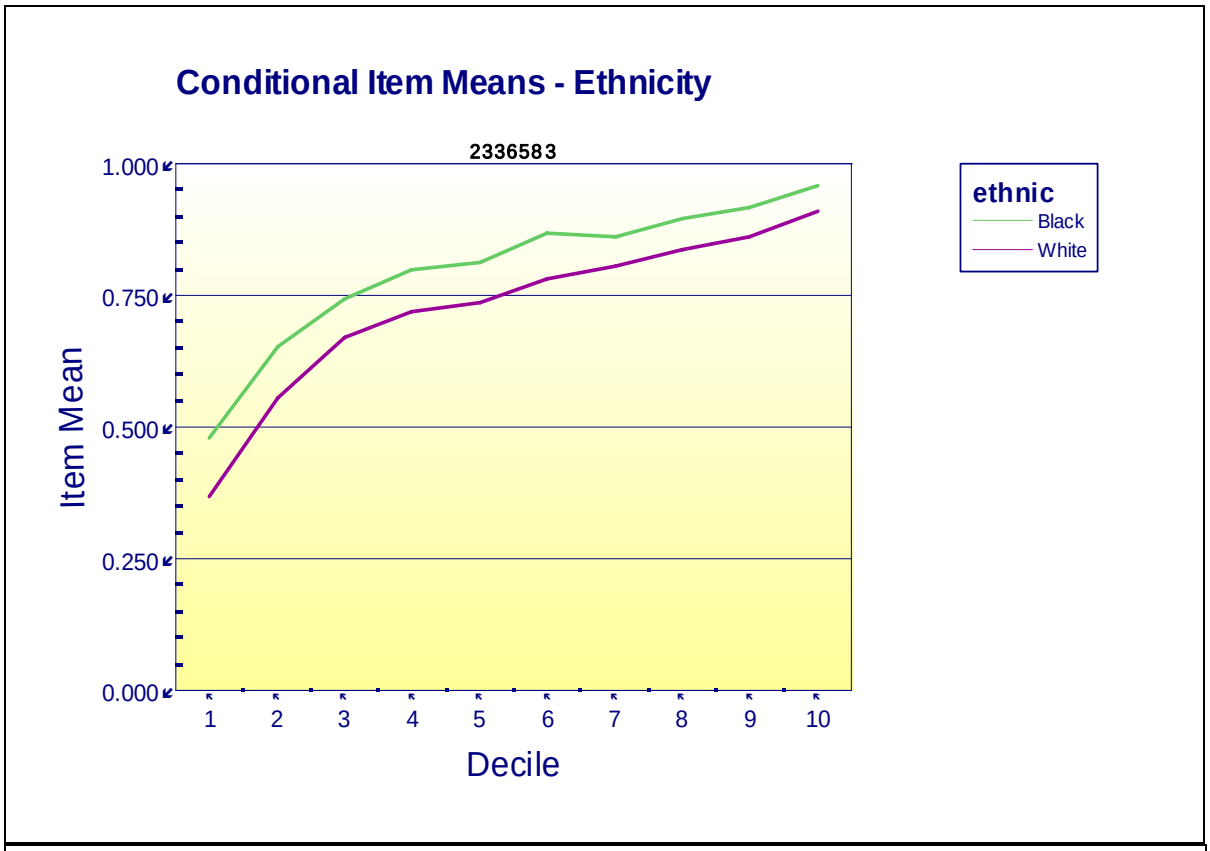
Explanation of DIF Flags
Blank - No or negligible DIF
B - Moderate DIF
C - Large DIF

Table 12.2. 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 or point biserial 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%**



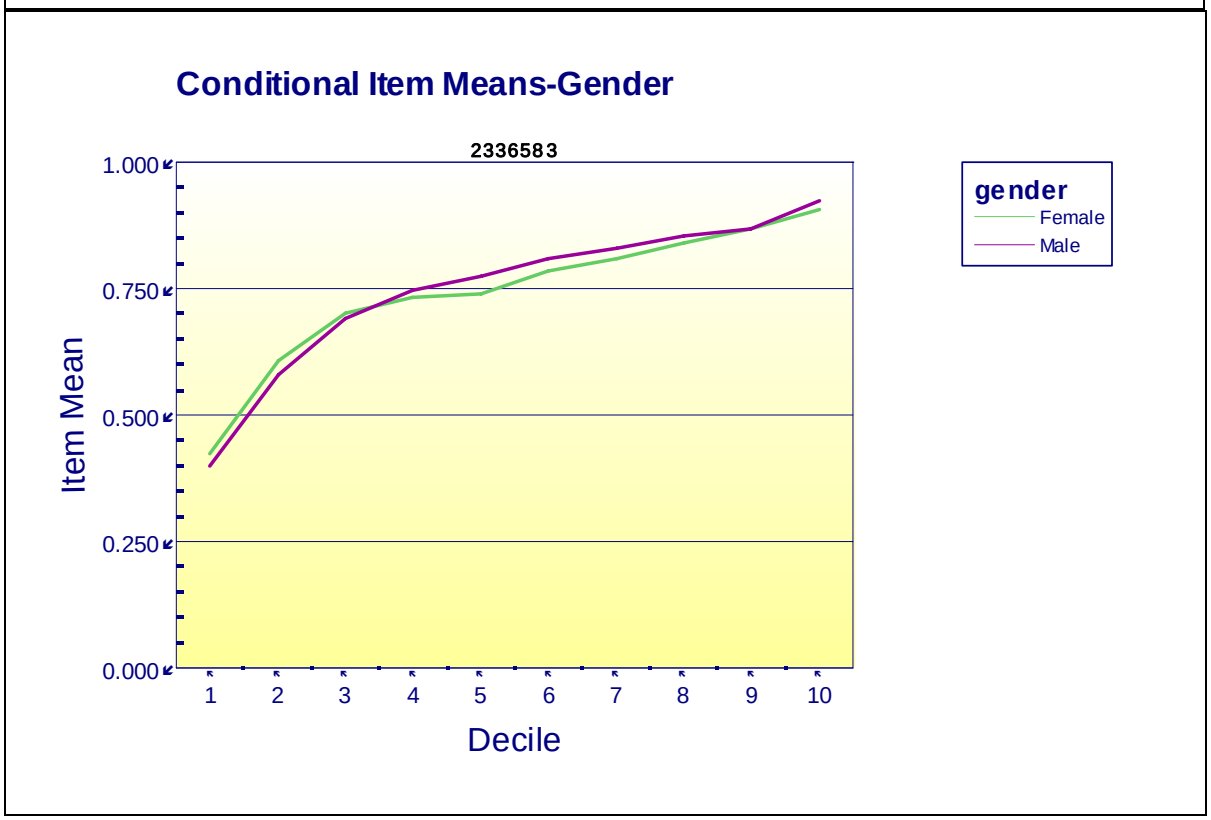


Figure 12.1. Conditional item mean plots for ethnicity and gender for MC items.

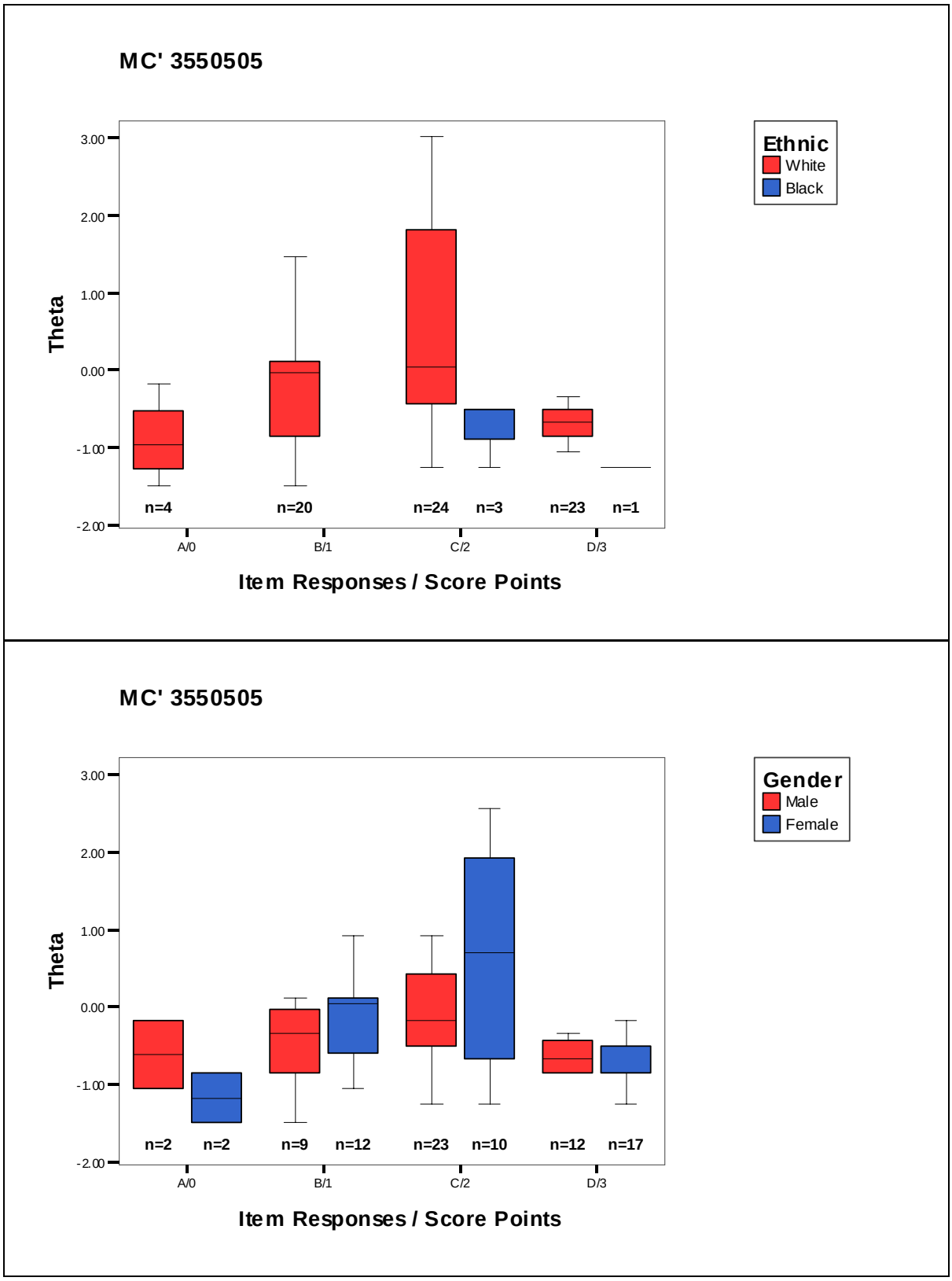


Figure 12.2. box & whisker plots for ethnicity and gender for MC items.

Chapter 13: Standard Setting

Standard setting is a complex and detailed procedure that requires extensive documentation, particularly given the high-stakes nature of standard setting for state-administered assessments. In order to provide the most complete information possible regarding standard setting, those interested in learning more about the standard setting process are asked to reference the complete standard setting document, found at http://www.michigan.gov/documents/mde/Item_BB_177386_7.pdf. Below is a brief overview of the main activities involved in standard setting for the MME. Standard setting was conducted for the MME in 2006 using the procedures outlined below.

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

The results of a modified item mapping procedure are described in the *Standard Setting Report* (Assessment and Examination Service, 2006). The following modified item mapping method was used: "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 MME standard setting were obtained from a group of 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. The MME cut scores are reported in Table 13.1. These cut scores were placed on the MME scale.

A *Michigan Department of Education Memorandum* in October 2006 described four possible 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 specifications. 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).

Table 13.1. Proficiency level cut scores by subject.

	Performance Standard Cut Score		
	Not Proficient/ Partially Proficient	Partially Proficient/ **Proficient**	Proficient/ **Advanced**
Writing	1051	1100	1146
Mathematics	1089	1100	1128
Reading	1078	1100	1158
Science	1087	1100	1143
Social Studies	1086	1100	1129

Chapter 14: Adequate Yearly Progress and Education YES

The major policy-based uses of assessment data from the MME, MEAP, MEAP-Access 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) 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 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 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 90%. 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 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 Michigan Student Data System (MSDS—formerly the 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. The initial 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.

Michigan's AYP targets for 2010–11 were:

74% - Elementary Mathematics
77% - Elementary Reading
66% - Middle School Mathematics
74% - Middle School Reading
67% - High School Mathematics
79% - High School Reading

Because NCLB is currently focused on mathematics and reading scores, the scores of all tested students must be used in the AYP determination for those subjects. 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 2010–11 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 standard error of measurement (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 reading 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 2010–11 score based on the school's scores from 2008, 2009 and 2010. 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 2010–11 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 (2009, 2010 and 2011) in the content area and grade level.

Schools and districts had opportunities throughout the school year to correct or appeal certain data related to student assessments and accountability. Appeals concerning issues and data that had earlier, separate appeals windows will not be accepted.

Earlier appeals/update windows were held for the following:

- Student demographics including enrollment, economic disadvantaged status, limited English proficient status, student with disability status, racial/ethnic status, and Primary Education Providing Entity (PEPE) status (Tested Roster)
- Missing tests/non-standard accommodations/prohibited behavior (Tested Roster)
- Student participation (Students Not Tested)
- Graduation rates (GAD Window)

A school district has the opportunity to appeal any data that do not have separate appeals windows (see above) that affect the AYP status of its schools if it has evidence that the data may be inaccurate. MDE will

process appeals submitted within the appropriate appeal window. The purpose of the appeal window is to address substantive issues regarding the draft School Report Cards. The school district must cite specific data that is challenged in the appeal.

Education YES! provides both a snapshot of current school performance and a roadmap for educators, supplying feedback and direction to assist them on a path of meaningful change. Michigan replaced the original 11 performance indicators with Indicators that are based on the School Improvement Framework. Based on a review of the research on school improvement, rubrics to measure 40 key characteristics have been selected as having the greatest effect on student achievement. The State Board of Education approved the Performance Indicators in December, 2006.

MDE works in partnership with the North Central Association Commission on Accreditation and School Improvement (NCA CASI), and AdvancED (the parent organization of NCA CASI). As one of the first steps in that partnership, a website was developed through AdvancED for schools, as well as for local and intermediate school districts. Through the site, all schools reported on their *Education YES!* School Performance Indicators, completed the School Process Rubrics (90) (SPR (90)) if applicable, and completed reports such as the NCA CASI ASSIST Self Assessment (ASSIST SA) and the NCA CASI Self Assessment (SA). The website is designed to reduce duplication of paperwork and reports. There are four scenarios for schools:

1. NCA CASI member schools that completed the ASSIST Self Assessment (ASSIST SA) fulfilled the building's *Education YES!* reporting requirements.

2. NCA CASI member schools that are not completing the ASSIST SA this year were required to complete the Self-Assessment developed by AdvancED. Completion of this report fulfilled the building's *Education YES!* reporting requirements.

3. Other schools (not members of NCA CASI) had the opportunity to update their *Education YES!* Performance Indicator self-assessment and rating on the website. The 40 Performance Indicators remained the same as last year and will inform the work schools do in the School Process Rubrics (90), if applicable.

4. Schools selected to complete the School Process Rubrics (90) completed that document on the website fulfilling the requirement to do a comprehensive needs assessment. The School Process Rubrics (90). The completion of this report fulfills Education *YES!* Reporting requirements.

All four scenarios fulfilled the *Education YES!* reporting requirements and, in the case of NCA CASI schools, eliminated the need for two reports to be filed. All reports are now available to MDE for inclusion in the calculations for the *Education YES!* School Report Card.

For school year 2010–2011, the Indicators will not be scored. Schools will get full "credit" for the Indicators (which count for 1/3 of the *EducationYES!* grades) for the completion of their assigned Indicators report. Note that all of the reporting scenarios will be treated in the same manner for the purposes of credit and scoring.

Scores and grades are calculated for each content area for each school. The content areas remain the same, using only reading 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 14.1. Education YES! Composite Score Weighting

Component	Point Value
School Performance Indicators	33
Achievement Status	34
Achievement Change	33
Total	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 14.2. 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 2009, 2010, and 2011 are summarized in Tables 14.3 through 14.6 below.

Table 14.3. Results of Accountability Analyses

Report on Michigan School AYP 2009, 2010 and 2011					
		Total Number of Schools	Elementary	Middle School	High Schools
Final Results for 2011					
Total Number of Schools		3,437	1,968	553	548
Made AYP		2,723	1,773	468	329
		79.2%	90.1%	84.6%	60.0%
Did Not Make AYP		714	195	85	219
		20.8%	9.9%	15.4%	40.0%
Final Results for 2010					
	Total Number of Schools	3,697	2,012	588	546
	Made AYP	3,188	1,833	547	447
		86%	93.6%	93.0%	81.9%
	Did Not Make AYP	509	129	41	99
		14%	6.4%	7%	9.1%

Final Results for 2009				
Total Number of Schools	3,671	1,968	597	537
Made AYP	3,147 85.7%	1,839 93.4%	569 95.3%	382 71.1%
Did Not Make AYP	524 14.3%	129 6.6%	28 4.7%	155 28.9%

Table 14.4. Report on School AYP 2008−09, 2009−10, and 2010−11

		2008-09	2009-10	2010-11
Total Number of Schools Assigned AYP status		3,671	3,697	3,437
Total Number of Schools Not Making AYP		524	509	714
Percent of Schools Not Making AYP		14.0%	14.0%	20.8%
Schools that make AYP using Interim Flexibility Option 1 - Students with Disabilities group		327	N/A*	N/A
Schools Identified for Improvement		514	453	386
Schools Identified for Improvement by Phase	Identified for Improvement	165	99	80
	Identified for Improvement (SES)	111	119	67
	Corrective Action	69	61	61
	Restructuring	61	68	176
Schools with Graduation Rates under 80%		204	243	229
Schools not meeting Participation target by group**	All Students	147	104	138
	Black	88	51	66
	American Indian	0	0	0
	Asian American	1	2	0
	Hispanic	8	1	4
	White	48	34	63
	Multiracial	0	0	3
	Limited English Proficient	76	2	5
	Students with Disabilities	76	49	65
	Economically Disadvantaged	117	87	113
Schools not meeting Proficiency target by group**	All Students	189	132	244
	Black	57	18	88
	Asian American	1	0	2
	Hispanic	5	0	12
	White	15	23	21
	Limited English Proficient	2	2	26
	Students with Disabilities	123	164	317
	Economically Disadvantaged	93	18	92

***Option 1—Schools that did not make AYP only because of special education proficiency used to get a 15% bonus, but MEAP-Access took the place of this in 2009−10.**

****A school is counted in all student subgroups that did not make AYP. So, there are duplicated counts of schools.**

Table 14.5. Report on Michigan District AYP 2009, 2010, 2011

		Total Number of Districts	Number Met AYP	Percent Met AYP	Number Not Met AYP	Percent Not Met AYP
Final Results for 2011						
	All School Districts	551	514	93.3%	37	6.7%
	K-12 Districts	491	473	96.3%	18	3.7%
	Charters	32	28	87.5%	4	12.5%
	ISDs	28	13	46.4%	15	53.6%
Final Results for 2010						
	All School Districts	550	522	94.0%	28	5.1%
	K-12 Districts	492	483	98.2%	9	1.8%
	Charters	31	29	93.5%	2	6.5%
	ISDs	27	10	37.0%	17	63.0%
Final Results for 2009						
	All School Districts	548	536	97.8%	12	2.2%
	K-12 Districts	492	486	98.8%	6	1.2%
	Charters	31	27	87.1%	4	12.9%
	ISDs	26	24	92.3%	2	7.7%

Table 14.6. State Accreditation Letter Grades 2009, 2010, and 2011

Grade	**2009**		**2010**		**2011**	
	Number of Schools	**Percent of Schools**	**Number of Schools**	**Percent of Schools**	**Number of Schools**	**Percent of Schools**
A	1,680	45.8%	1,842	49.8%	1,765	51.4%
B	1,002	27.3%	928	25.1%	888	25.8%
C	382	10.4%	350	9.5%	228	6.6%
D-Alert	142	3.9%	95	2.5%	130	3.8%
Unaccredited	5	0.1%	3	0.1%	0	0.0%
No Grade	460	12.5%	479	13.0%	426	12.4%
Total	3,671		3,697		3,437	

Chapter 15: State Summary Data

The summary data for the spring 2011 administration are presented in Table 15.1. For each content area, Table 15.1 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 15.1 through 15.5, and in Tables 15.2 through 15.6. Tables 15.7 through 15.11 present the summary statistics for the item parameter estimates for the items employed for MME scoring.

Table 15.1. Spring 2011 Michigan State Average Scores and Percentages in each Performance Level – All Examinees

			Percentages within Performance Levels			
Content Area	N	Average	Not Proficient	Partially Proficient	Proficient	Advanced
Reading	112,918	1108	14%	22%	61%	3%
Writing	113,615	1095	8%	44%	43%	5%
Mathematics	112,250	1095	32%	15%	40%	13%
Science	112,618	1104	24%	15%	52%	9%
Social Studies	112,714	1123	8%	14%	36%	42%

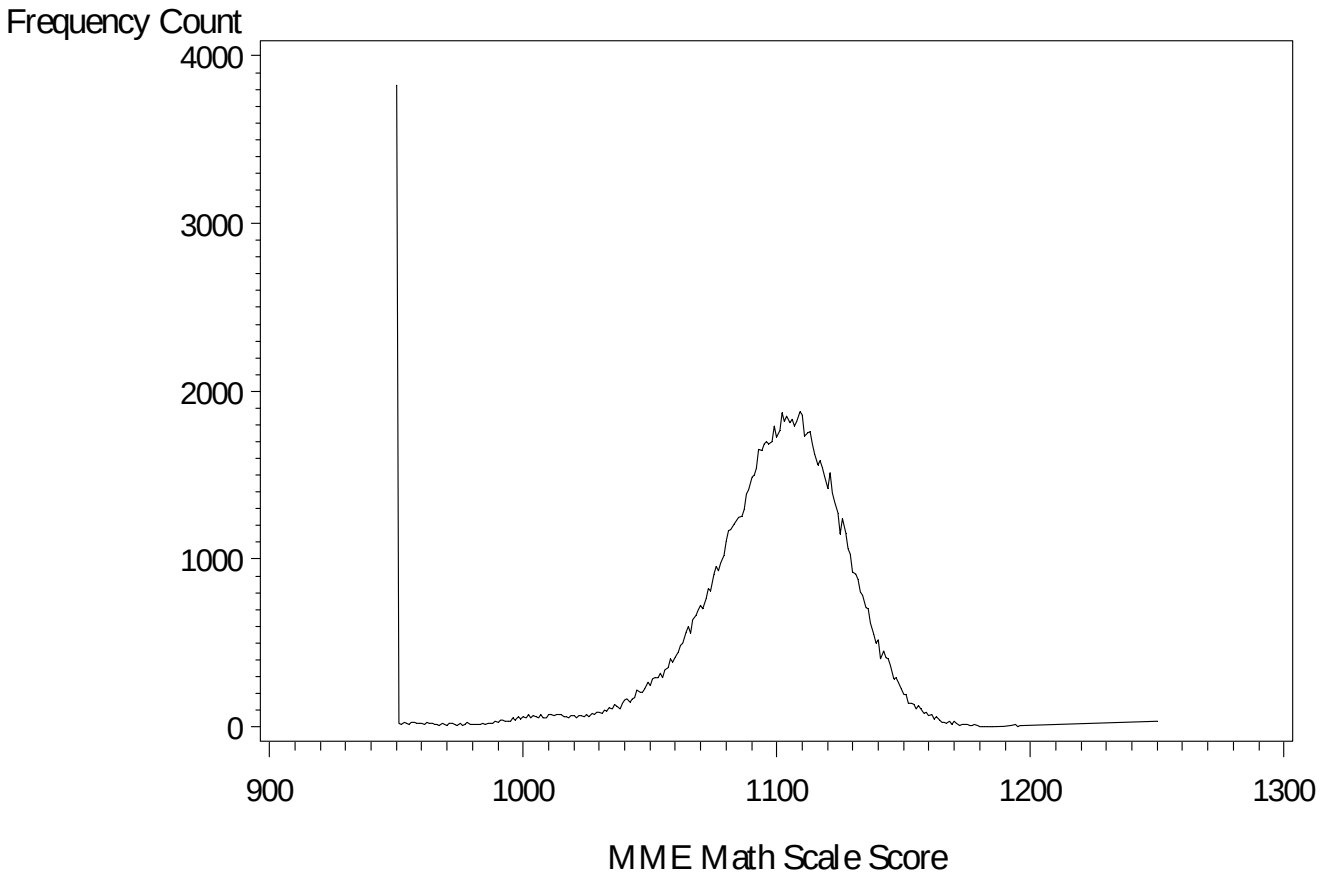


Figure 15.1. Frequency plot for MME Spring 2011 Mathematics scale score total group – All forms included.

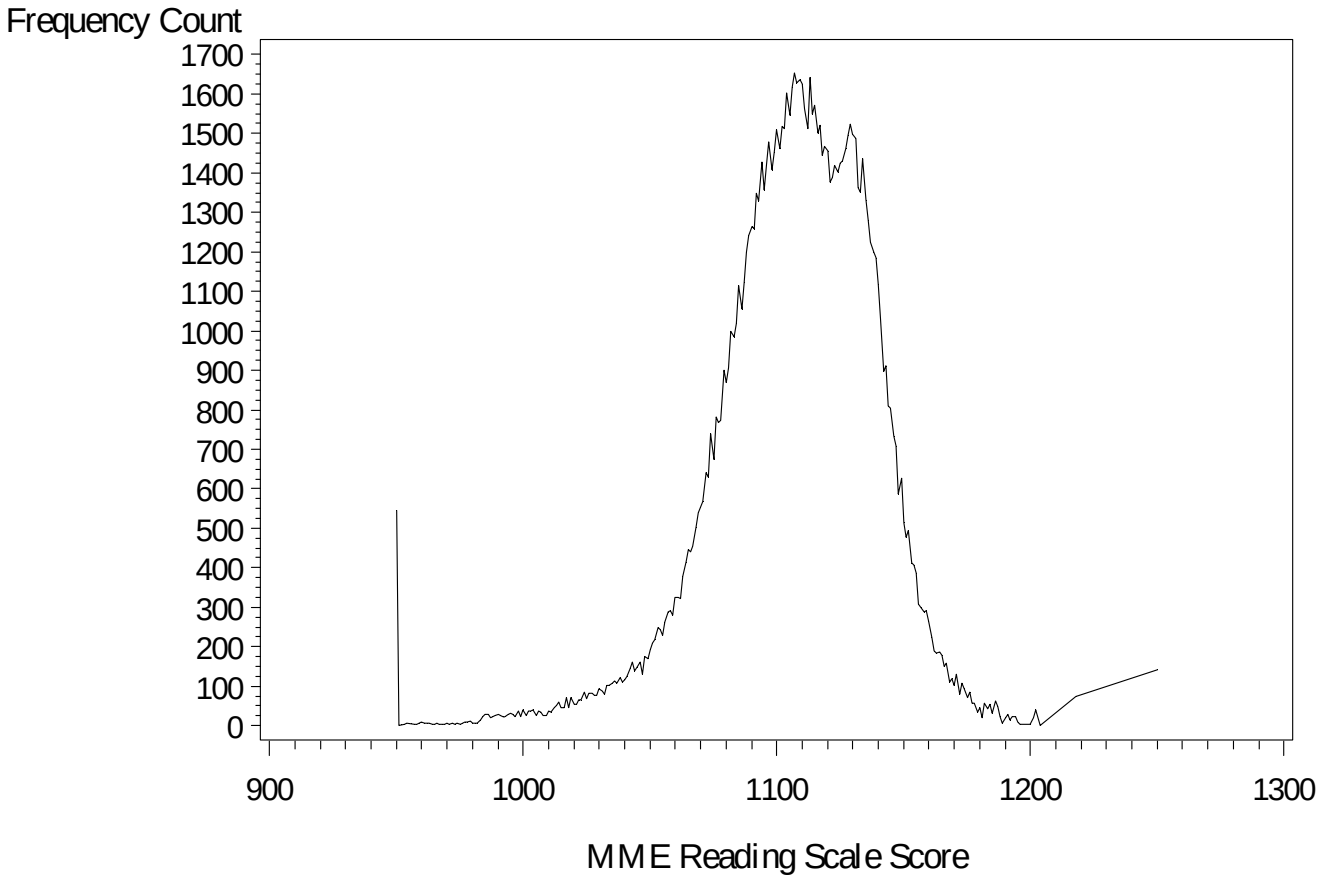


Figure 15.2. Frequency plot for MME Spring 2011 Reading scale score total group – All forms included.

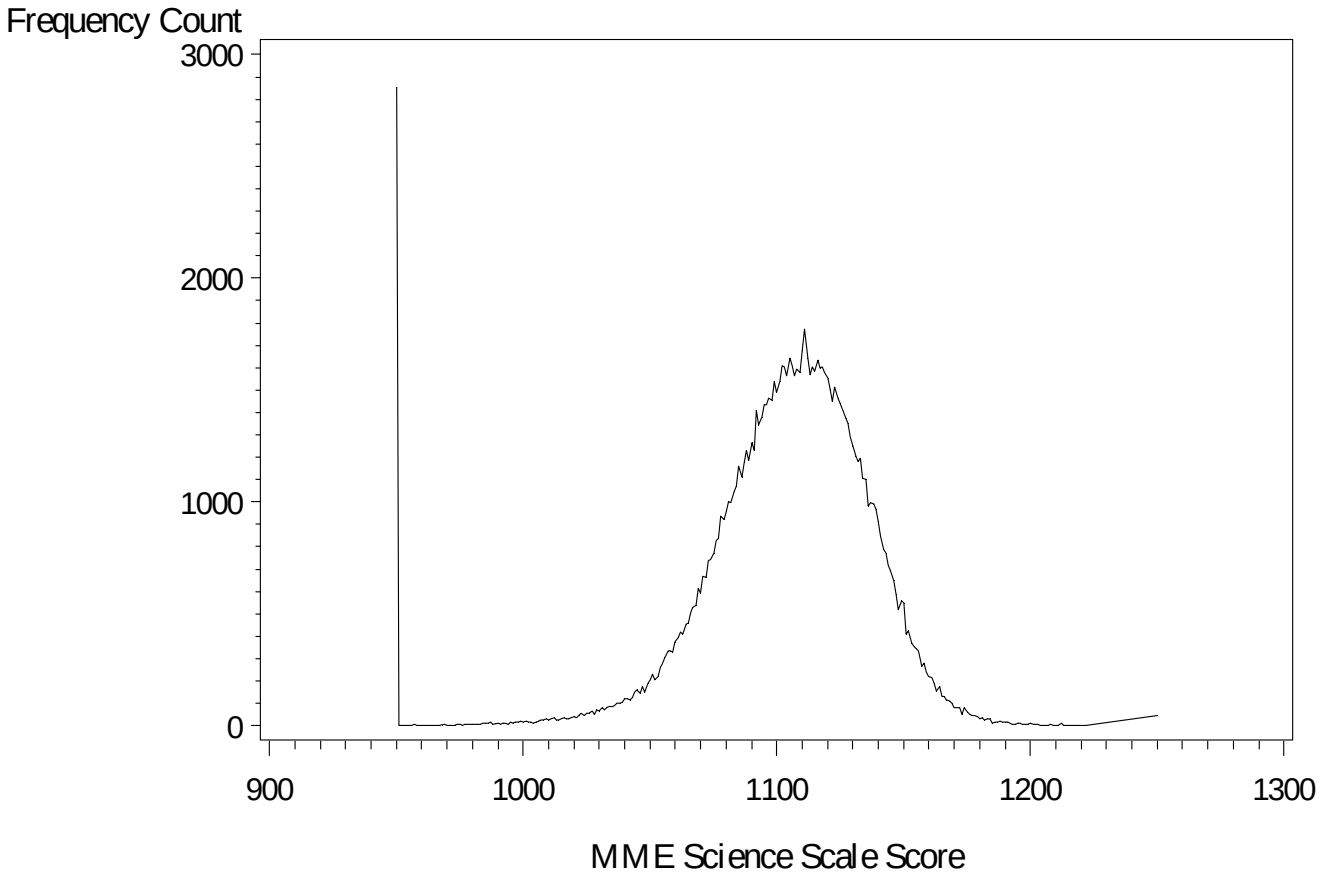


Figure 15.3. Frequency plot for MME Spring 2011 Science scale score total group – All forms included.

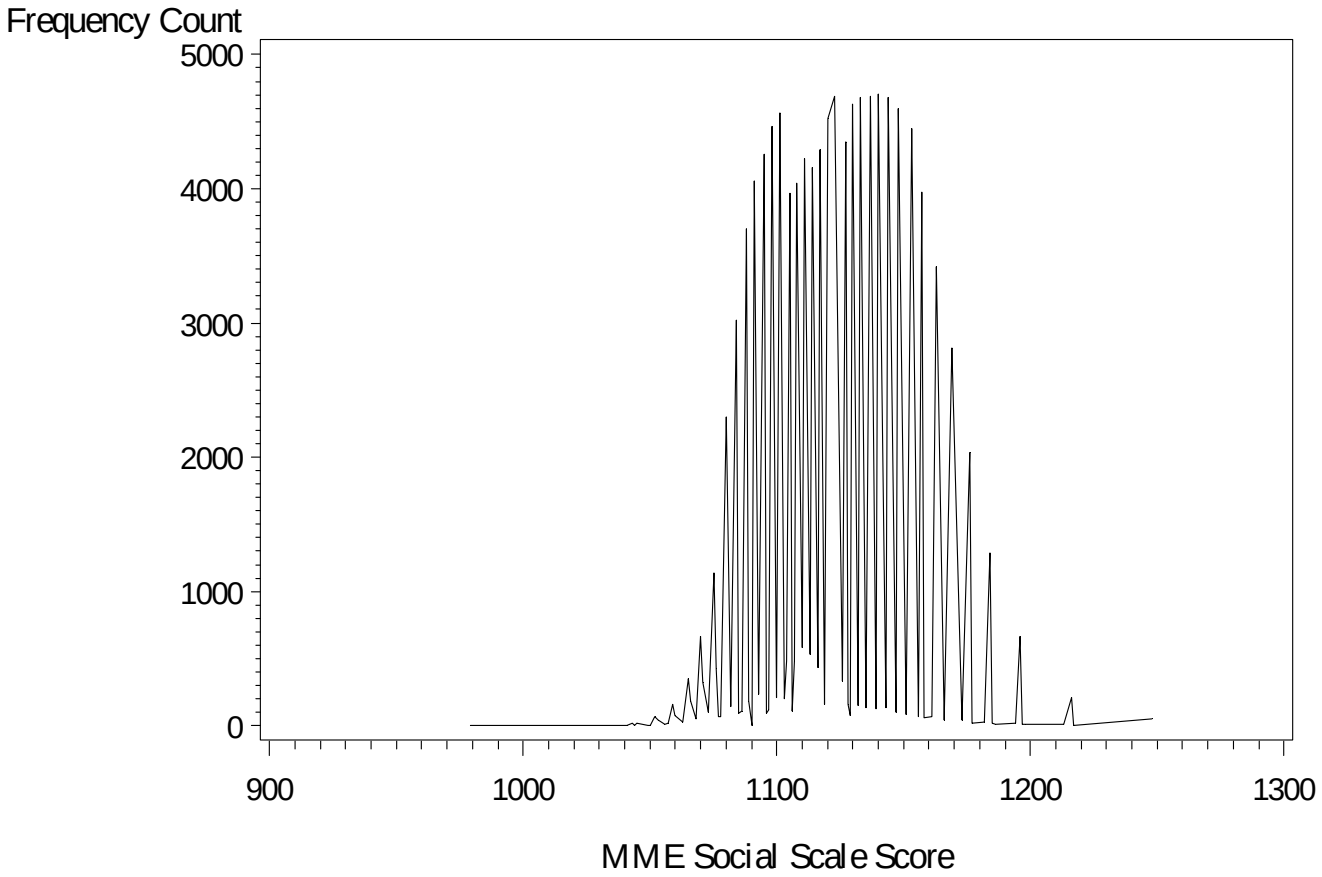


Figure 15.4. Frequency plot for MME Spring 2011 Social Studies scale score total group – All forms included.

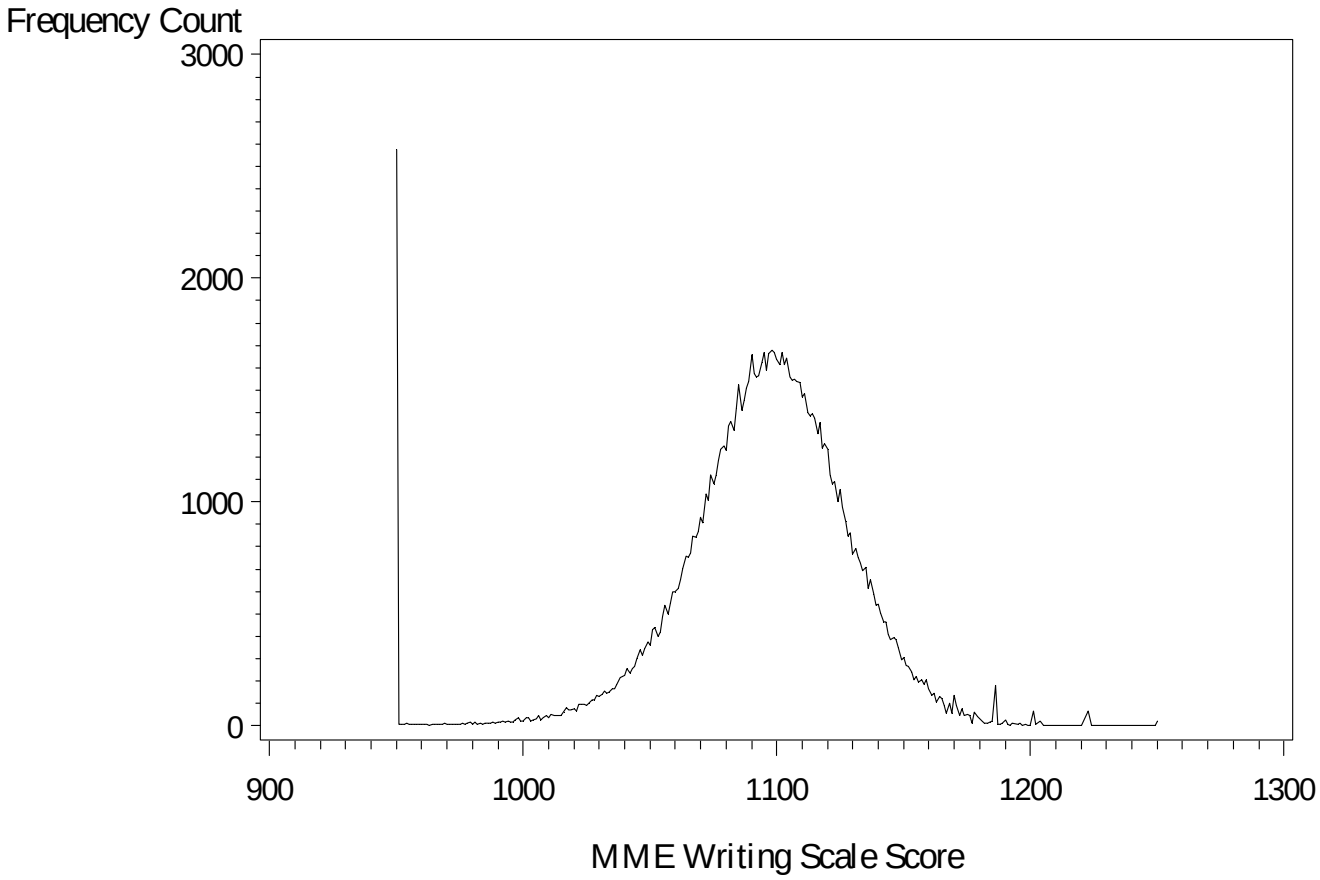


Figure 15.5. Frequency plot for MME Spring 2011 Writing scale score total group – All forms included.

Table 15.2. MME Spring 2011 Mathematics Frequencies for Total Group -- All Forms Included

Scale Score	Frequency	Percent
950	3823	3.28
951	20	0.02
952	16	0.01
953	29	0.02
954	22	0.02
955	15	0.01
956	27	0.02
957	29	0.02
958	24	0.02
959	18	0.02
960	19	0.02
961	17	0.01
962	25	0.02
963	20	0.02
964	20	0.02
965	16	0.01
966	13	0.01
967	8	0.01
968	18	0.02
969	14	0.01
970	9	0.01
971	21	0.02
972	19	0.02
973	16	0.01
974	8	0.01
975	18	0.02
976	11	0.01
977	13	0.01
978	26	0.02
979	14	0.01
980	17	0.01
981	17	0.01
982	17	0.01
983	15	0.01
984	19	0.02
985	16	0.01
986	18	0.02
987	22	0.02
988	24	0.02
989	32	0.03
990	29	0.02
991	43	0.04
992	40	0.03
993	32	0.03
994	36	0.03
995	34	0.03
996	54	0.05
997	42	0.04
998	60	0.05
999	45	0.04
1000	58	0.05
1001	56	0.05
1002	71	0.06
1003	51	0.04
1004	68	0.06
1005	64	0.05
1006	54	0.05
1007	75	0.06
1008	51	0.04
1009	56	0.05
1010	74	0.06
1011	71	0.06
1012	67	0.06
1013	74	0.06
1014	71	0.06
1015	72	0.06
1016	59	0.05
1017	63	0.05
1018	53	0.05
1019	67	0.06
1020	67	0.06
1021	51	0.04
1022	69	0.06
1023	70	0.06
1024	58	0.05
1025	71	0.06
1026	64	0.05
1027	81	0.07
1028	72	0.06
1029	88	0.08
1030	90	0.08
1031	81	0.07
1032	101	0.09
1033	94	0.08
1034	115	0.10
1035	105	0.09
1036	131	0.11
1037	119	0.10

1038	107	0.09
1039	138	0.12
1040	160	0.14
1041	167	0.14
1042	145	0.12
1043	170	0.15
1044	177	0.15
1045	220	0.19
1046	206	0.18
1047	210	0.18
1048	227	0.19
1049	264	0.23
1050	247	0.21
1051	289	0.25
1052	291	0.25
1053	292	0.25
1054	318	0.27
1055	293	0.25
1056	341	0.29
1057	356	0.31
1058	407	0.35
1059	388	0.33
1060	414	0.36
1061	449	0.39
1062	486	0.42
1063	502	0.43
1064	567	0.49
1065	596	0.51
1066	556	0.48
1067	639	0.55
1068	666	0.57
1069	697	0.60
1070	725	0.62
1071	704	0.60
1072	762	0.65
1073	823	0.71
1074	809	0.69
1075	913	0.78
1076	954	0.82
1077	927	0.79
1078	979	0.84
1079	1026	0.88
1080	1110	0.95
1081	1170	1.00
1082	1176	1.01
1083	1206	1.03
1084	1226	1.05
1085	1252	1.07

1086	1258	1.08
1087	1294	1.11
1088	1385	1.19
1089	1415	1.21
1090	1485	1.27
1091	1504	1.29
1092	1541	1.32
1093	1653	1.42
1094	1647	1.41
1095	1686	1.45
1096	1699	1.46
1097	1686	1.45
1098	1698	1.46
1099	1792	1.54
1100	1727	1.48
1101	1767	1.52
1102	1873	1.61
1103	1819	1.56
1104	1851	1.59
1105	1815	1.56
1106	1830	1.57
1107	1794	1.54
1108	1818	1.56
1109	1876	1.61
1110	1859	1.59
1111	1735	1.49
1112	1750	1.50
1113	1759	1.51
1114	1686	1.45
1115	1626	1.39
1116	1564	1.34
1117	1589	1.36
1118	1548	1.33
1119	1492	1.28
1120	1419	1.22
1121	1513	1.3
1122	1397	1.2
1123	1339	1.15
1124	1278	1.10
1125	1150	0.99
1126	1243	1.07
1127	1156	0.99
1128	1066	0.91
1129	1029	0.88
1130	926	0.79
1131	909	0.78
1132	877	0.75
1133	807	0.69

1134	784	0.67
1135	713	0.61
1136	706	0.61
1137	616	0.53
1138	554	0.48
1139	501	0.43
1140	519	0.45
1141	405	0.35
1142	453	0.39
1143	414	0.36
1144	409	0.35
1145	364	0.31
1146	287	0.25
1147	292	0.25
1148	264	0.23
1149	230	0.2
1150	196	0.17
1151	193	0.17
1152	138	0.12
1153	143	0.12
1154	133	0.11
1155	108	0.09
1156	128	0.11
1157	106	0.09
1158	84	0.07
1159	85	0.07
1160	69	0.06
1161	75	0.06
1162	48	0.04
1163	62	0.05
1164	40	0.03
1165	29	0.02
1166	26	0.02
1167	18	0.02
1168	33	0.03
1169	16	0.01
1170	37	0.03
1171	18	0.02
1172	11	0.01
1173	17	0.01
1174	14	0.01
1175	12	0.01
1176	5	0.00
1177	9	0.01
1178	14	0.01
1179	7	0.01
1180	4	0.00
1181	2	0.00

1182	1	0.00
1183	1	0.00
1184	3	0.00
1185	1	0.00
1186	2	0.00
1194	13	0.01
1195	4	0.00
1196	10	0.01
1197	5	0.00
1250	35	0.03

Table 15.3. MME Spring 2011 Reading Frequencies for Total Group -- All Forms Included

Scale Score	Frequency	Percent
950	544	0.46
951	1	0.00
952	2	0.00
953	3	0.00
954	7	0.01
955	6	0.01
956	5	0.00
957	4	0.00
958	4	0.00
959	6	0.01
960	9	0.01
961	6	0.01
962	6	0.01
963	6	0.01
964	4	0.00
965	4	0.00
966	5	0.00
967	3	0.00
968	3	0.00
969	3	0.00
970	7	0.01
971	3	0.00
972	5	0.00
973	3	0.00
974	7	0.01
975	3	0.00
976	7	0.01
977	10	0.01
978	8	0.01
979	13	0.01
980	5	0.00
981	5	0.00
982	7	0.01
983	16	0.01
984	24	0.02
985	28	0.02
986	30	0.03
987	20	0.02
988	24	0.02
989	26	0.02
990	29	0.02
991	26	0.02
992	22	0.02
993	23	0.02
994	29	0.02
995	31	0.03
996	29	0.02
997	23	0.02
998	36	0.03
999	22	0.02
1000	39	0.03
1001	27	0.02
1002	36	0.03
1003	37	0.03
1004	39	0.03
1005	27	0.02
1006	36	0.03
1007	33	0.03
1008	25	0.02
1009	27	0.02
1010	36	0.03
1011	35	0.03
1012	46	0.04
1013	50	0.04
1014	60	0.05
1015	46	0.04
1016	47	0.04
1017	72	0.06
1018	46	0.04
1019	72	0.06
1020	54	0.05
1021	55	0.05
1022	65	0.06
1023	66	0.06
1024	84	0.07
1025	68	0.06
1026	82	0.07
1027	82	0.07
1028	76	0.06
1029	78	0.07
1030	93	0.08
1031	87	0.07
1032	79	0.07
1033	102	0.09
1034	103	0.09
1035	109	0.09

1036	113	0.10
1037	109	0.09
1038	121	0.10
1039	110	0.09
1040	116	0.10
1041	124	0.11
1042	144	0.12
1043	161	0.14
1044	138	0.12
1045	148	0.13
1046	160	0.14
1047	129	0.11
1048	175	0.15
1049	169	0.14
1050	193	0.16
1051	210	0.18
1052	217	0.18
1053	248	0.21
1054	242	0.21
1055	230	0.20
1056	262	0.22
1057	288	0.25
1058	290	0.25
1059	280	0.24
1060	324	0.28
1061	325	0.28
1062	323	0.28
1063	378	0.32
1064	415	0.35
1065	446	0.38
1066	441	0.38
1067	454	0.39
1068	502	0.43
1069	540	0.46
1070	554	0.47
1071	568	0.48
1072	640	0.55
1073	629	0.54
1074	740	0.63
1075	676	0.58
1076	783	0.67
1077	769	0.66
1078	773	0.66
1079	901	0.77
1080	870	0.74
1081	907	0.77
1082	1000	0.85
1083	985	0.84

1084	1019	0.87
1085	1115	0.95
1086	1055	0.90
1087	1124	0.96
1088	1198	1.02
1089	1242	1.06
1090	1264	1.08
1091	1258	1.07
1092	1348	1.15
1093	1328	1.13
1094	1428	1.22
1095	1357	1.16
1096	1420	1.21
1097	1479	1.26
1098	1408	1.20
1099	1454	1.24
1100	1509	1.29
1101	1461	1.24
1102	1518	1.29
1103	1512	1.29
1104	1602	1.36
1105	1545	1.32
1106	1616	1.38
1107	1654	1.41
1108	1627	1.39
1109	1636	1.39
1110	1626	1.39
1111	1562	1.33
1112	1511	1.29
1113	1642	1.40
1114	1550	1.32
1115	1571	1.34
1116	1500	1.28
1117	1522	1.30
1118	1444	1.23
1119	1468	1.25
1120	1457	1.24
1121	1378	1.17
1122	1389	1.18
1123	1418	1.21
1124	1403	1.20
1125	1425	1.21
1126	1431	1.22
1127	1462	1.25
1128	1496	1.27
1129	1524	1.30
1130	1499	1.28
1131	1486	1.27

1132	1363	1.16
1133	1352	1.15
1134	1437	1.22
1135	1331	1.13
1136	1281	1.09
1137	1224	1.04
1138	1199	1.02
1139	1184	1.01
1140	1118	0.95
1141	1029	0.88
1142	897	0.76
1143	911	0.78
1144	810	0.69
1145	803	0.68
1146	735	0.63
1147	708	0.60
1148	586	0.50
1149	628	0.53
1150	515	0.44
1151	477	0.41
1152	494	0.42
1153	411	0.35
1154	408	0.35
1155	386	0.33
1156	309	0.26
1157	299	0.25
1158	289	0.25
1159	292	0.25
1160	267	0.23
1161	224	0.19
1162	189	0.16
1163	185	0.16
1164	188	0.16
1165	178	0.15
1166	151	0.13
1167	159	0.14

1168	110	0.09
1169	120	0.10
1170	101	0.09
1171	131	0.11
1172	79	0.07
1173	109	0.09
1174	94	0.08
1175	72	0.06
1176	86	0.07
1177	56	0.05
1178	56	0.05
1179	35	0.03
1180	45	0.04
1181	20	0.02
1182	57	0.05
1183	44	0.04
1184	54	0.05
1185	32	0.03
1186	62	0.05
1187	48	0.04
1188	22	0.02
1189	7	0.01
1190	20	0.02
1191	29	0.02
1192	14	0.01
1193	22	0.02
1194	24	0.02
1195	8	0.01
1196	3	0.00
1200	2	0.00
1201	19	0.02
1202	39	0.03
1204	1	0.00
1218	75	0.06
1250	142	0.12

Table 15.4. MME Spring 2011 Science Frequencies for Total Group—All Forms Included

Scale Score	Frequency	Percent
950	2855	2.44
951	2	0.00
952	1	0.00
953	2	0.00
956	1	0.00
957	4	0.00
958	2	0.00
959	1	0.00
960	1	0.00
961	1	0.00
962	2	0.00
963	2	0.00
964	2	0.00
965	1	0.00
966	2	0.00
967	2	0.00
968	3	0.00
969	4	0.00
970	2	0.00
971	2	0.00
972	3	0.00
973	2	0.00
974	4	0.00
975	8	0.01
976	2	0.00
977	8	0.01
978	5	0.00
979	5	0.00
980	7	0.01
981	8	0.01
982	7	0.01
983	7	0.01
984	10	0.01
985	9	0.01
986	10	0.01
987	16	0.01
988	8	0.01
989	9	0.01
990	10	0.01
991	7	0.01
992	10	0.01
993	12	0.01
994	6	0.01
995	18	0.02
996	11	0.01
997	14	0.01
998	16	0.01
999	19	0.02
1000	16	0.01
1001	21	0.02
1002	16	0.01
1003	15	0.01
1004	13	0.01
1005	18	0.02
1006	22	0.02
1007	25	0.02
1008	24	0.02
1009	30	0.03
1010	27	0.02
1011	29	0.02
1012	34	0.03
1013	25	0.02
1014	28	0.02
1015	32	0.03
1016	35	0.03
1017	33	0.03
1018	30	0.03
1019	38	0.03
1020	41	0.04
1021	36	0.03
1022	48	0.04
1023	57	0.05
1024	46	0.04
1025	57	0.05
1026	57	0.05
1027	64	0.05
1028	53	0.05
1029	73	0.06
1030	66	0.06
1031	80	0.07
1032	69	0.06
1033	83	0.07
1034	84	0.07
1035	84	0.07
1036	91	0.08
1037	99	0.08
1038	100	0.09
1039	104	0.09

1040	118	0.10
1041	119	0.10
1042	113	0.10
1043	124	0.11
1044	149	0.13
1045	158	0.13
1046	147	0.13
1047	173	0.15
1048	152	0.13
1049	188	0.16
1050	203	0.17
1051	228	0.19
1052	206	0.18
1053	220	0.19
1054	258	0.22
1055	282	0.24
1056	303	0.26
1057	333	0.28
1058	332	0.28
1059	330	0.28
1060	375	0.32
1061	394	0.34
1062	419	0.36
1063	411	0.35
1064	456	0.39
1065	457	0.39
1066	505	0.43
1067	529	0.45
1068	537	0.46
1069	611	0.52
1070	594	0.51
1071	668	0.57
1072	665	0.57
1073	739	0.63
1074	741	0.63
1075	770	0.66
1076	826	0.71
1077	836	0.71
1078	936	0.80
1079	920	0.79
1080	956	0.82
1081	1003	0.86
1082	996	0.85
1083	1045	0.89
1084	1072	0.92
1085	1163	0.99
1086	1112	0.95
1087	1175	1.00

1088	1228	1.05
1089	1183	1.01
1090	1267	1.08
1091	1232	1.05
1092	1409	1.20
1093	1347	1.15
1094	1382	1.18
1095	1432	1.22
1096	1432	1.22
1097	1465	1.25
1098	1452	1.24
1099	1541	1.32
1100	1487	1.27
1101	1537	1.31
1102	1608	1.37
1103	1604	1.37
1104	1564	1.34
1105	1645	1.41
1106	1607	1.37
1107	1564	1.34
1108	1594	1.36
1109	1580	1.35
1110	1679	1.43
1111	1772	1.51
1112	1642	1.40
1113	1567	1.34
1114	1606	1.37
1115	1584	1.35
1116	1632	1.39
1117	1599	1.37
1118	1603	1.37
1119	1578	1.35
1120	1552	1.33
1121	1503	1.28
1122	1451	1.24
1123	1512	1.29
1124	1463	1.25
1125	1439	1.23
1126	1416	1.21
1127	1374	1.17
1128	1349	1.15
1129	1291	1.10
1130	1248	1.07
1131	1205	1.03
1132	1181	1.01
1133	1197	1.02
1134	1108	0.95
1135	1103	0.94

1136	982	0.84
1137	996	0.85
1138	991	0.85
1139	966	0.83
1140	910	0.78
1141	845	0.72
1142	788	0.67
1143	771	0.66
1144	717	0.61
1145	692	0.59
1146	648	0.55
1147	590	0.50
1148	517	0.44
1149	559	0.48
1150	547	0.47
1151	409	0.35
1152	424	0.36
1153	371	0.32
1154	353	0.30
1155	344	0.29
1156	334	0.29
1157	263	0.22
1158	281	0.24
1159	239	0.20
1160	222	0.19
1161	214	0.18
1162	192	0.16
1163	154	0.13
1164	173	0.15
1165	129	0.11
1166	131	0.11
1167	116	0.10
1168	112	0.10
1169	101	0.09
1170	82	0.07
1171	81	0.07
1172	80	0.07
1173	51	0.04
1174	80	0.07
1175	61	0.05
1176	51	0.04
1177	46	0.04
1178	47	0.04
1179	40	0.03
1180	32	0.03
1181	35	0.03
1182	27	0.02
1183	30	0.03

1184	30	0.03
1185	9	0.01
1186	18	0.02
1187	14	0.01
1188	19	0.02
1189	15	0.01
1190	16	0.01
1191	14	0.01
1192	13	0.01
1193	8	0.01
1194	8	0.01
1195	10	0.01
1196	10	0.01
1197	5	0.00
1198	6	0.01
1199	4	0.00
1200	9	0.01
1201	7	0.01
1202	8	0.01
1203	8	0.01
1204	1	0.00
1205	2	0.00
1206	2	0.00
1207	1	0.00
1208	6	0.01
1209	1	0.00
1210	2	0.00
1211	3	0.00
1212	13	0.01
1213	2	0.00
1219	1	0.00
1220	1	0.00
1222	3	0.00
1250	47	0.04

Table 15.5. MME Spring 2011 Social Studies Frequencies for Total Group—All Forms Included

Scale Score	Frequency	Percent
979	2	0.00
1011	1	0.00
1031	5	0.00
1033	2	0.00
1041	1	0.00
1043	20	0.02
1044	2	0.00
1045	16	0.01
1049	2	0.00
1050	5	0.00
1052	65	0.06
1053	42	0.04
1056	7	0.01
1057	14	0.01
1059	159	0.14
1060	73	0.06
1063	27	0.02
1065	346	0.3
1066	188	0.16
1068	55	0.05
1070	662	0.57
1071	329	0.28
1073	99	0.08
1075	1137	0.97
1076	421	0.36
1077	66	0.06
1078	64	0.05
1080	2298	1.96
1082	146	0.12
1084	3019	2.58
1085	93	0.08
1086	111	0.09
1088	3704	3.16
1089	183	0.16
1090	2	0.00
1091	4054	3.46
1093	230	0.2
1095	4253	3.63
1096	95	0.08
1097	120	0.10
1098	4463	3.81
1100	213	0.18
1101	4560	3.89
1103	197	0.17
1104	481	0.41
1105	3966	3.39
1106	110	0.09
1107	481	0.41
1108	4043	3.45
1110	582	0.50
1111	4228	3.61
1113	536	0.46
1114	4155	3.55
1116	431	0.37
1117	4287	3.66
1119	162	0.14
1120	4520	3.86
1123	4688	4.00
1126	335	0.29
1127	4352	3.71
1128	157	0.13
1129	74	0.06
1130	4631	3.95
1132	154	0.13
1133	4683	4.00
1135	133	0.11
1137	4689	4.00
1139	126	0.11
1140	4701	4.01
1143	137	0.12
1144	4679	3.99
1147	99	0.08
1148	4594	3.92
1151	83	0.07
1153	4444	3.79
1156	68	0.06
1157	3976	3.39
1158	63	0.05
1161	72	0.06
1163	3418	2.92
1166	45	0.04
1169	2815	2.40
1173	41	0.03
1176	2036	1.74
1177	17	0.01
1182	30	0.03
1184	1286	1.10
1185	15	0.01

1186	10	0.01
1194	20	0.02
1196	666	0.57
1197	7	0.01
1198	6	0.01
1213	8	0.01
1216	208	0.18
1217	4	0.00
1248	49	0.04

Table 15.6. MME Spring 2011 Writing Frequencies for Total Group—All Forms Included

Scale Score	Frequency	Percent
950	2574	2.18
951	7	0.01
952	4	0.00
953	6	0.01
954	9	0.01
955	4	0.00
956	7	0.01
957	6	0.01
958	7	0.01
959	5	0.00
960	5	0.00
961	8	0.01
962	5	0.00
963	3	0.00
964	7	0.01
965	7	0.01
966	7	0.01
967	6	0.01
968	5	0.00
969	11	0.01
970	7	0.01
971	7	0.01
972	8	0.01
973	8	0.01
974	8	0.01
975	7	0.01
976	10	0.01
977	7	0.01
978	9	0.01
979	14	0.01
980	7	0.01
981	15	0.01
982	8	0.01
983	13	0.01
984	5	0.00
985	11	0.01
986	10	0.01
987	9	0.01
988	14	0.01
989	12	0.01
990	15	0.01
991	17	0.01
992	23	0.02
993	17	0.01
994	21	0.02
995	18	0.02
996	18	0.02
997	26	0.02
998	37	0.03
999	23	0.02
1000	23	0.02
1001	34	0.03
1002	38	0.03
1003	20	0.02
1004	26	0.02
1005	33	0.03
1006	44	0.04
1007	28	0.02
1008	37	0.03
1009	45	0.04
1010	36	0.03
1011	50	0.04
1012	44	0.04
1013	46	0.04
1014	44	0.04
1015	44	0.04
1016	63	0.05
1017	82	0.07
1018	72	0.06
1019	69	0.06
1020	76	0.06
1021	66	0.06
1022	93	0.08
1023	96	0.08
1024	97	0.08
1025	90	0.08
1026	101	0.09
1027	113	0.10
1028	115	0.10
1029	135	0.11
1030	132	0.11
1031	138	0.12
1032	153	0.13
1033	147	0.12
1034	148	0.13
1035	166	0.14
1036	165	0.14
1037	183	0.15

1038	217	0.18
1039	218	0.18
1040	225	0.19
1041	257	0.22
1042	237	0.20
1043	255	0.22
1044	267	0.23
1045	301	0.25
1046	338	0.29
1047	313	0.26
1048	345	0.29
1049	376	0.32
1050	359	0.30
1051	427	0.36
1052	439	0.37
1053	400	0.34
1054	417	0.35
1055	491	0.42
1056	539	0.46
1057	497	0.42
1058	548	0.46
1059	598	0.51
1060	597	0.50
1061	611	0.52
1062	654	0.55
1063	704	0.60
1064	759	0.64
1065	750	0.63
1066	772	0.65
1067	848	0.72
1068	841	0.71
1069	865	0.73
1070	934	0.79
1071	909	0.77
1072	1038	0.88
1073	1005	0.85
1074	1122	0.95
1075	1083	0.92
1076	1119	0.95
1077	1184	1.00
1078	1236	1.05
1079	1248	1.06
1080	1230	1.04
1081	1341	1.13
1082	1360	1.15
1083	1319	1.12
1084	1420	1.20
1085	1525	1.29

1086	1408	1.19
1087	1452	1.23
1088	1509	1.28
1089	1539	1.30
1090	1660	1.40
1091	1575	1.33
1092	1560	1.32
1093	1566	1.32
1094	1623	1.37
1095	1670	1.41
1096	1591	1.35
1097	1661	1.40
1098	1678	1.42
1099	1669	1.41
1100	1639	1.39
1101	1612	1.36
1102	1668	1.41
1103	1615	1.37
1104	1643	1.39
1105	1558	1.32
1106	1543	1.30
1107	1550	1.31
1108	1540	1.30
1109	1536	1.30
1110	1467	1.24
1111	1482	1.25
1112	1399	1.18
1113	1384	1.17
1114	1395	1.18
1115	1375	1.16
1116	1307	1.11
1117	1357	1.15
1118	1240	1.05
1119	1259	1.06
1120	1235	1.04
1121	1122	0.95
1122	1081	0.91
1123	1091	0.92
1124	1000	0.85
1125	1055	0.89
1126	978	0.83
1127	910	0.77
1128	847	0.72
1129	864	0.73
1130	768	0.65
1131	794	0.67
1132	754	0.64
1133	729	0.62

1134	693	0.59
1135	710	0.60
1136	614	0.52
1137	655	0.55
1138	586	0.50
1139	539	0.46
1140	541	0.46
1141	503	0.43
1142	463	0.39
1143	466	0.39
1144	410	0.35
1145	385	0.33
1146	395	0.33
1147	382	0.32
1148	349	0.30
1149	295	0.25
1150	303	0.26
1151	268	0.23
1152	263	0.22
1153	242	0.20
1154	207	0.18
1155	219	0.19
1156	194	0.16
1157	204	0.17
1158	185	0.16
1159	204	0.17
1160	164	0.14
1161	136	0.11
1162	147	0.12
1163	106	0.09
1164	130	0.11
1165	119	0.10
1166	92	0.08
1167	58	0.05
1168	98	0.08
1169	54	0.05
1170	134	0.11
1171	92	0.08
1172	48	0.04
1173	77	0.07
1174	44	0.04
1175	51	0.04
1176	44	0.04
1177	11	0.01
1178	62	0.05

1179	41	0.03
1180	33	0.03
1181	20	0.02
1182	13	0.01
1183	12	0.01
1184	14	0.01
1185	22	0.02
1186	180	0.15
1187	7	0.01
1188	7	0.01
1189	13	0.01
1190	28	0.02
1191	5	0.00
1192	3	0.00
1193	13	0.01
1195	4	0.00
1196	11	0.01
1197	2	0.00
1198	7	0.01
1199	2	0.00
1200	3	0.00
1201	64	0.05
1202	7	0.01
1204	20	0.02
1205	1	0.00
1206	1	0.00
1207	3	0.00
1208	1	0.00
1209	1	0.00
1210	1	0.00
1211	1	0.00
1214	1	0.00
1217	1	0.00
1218	1	0.00
1219	1	0.00
1220	1	0.00
1223	65	0.05
1224	1	0.00
1228	2	0.00
1233	1	0.00
1245	2	0.00
1249	1	0.00
1250	23	0.02

Table 15.7. Mean and SD of Item Parameter Estimates for Mathematics

	2011 Spring Mathematics		
	a	b	c
	Initial Form 1		
mean	1.537	0.221	0.205
SD	0.473	0.615	0.076
	Initial Form 2		
mean	1.526	0.244	0.205
SD	0.5	0.616	0.078
	Initial Form 3		
mean	1.521	0.263	0.201
SD	0.477	0.61	0.077
	Initial Form 4		
mean	1.528	0.227	0.206
SD	0.482	0.611	0.082
	Initial Form 5		
mean	1.535	0.237	0.207
SD	0.472	0.627	0.076
	Initial Form 6		
mean	1.524	0.248	0.205
SD	0.485	0.605	0.079
	Initial Form 7		
mean	1.539	0.253	0.201
SD	0.471	0.606	0.077
	Initial Form 8		
mean	1.534	0.248	0.206
SD	0.466	0.605	0.078
	Initial Form 9		
mean	1.533	0.229	0.207
SD	0.469	0.636	0.077
	Initial Form 10		
mean	1.538	0.256	0.196
SD	0.468	0.611	0.075
	Makeup Form		
mean	1.348	0.028	0.191
SD	0.433	0.827	0.077
	Accommodated Form		
mean	1.544	0.044	0.203
SD	0.519	0.713	0.072

Table 15.8. Mean and SD of Item Parameter Estimates for Reading

	2011 Spring Reading		
	a	b	c
	Initial Form		
mean	1.017	0.058	0.199
SD	0.507	1.261	0.07
	Makeup Form		
mean	0.947	-0.013	0.199
SD	0.458	1.341	0.077
	Accommodated Form		
mean	0.959	0.011	0.19
SD	0.422	1.389	0.079

Table 15.9. Mean and SD of Item Parameter Estimates for Science

	2011 Spring Science		
	a	b	c
	Initial Form 1		
mean	0.931	0.612	0.222
SD	0.310	0.928	0.090
	Initial Form 2		
mean	0.879	0.682	0.209
SD	0.289	0.895	0.085
	Initial Form 3		
mean	0.905	0.468	0.206
SD	0.280	0.896	0.084
	Initial Form 4		
mean	0.940	0.711	0.223
SD	0.278	0.845	0.094
	Initial Form 5		
mean	0.906	0.575	0.215
SD	0.281	0.837	0.084
	Initial Form 6		
mean	0.881	0.556	0.214
SD	0.290	0.775	0.084
	Initial Form 7		
mean	0.941	0.650	0.218
SD	0.280	0.766	0.086
	Initial Form 8		
mean	0.909	0.605	0.217
SD	0.283	0.879	0.092
	Initial Form 9		
mean	0.913	0.516	0.220
SD	0.286	0.873	0.089
	Initial Form 10		
mean	0.921	0.683	0.217
SD	0.275	0.840	0.088
	Makeup Form		
mean	0.85	0.685	0.224
SD	0.279	1.082	0.072
	Accommodated Form		
mean	0.876	0.619	0.226
SD	0.297	1.034	0.081

Table 15.10. Mean and SD of Item Parameter Estimates for Writing

	2011 Spring Writing MC Items		
	a	b	c
	Initial Form		
mean	0.827	0.361	0.221
SD	0.227	0.864	0.071
	Makeup Form		
mean	0.855	0.467	0.217
SD	0.262	0.693	0.089
	Accommodated Form		
mean	0.895	0.339	0.201
SD	0.285	0.71	0.065

	2011 Spring Writing CR Items						
	a	b	Step1	Step2	Step3	Step4	Step5
	Initial Form						
mean	0.483	1.002	3.337	3.032	0.858	-2.395	-4.832
SD	0.004	0.014	0.002	0.012	0.01	0.015	0.039
	Makeup Form						
mean	0.494	1.14	3.35	2.56	0.778	-2.441	-4.247
SD	0.009	0.033	0.014	0.08	0.006	0.061	0.149
	Accommodated Form						
mean	0.398	1.425	3.208	2.659	0.925	-2.336	-4.456
SD	0.005	0.246	0.23	0.275	0.238	0.358	1.101

Table 15.11. Mean and SD of Item Parameter Estimates for Social Studies

	2011 Spring Social Studies
	b
	Initial Form
mean	0.1191
SD	0.7373
	Makeup Form
mean	0.0915
SD	0.8375
	Accommodated Form
mean	0.1036
SD	0.5858

Chapter 16: MME Scale Score History

The first MME assessment was administered statewide in spring 2007. For each content area, Tables 16.1 to 16.5 present the average scores and the percentages of students in each of the four performance levels. The MME proficiency level cut scores by subjects are reported in table 13.1.Yearly samples from a state may vary, and changes to a testing program (such as the inclusion of a new measurement instrument like *Locating Information* in 2009) contribute to annual scale score means. Therefore, changes in means across years need to be considered in context. The BAA encourages those interested in using MME scale scores to reference the informational materials contained in this technical report, as well as information posted on the MME website (www.michigan.gov/mme). Additionally, the Guide to Reports should be used when using or referencing assessment data (see http://www.michigan.gov/documents/mde/2011_MME_GTR_Final_355746_7.pdf).

Table 16.1. Spring 2007 Michigan State Average Scores and Percentages in each Performance Level

			Percentages within Performance Levels			
Content Area	**N**	**Average**	**Apprentice**	**Basic**	**Met Standards**	**Exceeded Standards**
Reading	113,956	1104	17%	24%	58%	2%
Writing	111,479	1090	10%	50%	38%	2%
ELA	111,000	1098	12%	37%	49%	2%
Mathematics	113,839	1093	38%	16%	37%	10%
Science	113,630	1098	28%	16%	50%	6%
Social Studies	113,718	1124	7%	9%	42%	41%

Table 16.2. 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**	**Advanced**
Reading	130,226	1106	17%	21%	60%	3%
Writing	129,400	1090	11%	48%	39%	3%
ELA	128,818	1099	13%	34%	50%	2%
Mathematics	129,803	1093	38%	16%	36%	10%
Science	129,691	1099	27%	16%	51%	6%
Social Studies	130,957	1123	7%	13%	39%	41%

Table 16.3. Spring 2009 Michigan State Average Scores and Percentages in each Performance Level

			Percentages within Performance Levels			
Content Area	**N**	**Average**	**Not Proficient**	**Partially Proficient**	**Proficient**	**Advanced**
Reading	124,385	1106	16%	23%	58%	3%
Writing	125,579	1091	10%	45%	40%	4%
ELA	124,099	1099	12%	35%	50%	3%
Mathematics	123,284	1095	35%	15%	37%	13%
Science	123,873	1099	29%	15%	48%	8%
Social Studies	123,969	1127	8%	11%	39%	43%

Table 16.4. Spring 2010 Michigan State Average Scores and Percentages in each Performance Level

			Percentages within Performance Levels			
Content Area	**N**	**Average**	**Not Proficient**	**Partially Proficient**	**Proficient**	**Advanced**
Reading	120,488	1108	14%	21%	62%	3%
Writing	121,181	1093	8%	47%	41%	4%
Mathematics	119,672	1094	33%	16%	39%	12%
Science	120,084	1101	27%	15%	49%	9%
Social Studies	120,233	1123	7%	14%	42%	38%

Table 16.5. Spring 2011 Michigan State Average Scores and Percentages in each Performance Level

			Percentages within Performance Levels			
Content Area	**N**	**Average**	**Not Proficient**	**Partially Proficient**	**Proficient**	**Advanced**
Reading	112,918	1108	14%	22%	61%	3%
Writing	113,615	1095	8%	44%	43%	5%
Mathematics	112,250	1095	32%	15%	40%	13%
Science	112,618	1104	24%	15%	52%	9%
Social Studies	112,714	1123	8%	14%	36%	42%

It is worth mentioning that beginning from Spring 2010 administration, total ELA was not reported and only separate MME Reading and MME Writing scale scores were reported, thus leaving five subjects in the "Content Area" column in Tables 16.4 and 16.5 above as compared with six subjects for the previous years.

References

ACT. (2007a). *ACT National Curriculum Survey 2005-2006.* Iowa City, IA: Author.

ACT. (2009). *ACT Writing Test Technical Report*. Iowa City, IA: Author.

ACT. (2007b). *The ACT technical manual*. Iowa City, IA: Author.

ACT. (2008a). *WorkKeys Applied Mathematics technical manual*. Iowa City, IA: Author.

ACT. (2008b). *WorkKeys Locating Information technical manual*. Iowa City, IA: Author.

ACT. (2008c). *WorkKeys Reading for Information technical manual*. Iowa City, IA: Author.

ACT (2010). *2009–2010 Fairness Report for the ACT Tests.* Iowa City, IA: Author

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

Anastasi, A. (1982). *Psychological testing (5th ed.).* New York: Macmillan.

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.

de Ayala, R.J. (2009). *The theory and practice of item response theory*. NY: Guilford Press.

Dorans, N.J., & Schmitt, A. P. (1991). Constructed response and differential item functioning: A pragmatic approach (Research Rep. No. 91-47). Princeton, NJ: Educational Testing Service.

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

Drasgow, F., Levine, M. V., & Williams, E. A. (1985). Appropriateness measurement with polytomous item response models and standardized indices. *British Journal of Mathematical and Statistical Psychology*, 38, 67-68.

Hambleton, R.K., Merenda, P. F., & Spielberger, C. D. (2005). Adapting educational and psychological tests for cross-cultural assessments. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Hambleton, R.K., & Swaminathan, H. (1985). *Item response theory: Principles and Applications.* Boston: Kluwer.

Holland, P.W., & Wainer, H. (1993). *Differential item functioning*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Joint Committee on Testing Practices. (2004). *Code of fair testing practices in education*. Washington, DC: Author.

Kolen, M.J. (1984). Effectiveness of analytic smoothing in equipercentile equating. *Journal of Educational Statistics*, 9, 25 – 44.

Kolen, M. J., & Brennan, R. L. (2004). *Test equating, scaling, and linking: Methods and practices*, New York: Springer.

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.

Li, M. F., & Olejnik, S. (1997). The power of Rasch person-fit statistics in detecting unusual response patterns. *Applied Psychological Measurement*, 21, 215-231.

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/.

Millman, J., & Greene, U. (1989). *The specification and development of tests of achievement and ability.* In R.L. Linn (Ed.) Educational Measurement (3rd edition, pp. 335-366). New York: American Council on Education and Macmillan.

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

Muraki, E. (1996). A generalized partial credit model. In Van der Linden, W.J. & Hambleton, R.K. (Eds). *Handbook of modern item response theory* (pp. 153 – 168). New York: Springer.

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

Nering, M. L., & Meijer, M. D. (1998). A comparison of the person response function and the lz person fit statistic. *Applied Psychological Measurement*, 22, 53-69.

Office of Educational Assessment and Accountability, Michigan Department of Education. (2009). Assessment integrity guide—Michigan standards for professional and ethnical conduct in assessment administration and reporting. Retrieved October 13, 2010, from http://www.michigan.gov/documents/mde/Assessment_Integrity_Guide_291950_7.pdf

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. (2005). *Review of ACT Coverage of Michigan Language Arts Standards.*

Spray, J. A.(1989). *Performance of Three Conditional DIF Statistics in Detecting Differential Item Functioning on Simulated Tests*. ACT Research Report No.89-7, Iowa City, IA.

Stocking, M. L. & Lord, F. M. (1983). Developing a common metric in item response theory. *Applied Psychological Measurement, 7*, 201-210.

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*.

Webb, N.L.(2010). *Alignment Analysis of Mathematics Standards and Two Forms of ACT, WorkKeys, and Augmented Assessments (Forms M1 and M2), One Form of Augmented Assessment with Sixteen Common and Unique Items (Form 1011), and Eight Forms of Augmented Assessment with Six Unique Items (Forms 1003-1010)*. An Independent Review Report Prepared for the Office of Educational Assessment and Accountability, Michigan Department of Education. Lansing, MI.

Webb, N.L. (2010). *Alignment Analysis of English Language Arts Standards and Two Forms of ACT/WorkKeys Reading Assessment and Two Forms of ACT English Assessment*. An Independent Review Report Prepared for the Office of Educational Assessment and Accountability, Michigan Department of Education. Lansing, MI.

Webb, N.L. (2010). *Alignment Analysis of Science Standards and Two Assessment Forms of ACT and Michigan Items*. An Independent Review Report Prepared for the Office of Educational Assessment and Accountability, Michigan Department of Education. Lansing, MI.

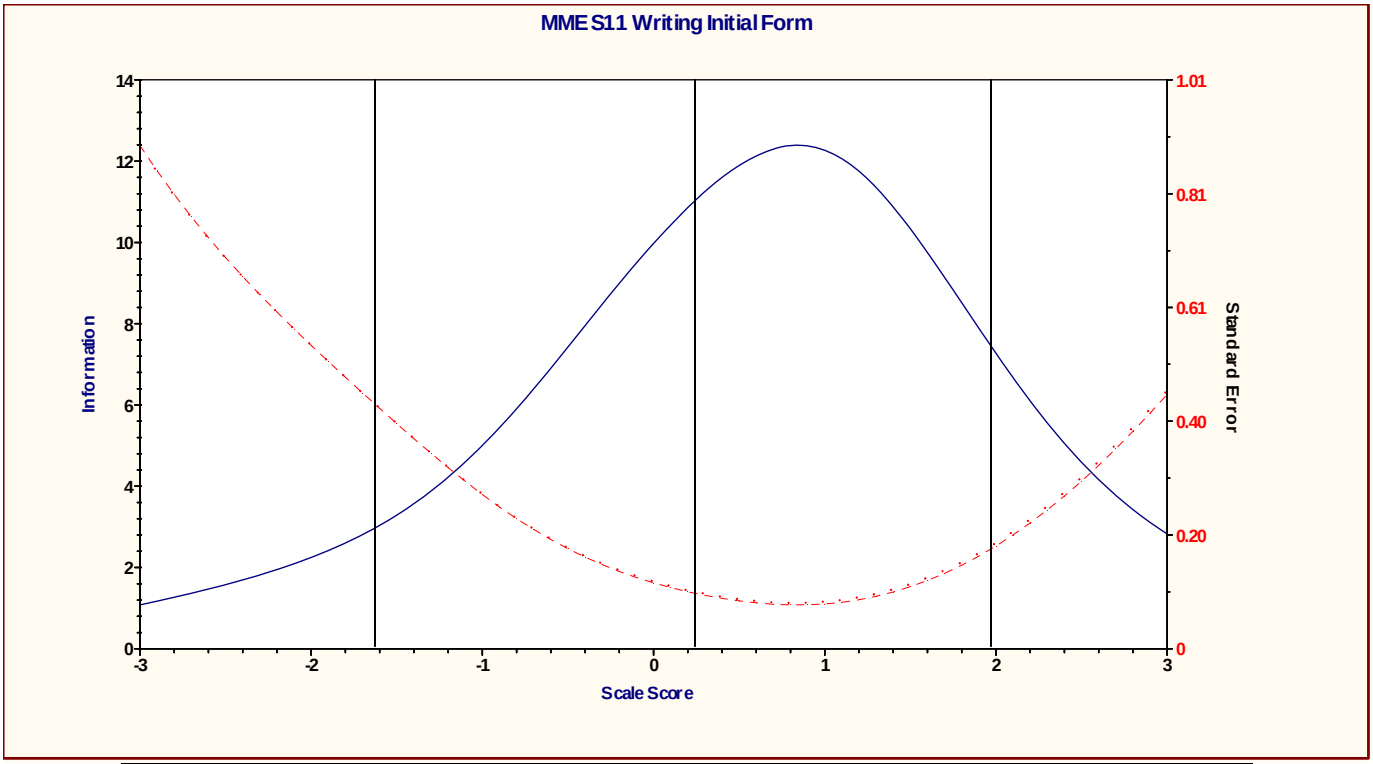
White, E.M. (1994). *Teaching and assessing writing: Recent advances in understanding, evaluating, and improving student performance* (2nd ed.). Portland, Maine: Calendar Islands.

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 MME Test Information Functions

Spring 2011 Writing Initial Form

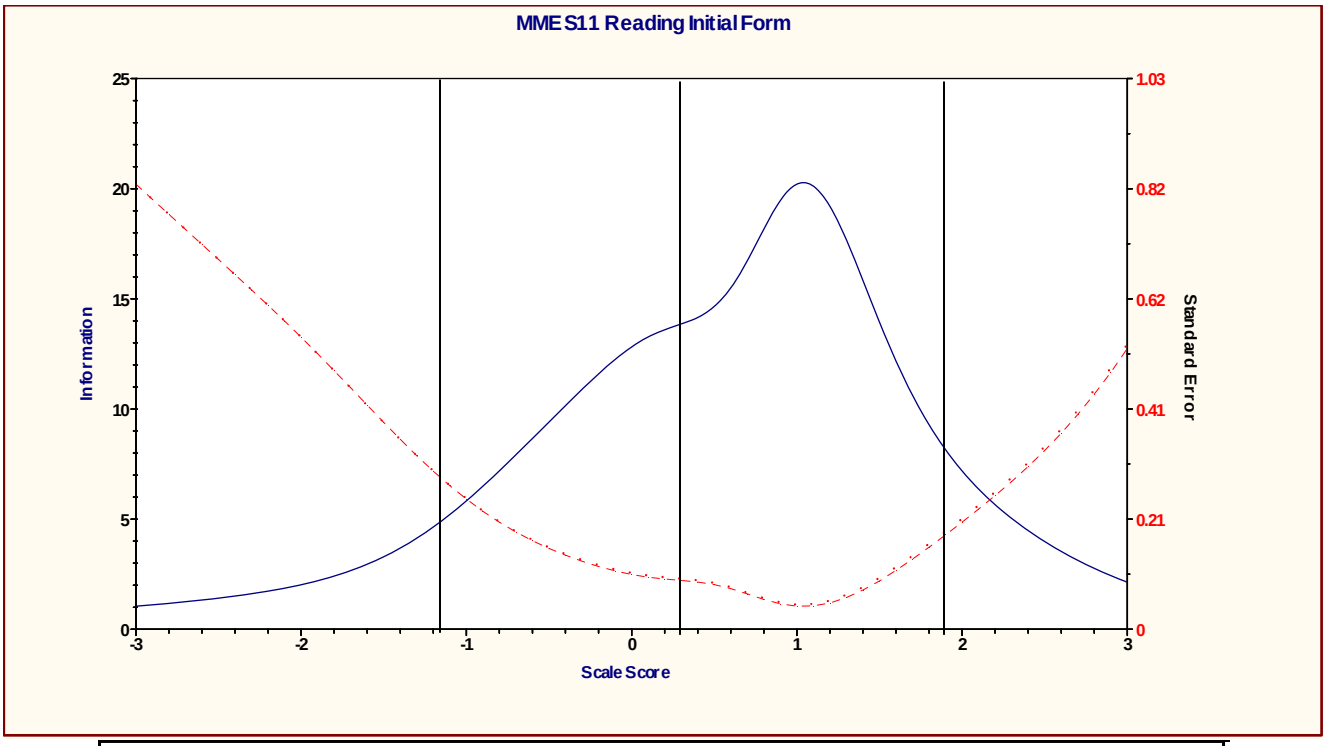


Test information curve: solid line **Standard error curve: dotted line**

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 2011 Reading Initial Form

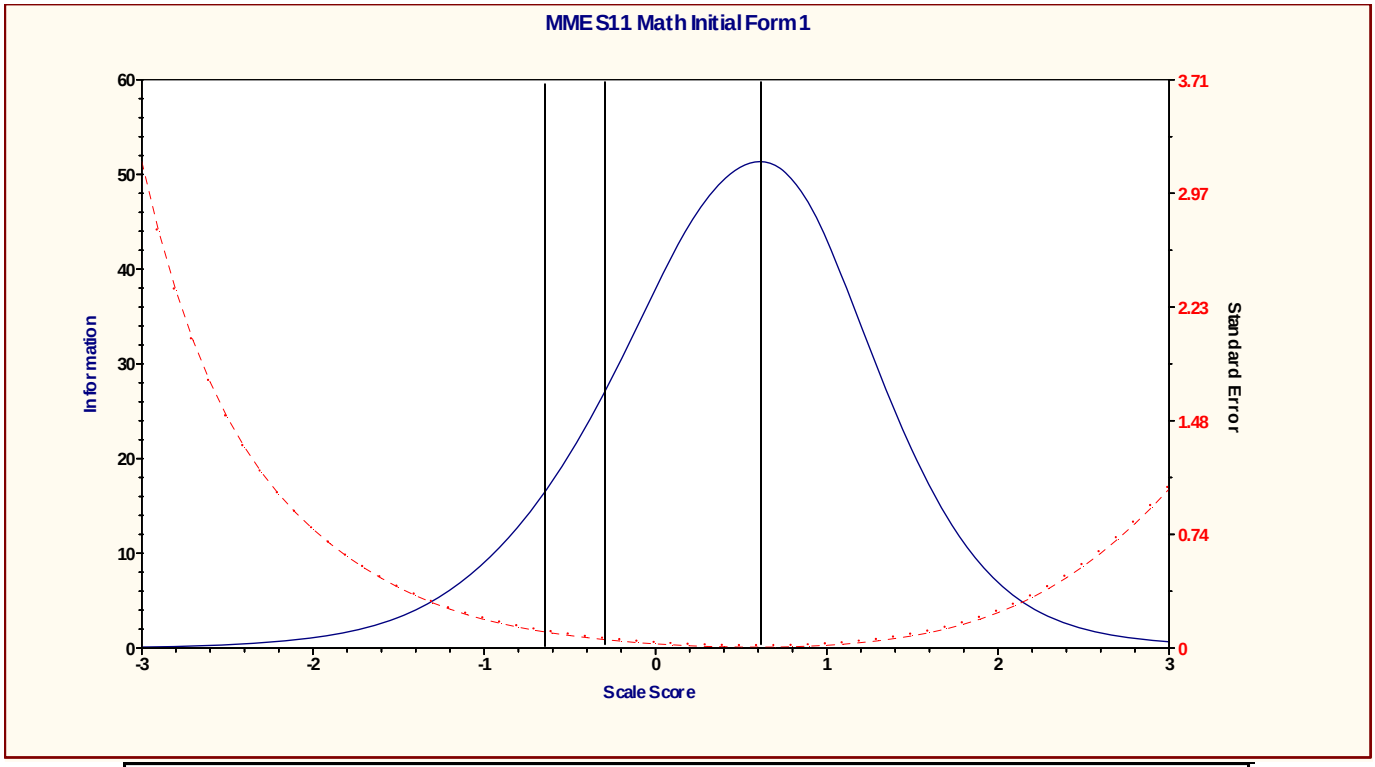


Test information curve: solid line **Standard error curve: dotted line**

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 2011 Mathematics Initial Form 1101

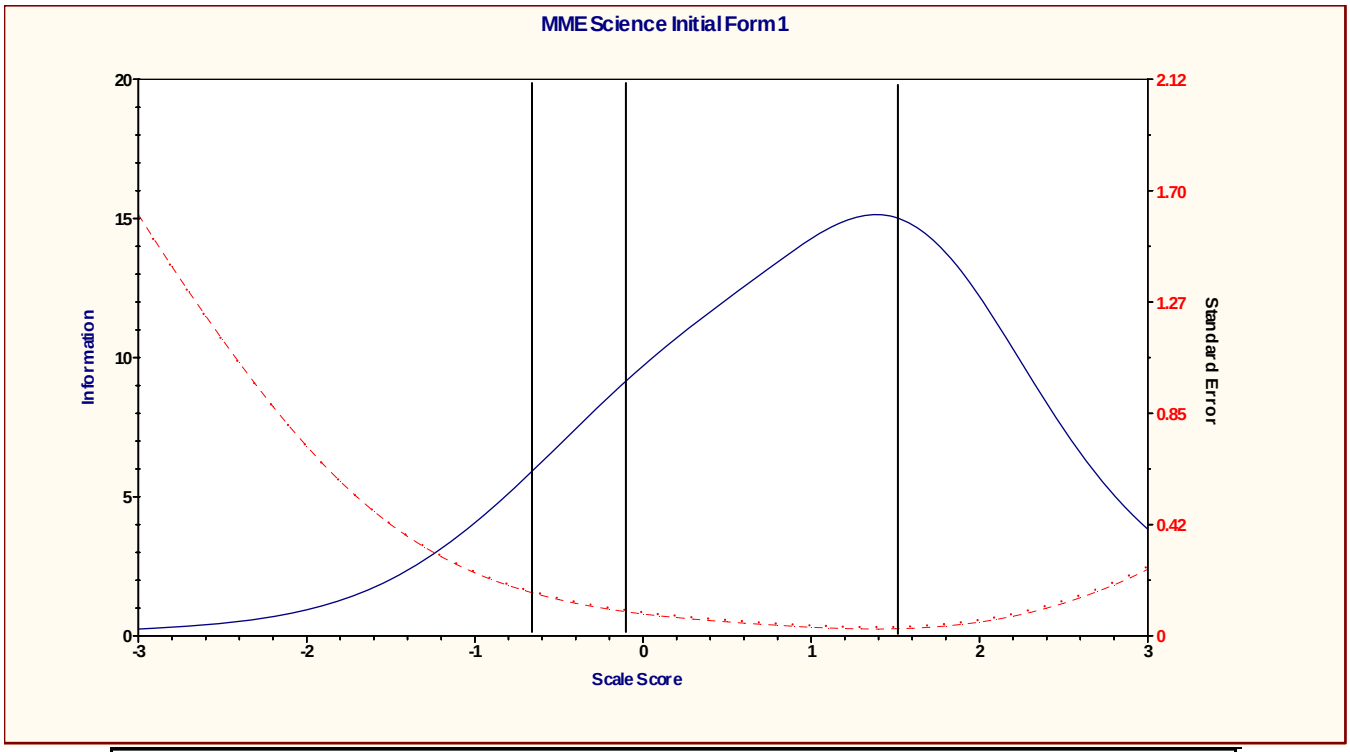


Test information curve: solid line **Standard error curve: dotted line**

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 2011 Science Initial Form 1101

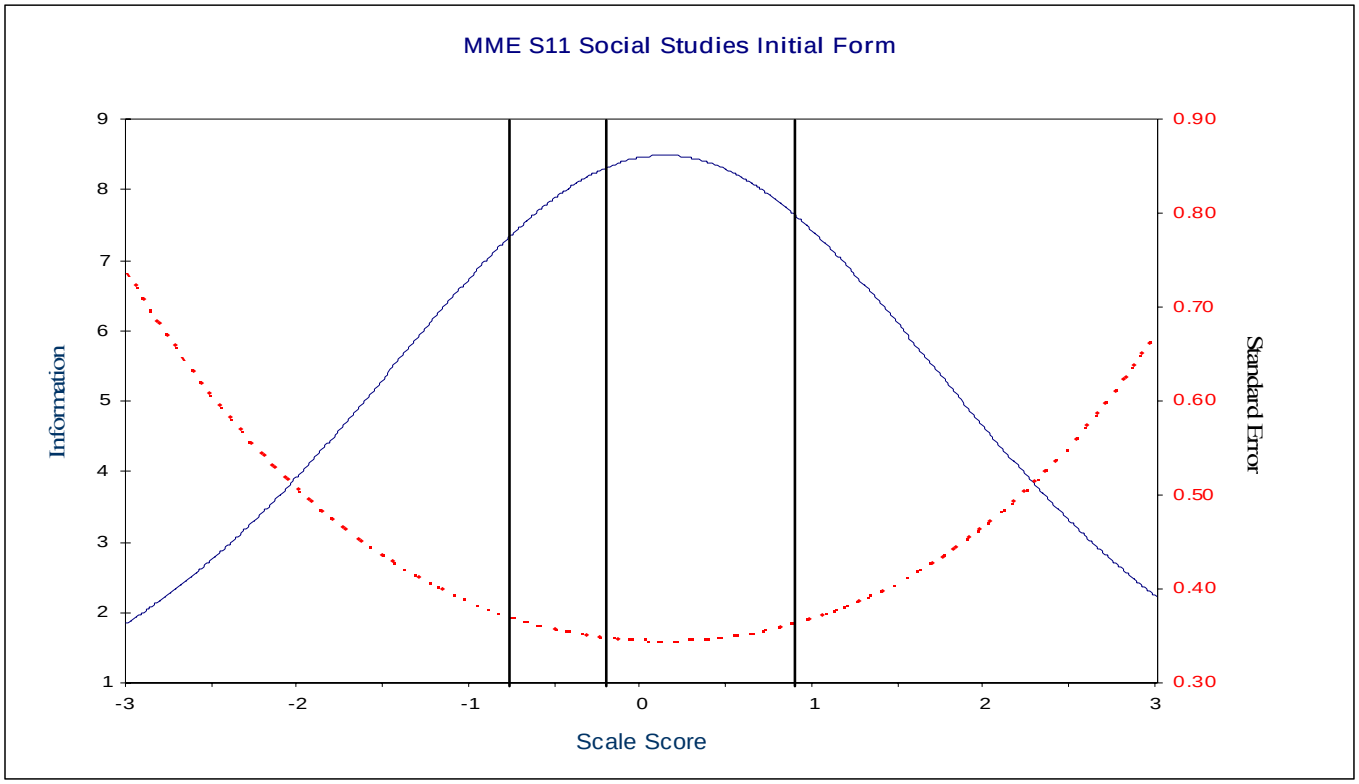


Test information curve: solid line **Standard error curve: dotted line**

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 2011 Social Studies 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 Description	Sample Values	Source	Mandatory (item bank use)	Ignore When pulling from Vendor data (Y/N)? (Item Bank use)	In MME Data	Note
Unique IBS defined ID for the Statistic row			Y	Y		
Unique IBS defined ID for the Item			Y	Y		
Unique IBS defined ID for the program			Y	Y		
Unique IBS defined ID for the Cycle			Y	Y		
BAA provided Cycle ID from Master Data	TestcycleID	Test map	Y		Y	
Unique IBS defined ID for the ContentArea			Y	Y		
Unique IBS defined ID for the Grade			Y	Y		
Unique IBS defined ID for the Administration Type			Y	Y		
Unique IBS defined ID for the blueprint			Y	Y		
Name of the administration cycle	Possible values: Fall 2010, Fall 2011, Spring 2012		Y	Y		
Administration Year	Possible values: 2010, 2011, 2012		Y	Y		
Indicator used in the Statistics upload into BAA and then into IBS tables	Possible values: Test map created, Data uploaded from vendor, Data Pulled by IBS		Y	Y		
MEAP Item ID (13 characters)		Test map	Y		Y	

Legacy CID/MEAP ID (both 7 and 12 digits used) from vendor system	CID/Itemcode	Test map	Y		Y	
Program Shortname/Long name in the IBS	MME		Y	Y	Y	
Subject (RE, WR, LI, MA, SC, SS)	**For MME: MA, SC, SS.** Possible values: Mathematics,Reading,Writing,Science,Social Studies,Content Area, English Language Arts,Science Listening, Speaking, Algebra I, Algebra II, Probability & Statistics, Pre-Calculus, English 9, English 10, English 11, English 12,Biology,Chemistry,Physics,Earth Science, US History + Geography, World History + Geography, Civics, Economics				Y	
Grade-Level of GLCE		Test map			Y	
Grade level the item will be tested in	11	Test map	Y	Y	Y	
Order for item processing	ACT variable that contains the item processing order within each subject				Y	
Name of the Blueprint			Y	Y		
Name of the Administration Type	Possible values: Standard (Initial), Emergency, Accommodated, Braille, Makeup			Y	Y	
Item Type (MC, CR, SR, AB)	for MME Day 3: MC only		Y	Y	Y	

Scenario Title/Passage type		Test map			Y	
Section of the test		Test map			Y	
Item Answer Key	Mandatory when item is MC, Possible values: A, B, C, D	Test map	Y	Y	Y	
Item Maximal Score points						
Unique IBS defined ID for the Primary Content standard of the item			Y	Y		
Primary Content Standard of the Item	Possible values: D.RE.02.01		Y	Y		
Level 1 of the Content Standard	Possible values: D		Y	Y		
Level 2 of the Content Standard	Possible values: D.RE		Y	Y		
Level 3 of the Content Standard	Possible values: D.RE.02.01		Y	Y		
If applicable						
Depth of Knowledge, EDOK, LDL			Y	Y		
Item reported standard		Test map			Y	
item reported domain		Test map			Y	
item reported benchmark		Test map			Y	
High School Content Expectation		Test map			Y	
Calculator (Y) or None		Test map			Y	

	OP, FT for MME. Possible values: Assigned,Initial Development,Submitted,Accepted,Banked,Ready for Pilot Test,Selected for Pilot Test,Pilot Tested,Ready for Field Review,Selected for Field Review,Field Reviewed,Ready for Field Test,Selected for Field Test,Field Tested,Selected for Operational,Ready for Operational,Used Operationally, Retired,Released,Do Not Use	Test map	Y	Y	Y	
Item Function in Current Administration	Possible Values: Equating, Common, Field Test and Matrix		Y	Y	Y	
Form the item appears on	For matrix items the first form	Test map	Y	Y	Y	
Number of Forms item appeared	Possible Values: 1,2,3,4,5,6,7,8,9,10	Test map	Y	Y	Y	
Forms in which the item appeared	Possible Values: 1101,1102,1110	Test map	Y	Y	Y	
Item position on the test form	for matrix items the first form	Test map			Y	
Test Positions in the forms used	Possible Values: 21,23,29,36,27	Test map	Y	Y	Y	
anchor item (0=No, 1=Yes); 1 Means "Use for pre-equating"	0, 1	Test map			Y	
	0.1	Test map			Y	
	Mandatory when item is CR, SR or AB		Y	Y		

	Mandatory when item is CR, SR or AB		Y	Y		
N-count		Total number of calibration students	Y	Y	Y	
Ncount Males		N-counts for bread-down groups	Y	Y	Y	
Ncount Females		N-counts for bread-down groups	Y	Y	Y	
Ncount White		N-counts for bread-down groups	Y	Y	Y	
Ncount Black		N-counts for bread-down groups	Y	Y	Y	
Ncount Hispanic		N-counts for bread-down groups	Y	Y		
Ncount Asian/Pacific Island		N-counts for bread-down groups	Y	Y		
Ncount Accommodated		N-counts for bread-down groups	Y	Y	Y	in a separate file
Ncount Non-accommodated		N-counts for bread-down groups	Y	Y		
Ncount Economically Disadvantaged		N-counts for bread-down groups	Y	Y		
Ncount Non-Economically		N-counts	Y	Y		

Disadvantaged		for bread-down groups				
Ncount English Language Learner		N-counts for bread-down groups	Y	Y		
Ncount Non-English Language Learner		N-counts for bread-down groups	Y	Y		
Ncount Disabled Learner						
Ncount NonDisabled Learner						
Percent for Comment Code 0			Y			
Percent for Comment Code 1			Y			
Percent for Comment Code 2			Y			
Percent for Comment Code 3			Y			
Percent for Comment Code 4			Y			
Percent for Comment Code 5			Y			
Percent for Comment Code 6			Y			
Percent for Comment Code 7			Y			
Percent for Comment Code 8			Y			
Percent for Comment Code 9			Y			
Percent for Comment Code 10 (not used yet)			Y			
Percent for Comment Code 11 (not used yet)			Y			
Percent for Comment Code 12 (not used yet)			Y			
Percent for Condition Code A						

Percent for Condition Code B						
Percent for Condition Code C						
Percent for Condition Code D						
Percent for Condition Code E						
Percent for Condition Code F (not used yet)						
Percent for Condition Code G (not used yet)						
Percent for Condition Code H (not used yet)						
Percent for Condition Code I (not used yet)						
Percent for Condition Code J (not used yet)						
Percent for Condition Code K (not used yet)						
Percent for Condition Code L (not used yet)						
Percent (option A)			Y		Y	
Percent (option B)			Y		Y	
Percent (option C)			Y		Y	
Percent (option D)			Y		Y	
Percent (option E)			Y			
Percent (mult. marks)			Y		Y	
Percent (Omits)			Y		Y	
Percent with (scorepoint 0)			Y			
Percent with (scorepoint 1)			Y			
Percent with (scorepoint 2)			Y			
Percent with (scorepoint 3)			Y			
Percent with (scorepoint 4)			Y			
Percent with (scorepoint 5)			Y			

Percent with (scorepoint 6)			Y			
Percent with (scorepoint 7)			Y			
Percent with (scorepoint 8)			Y			
Percent with (scorepoint 9)			Y			
Percent with (scorepoint 10)			Y			
Percent with (scorepoint 11)			Y			
Percent with (scorepoint 12)			Y			
Male Percent (A)			Y		Y	
Male Percent (B)			Y		Y	
Male Percent (C)			Y		Y	
Male Percent (D)			Y		Y	
Male Percent (E)			Y			
Male Percent (mult. marks)			Y		Y	
Male Percent (Omits)			Y		Y	
Male Percent (scorepoint 0)			Y			
Male Percent (scorepoint 1)			Y			
Male Percent (scorepoint 2)			Y			
Male Percent (scorepoint 3)			Y			
Male Percent (scorepoint 4)			Y			
Male Percent (scorepoint 5)			Y			
Male Percent (scorepoint 6)			Y			
Male Percent (scorepoint 7)			Y			
Male Percent (scorepoint 8)			Y			

Male Percent (scorepoint 9)			Y			
Male Percent (scorepoint 10)			Y			
Male Percent (scorepoint 11)			Y			
Male Percent (scorepoint 12)			Y			
Female Percent (A)			Y		Y	
Female Percent (B)			Y		Y	
Female Percent (C)			Y		Y	
Female Percent (D)			Y		Y	
Female Percent (E)			Y			
Female Percent (mult. marks)			Y		Y	
Female Percent (Omits)			Y		Y	
Female Percent (scorepoint 0)			Y			
Female Percent (scorepoint 1)			Y			
Female Percent (scorepoint 2)			Y			
Female Percent (scorepoint 3)			Y			
Female Percent (scorepoint 4)			Y			
Female Percent (scorepoint 5)			Y			
Female Percent (scorepoint 6)			Y			
Female Percent (scorepoint 7)			Y			
Female Percent (scorepoint 8)			Y			
Female Percent (scorepoint 9)			Y			
Female Percent (scorepoint 10)			Y			
Female Percent (scorepoint 11)			Y			

Female Percent (scorepoint 12)			Y			
White Percent (A)			Y		Y	
White Percent (B)			Y		Y	
White Percent (C)			Y		Y	
White Percent (D)			Y		Y	
White Percent (E)			Y			
White Percent (mult. marks)			Y		Y	
White Percent (Omits)			Y		Y	
White Percent (scorepoint 0)			Y			
White Percent (scorepoint 1)			Y			
White Percent (scorepoint 2)			Y			
White Percent (scorepoint 3)			Y			
White Percent (scorepoint 4)			Y			
White Percent (scorepoint 5)			Y			
White Percent (scorepoint 6)			Y			
White Percent (scorepoint 7)			Y			
White Percent (scorepoint 8)			Y			
White Percent (scorepoint 9)			Y			
White Percent (scorepoint 10)			Y			
White Percent (scorepoint 11)			Y			
White Percent (scorepoint 12)			Y			
Black Percent (A)			Y		Y	
Black Percent (B)			Y		Y	
Black Percent (C)			Y		Y	
Black Percent (D)			Y		Y	
Black Percent (E)			Y			

Black Percent (mult. marks)			Y		Y	
Black Percent (Omits)			Y		Y	
Black Percent (scorepoint 0)			Y			
Black Percent (scorepoint 1)			Y			
Black Percent (scorepoint 2)			Y			
Black Percent (scorepoint 3)			Y			
Black Percent (scorepoint 4)			Y			
Black Percent (scorepoint 5)			Y			
Black Percent (scorepoint 6)			Y			
Black Percent (scorepoint 7)			Y			
Black Percent (scorepoint 8)			Y			
Black Percent (scorepoint 9)			Y			
Black Percent (scorepoint 10)			Y			
Black Percent (scorepoint 11)			Y			
Black Percent (scorepoint 12)			Y			
Hispanic Percent (A)						
Hispanic Percent (B)						
Hispanic Percent (C)						
Hispanic Percent (D)						
Hispanic Percent (E)						
Hispanic Percent (mult. marks)						
Hispanic Percent (Omits)						
Hispanic Percent (scorepoint 0)						
Hispanic Percent (scorepoint 1)						

Hispanic Percent (scorepoint 2)						
Hispanic Percent (scorepoint 3)						
Hispanic Percent (scorepoint 4)						
Hispanic Percent (scorepoint 5)						
Hispanic Percent (scorepoint 6)						
Hispanic Percent (scorepoint 7)						
Hispanic Percent (scorepoint 8)						
Hispanic Percent (scorepoint 9)						
Hispanic Percent (scorepoint 10)						
Hispanic Percent (scorepoint 11)						
Hispanic Percent (scorepoint 12)						
Asian/Pacific Islander Percent (A)						
Asian/Pacific Islander Percent (B)						
Asian/Pacific Islander Percent (C)						
Asian/Pacific Islander Percent (D)						
Asian/Pacific Islander Percent (E)						
Asian/Pacific Islander Percent (mult. marks)						
Asian/Pacific Islander Percent (Omits)						
Asian/Pacific Islander Percent (scorepoint 0)						
Asian/Pacific Islander Percent (scorepoint 1)						

Asian/Pacific Islander Percent (scorepoint 2)						
Asian/Pacific Islander Percent (scorepoint 3)						
Asian/Pacific Islander Percent (scorepoint 4)						
Asian/Pacific Islander Percent (scorepoint 5)						
Asian/Pacific Islander Percent (scorepoint 6)						
Asian/Pacific Islander Percent (scorepoint 7)						
Asian/Pacific Islander Percent (scorepoint 8)						
Asian/Pacific Islander Percent (scorepoint 9)						
Asian/Pacific Islander Percent (scorepoint 10)						
Asian/Pacific Islander Percent (scorepoint 11)						
Asian/Pacific Islander Percent (scorepoint 12)						
Accommodated Percent (A)			Y		Y	In a separate analysis data file
Accommodated Percent (B)			Y		Y	In a separate analysis data file
Accommodated Percent (C)			Y		Y	In a separate analysis data file
Accommodated Percent (D)			Y		Y	In a separate analysis data file
Accommodated Percent (E)			Y			
Accommodated Percent (mult. marks)			Y		Y	In a separate analysis data file
Accommodated Percent (Omits)			Y		Y	In a separate analysis data file
Accommodated Percent (scorepoint 0)			Y			
Accommodated Percent (scorepoint 1)			Y			

Accommodated Percent (scorepoint 2)			Y			
Accommodated Percent (scorepoint 3)			Y			
Accommodated Percent (scorepoint 4)			Y			
Accommodated Percent (scorepoint 5)			Y			
Accommodated Percent (scorepoint 6)			Y			
Accommodated Percent (scorepoint 7)			Y			
Accommodated Percent (scorepoint 8)			Y			
Accommodated Percent (scorepoint 9)			Y			
Accommodated Percent (scorepoint 10)			Y			
Accommodated Percent (scorepoint 11)			Y			
Accommodated Percent (scorepoint 12)			Y			
Non-Accommodated Percent (A)			Y			
Non-Accommodated Percent (B)			Y			
Non-Accommodated Percent (C)			Y			
Non-Accommodated Percent (D)			Y			
Non-Accommodated Percent (E)			Y			
Non-Accommodated Percent (mult. marks)			Y			
Non-Accommodated Percent (Omits)			Y			
Non-Accommodated Percent (scorepoint 0)			Y			
Non-Accommodated Percent (scorepoint 1)			Y			

Non-Accommodated Percent (scorepoint 2)			Y			
Non-Accommodated Percent (scorepoint 3)			Y			
Non-Accommodated Percent (scorepoint 4)			Y			
Non-Accommodated Percent (scorepoint 5)			Y			
Non-Accommodated Percent (scorepoint 6)			Y			
Non-Accommodated Percent (scorepoint 7)			Y			
Non-Accommodated Percent (scorepoint 8)			Y			
Non-Accommodated Percent (scorepoint 9)			Y			
Non-Accommodated Percent (scorepoint 10)			Y			
Non-Accommodated Percent (scorepoint 11)			Y			
Non-Accommodated Percent (scorepoint 12)			Y			
Economically Disadvantaged Percent (A)			Y			
Economically Disadvantaged Percent (B)			Y			
Economically Disadvantaged Percent (C)			Y			
Economically Disadvantaged Percent (D)			Y			
Economically Disadvantaged Percent (E)			Y			
Economically Disadvantaged Percent (mult. marks)			Y			

Economically Disadvantaged Percent (Omits)			Y			
Economically Disadvantaged Percent (scorepoint 0)			Y			
Economically Disadvantaged Percent (scorepoint 1)			Y			
Economically Disadvantaged Percent (scorepoint 2)			Y			
Economically Disadvantaged Percent (scorepoint 3)			Y			
Economically Disadvantaged Percent (scorepoint 4)			Y			
Economically Disadvantaged Percent (scorepoint 5)			Y			
Economically Disadvantaged Percent (scorepoint 6)			Y			
Economically Disadvantaged Percent (scorepoint 7)			Y			
Economically Disadvantaged Percent (scorepoint 8)			Y			
Economically Disadvantaged Percent (scorepoint 9)			Y			
Economically Disadvantaged Percent (scorepoint 10)			Y			
Economically Disadvantaged Percent (scorepoint 11)			Y			

Economically Disadvantaged Percent (scorepoint 12)			Y			
Non Economically Disadvantaged Percent (A)			Y			
Non Economically Disadvantaged Percent (B)			Y			
Non Economically Disadvantaged Percent (C)			Y			
Non Economically Disadvantaged Percent (D)			Y			
Non Economically Disadvantaged Percent (E)			Y			
Non Economically Disadvantaged Percent (mult. marks)			Y			
Non Economically Disadvantaged Percent (Omits)			Y			
Non Economically Disadvantaged Percent (scorepoint 0)			Y			
Non Economically Disadvantaged Percent (scorepoint 1)			Y			
Non Economically Disadvantaged Percent (scorepoint 2)			Y			
Non Economically Disadvantaged Percent (scorepoint 3)			Y			
Non Economically Disadvantaged Percent (scorepoint 4)			Y			

Non Economically Disadvantaged Percent (scorepoint 5)			Y			
Non Economically Disadvantaged Percent (scorepoint 6)			Y			
Non Economically Disadvantaged Percent (scorepoint 7)			Y			
Non Economically Disadvantaged Percent (scorepoint 8)			Y			
Non Economically Disadvantaged Percent (scorepoint 9)			Y			
Non Economically Disadvantaged Percent (scorepoint 10)			Y			
Non Economically Disadvantaged Percent (scorepoint 11)			Y			
Non Economically Disadvantaged Percent (scorepoint 12)			Y			
English Language Learner Percent (A)			Y			
English Language Learner Percent (B)			Y			
English Language Learner Percent (C)			Y			
English Language Learner Percent (D)			Y			
English Language Learner Percent (E)			Y			
English Language Learner Percent (mult. marks)			Y			
English Language Learner Percent (Omits)			Y			

English Language Learner Percent (scorepoint 0)			Y			
English Language Learner Percent (scorepoint 1)			Y			
English Language Learner Percent (scorepoint 2)			Y			
English Language Learner Percent (scorepoint 3)			Y			
English Language Learner Percent (scorepoint 4)			Y			
English Language Learner Percent (scorepoint 5)			Y			
English Language Learner Percent (scorepoint 6)			Y			
English Language Learner Percent (scorepoint 7)			Y			
English Language Learner Percent (scorepoint 8)			Y			
English Language Learner Percent (scorepoint 9)			Y			
English Language Learner Percent (scorepoint 10)			Y			
English Language Learner Percent (scorepoint 11)			Y			
English Language Learner Percent (scorepoint 12)			Y			
Non-English Language Learner Percent (A)			Y			

Non-English Language Learner Percent (B)			Y			
Non-English Language Learner Percent (C)			Y			
Non-English Language Learner Percent (D)			Y			
Non-English Language Learner Percent (E)			Y			
Non-English Language Learner Percent (mult. marks)			Y			
Non-English Language Learner Percent (Omits)			Y			
Non-English Language Learner Percent (scorepoint 0)			Y			
Non-English Language Learner Percent (scorepoint 1)			Y			
Non-English Language Learner Percent (scorepoint 2)			Y			
Non-English Language Learner Percent (scorepoint 3)			Y			
Non-English Language Learner Percent (scorepoint 4)			Y			
Non-English Language Learner Percent (scorepoint 5)			Y			
Non-English Language Learner Percent (scorepoint 6)			Y			
Non-English Language Learner Percent (scorepoint 7)			Y			
Non-English Language Learner Percent (scorepoint 8)			Y			

Non-English Language Learner Percent (scorepoint 9)			Y			
Non-English Language Learner Percent (scorepoint 10)			Y			
Non-English Language Learner Percent (scorepoint 11)			Y			
Non-English Language Learner Percent (scorepoint 12)			Y			
Disabled Learner Percent (A)			Y			
Disabled Learner Percent (B)			Y			
Disabled Learner Percent (C)			Y			
Disabled Learner Percent (D)			Y			
Disabled Learner Percent (E)			Y			
Disabled Learner Percent (mult. marks)			Y			
Disabled Learner Percent (Omits)			Y			
Disabled Learner Percent (scorepoint 0)			Y			
Disabled Learner Percent (scorepoint 1)			Y			
Disabled Learner Percent (scorepoint 2)			Y			
Disabled Learner Percent (scorepoint 3)			Y			
Disabled Learner Percent (scorepoint 4)			Y			
Disabled Learner Percent (scorepoint 5)			Y			
Disabled Learner Percent (scorepoint 6)			Y			

Disabled Learner Percent (scorepoint 7)			Y			
Disabled Learner Percent (scorepoint 8)			Y			
Disabled Learner Percent (scorepoint 9)			Y			
Disabled Learner Percent (scorepoint 10)			Y			
Disabled Learner Percent (scorepoint 11)			Y			
Disabled Learner Percent (scorepoint 12)			Y			
Non-Disabled Learner Percent (A)			Y			
Non-Disabled Learner Percent (B)			Y			
Non-Disabled Learner Percent (C)			Y			
Non-Disabled Learner Percent (D)			Y			
Non-Disabled Learner Percent (E)			Y			
Non-Disabled Learner Percent (mult. marks)			Y			
Non-Disabled Learner Percent (Omits)			Y			
Non-Disabled Learner Percent (scorepoint 0)			Y			
Non-Disabled Learner Percent (scorepoint 1)			Y			
Non-Disabled Learner Percent (scorepoint 2)			Y			
Non-Disabled Learner Percent (scorepoint 3)			Y			
Non-Disabled Learner Percent (scorepoint 4)			Y			
Non-Disabled Learner Percent (scorepoint 5)			Y			
Non-Disabled Learner Percent (scorepoint 6)			Y			

Non-Disabled Learner Percent (scorepoint 7)			Y			
Non-Disabled Learner Percent (scorepoint 8)			Y			
Non-Disabled Learner Percent (scorepoint 9)			Y			
Non-Disabled Learner Percent (scorepoint 10)			Y			
Non-Disabled Learner Percent (scorepoint 11)			Y			
Non-Disabled Learner Percent (scorepoint 12)			Y			
P-value or Item Mean			Y		Y	
P-value or Item Mean for Male			Y		Y	
P-value or Item Mean for Female			Y		Y	
P-value or Item Mean for White			Y		Y	
P-value or Item Mean for Black			Y		Y	
P-value or Item Mean for Hispanic			Y			
P-value or Item Mean for Asian/Pacific Islander			Y			
P-value or Item Mean for Accommodated			Y		Y	In a separate analysis data file
P-value or Item Mean for Non-Accommodated			Y			
P-value or Item Mean for Economically Disadvantage			Y			
P-value or Item Mean for Non-Economically Disadvantage			Y			
P-value or Item Mean for English Language Learner			Y			
P-value or Item Mean for Non-English Language Learner			Y			

P-value or Item Mean for Disabled Leaner						
P-value or Item Mean for Non-Disabled Leaner						
Adjusted P-value	same as pvalue for MC items		Y		Y	
Adjusted P-value for Male			Y		Y	
Adjusted P-value for Female			Y		Y	
Adjusted P-value for White			Y		Y	
Adjusted P-value for Black			Y		Y	
Adjusted P-value for Hispanic			Y			
Adjusted P-value for Asian/Pacific Islander			Y			
Adjusted P-value for Accommodated			Y		Y	In a separate data file
Adjusted P-value for Non-Accommodated			Y			
Adjusted P-value for Economically Disadvantage			Y			
Adjusted P-value for Non-Economically Disadvantage			Y			
Adjusted P-value for English Language Learner			Y			
Adjusted P-value for Non-English Language Learner			Y			
Adjusted P-value for Disabled Learner						
Adjusted P-value for Non-Disabled Learner						
Difficulty flag			Y		Y	
Item Standard Deviation			Y		Y	
Item-Total Correlation			Y		Y	
Biserial / Polyserial Correlation			Y		Y	

Point-Biserial Correlation (corrected)			Y		Y	
Item-Total correlation flag			Y		Y	
P-b correlation for option A			Y		Y	
P-b correlation for option B			Y		Y	
P-b correlation for option C			Y		Y	
P-b correlation for option D			Y		Y	
P-b correlation for option E			Y			
P-b correlation for Omits			Y		Y	
P-b correlation for option A-Male			Y			
P-b correlation for option B-Male			Y			
P-b correlation for option C-Male			Y			
P-b correlation for option D-Male			Y			
P-b correlation for option E-Male			Y			
P-b correlation for Omits-Male			Y			
P-b correlation for option A-Female			Y			
P-b correlation for option B-Female			Y			
P-b correlation for option C-Female			Y			
P-b correlation for option D-Female			Y			
P-b correlation for option E-Female			Y			
P-b correlation for Omits-Female			Y			
P-b correlation for option A-White			Y			
P-b correlation for option B--White			Y			

P-b correlation for option C-White			Y			
P-b correlation for option D-White			Y			
P-b correlation for option E-White			Y			
P-b correlation for Omits O-White			Y			
P-b correlation for option A-Black			Y			
P-b correlation for option B-Black			Y			
P-b correlation for option C-Black			Y			
P-b correlation for option D-Black			Y			
P-b correlation for option E-Black			Y			
P-b correlation for Omits O-Black			Y			
P-b correlation for option A_HIS						
P-b correlation for option B_HIS						
P-b correlation for option C_HIS						
P-b correlation for option D_HIS						
P-b correlation for option E_HIS						
P-b correlation for Omits_HIS						
P-b correlation for option A_API						
P-b correlation for option B_API						
P-b correlation for option C_API						
P-b correlation for option D_API						
P-b correlation for option E_API						
P-b correlation for Omits_API						

P-b correlation for option A-ACCOM			Y		Y	in a separate file
P-b correlation for option B-ACCOM			Y		Y	in a separate file
P-b correlation for option C-ACCOM			Y		Y	in a separate file
P-b correlation for option D-ACCOM			Y		Y	in a separate file
P-b correlation for option E-ACCOM			Y			
P-b correlation for Omits O-ACCOM			Y		Y	in a separate file
P-b correlation for option A-NACCOM			Y			
P-b correlation for option B-NACCOM			Y			
P-b correlation for option C-NACCOM			Y			
P-b correlation for option D-NACCOM			Y			
P-b correlation for option E-NACCOM			Y			
P-b correlation for Omits O-NACCOM			Y			
P-b correlation for option A_ED			Y			
P-b correlation for option B-ED			Y			
P-b correlation for option C-ED			Y			
P-b correlation for option D-ED			Y			
P-b correlation for option E-ED			Y			
P-b correlation for Omits O-ED			Y			
P-b correlation for option A-NED			Y			
P-b correlation for option B-NED			Y			

P-b correlation for option C-NED			Y			
P-b correlation for option D-NED			Y			
P-b correlation for option E-NED			Y			
P-b correlation for Omits-NED			Y			
P-b correlation for option A-ELL			Y			
P-b correlation for option B-ELL			Y			
P-b correlation for option C-ELL			Y			
P-b correlation for option D-ELL			Y			
P-b correlation for option E-ELL			Y			
P-b correlation for Omits O-ELL			Y			
P-b correlation for option A-NELL			Y			
P-b correlation for option B-NELL			Y			
P-b correlation for option C-NELL			Y			
P-b correlation for option D-NELL			Y			
P-b correlation for option E-NELL			Y			
P-b correlation for Omits O-NELL			Y			
P-b correlation for option A-DIS			Y			
P-b correlation for option B-DIS			Y			
P-b correlation for option C-DIS			Y			
P-b correlation for option D-DIS			Y			

P-b correlation for option E-DIS			Y			
P-b correlation for Omits O-DIS			Y			
P-b correlation for option A-NDIS			Y			
P-b correlation for option B-NDIS			Y			
P-b correlation for option C-NDIS			Y			
P-b correlation for option D-NDIS			Y			
P-b correlation for option E-NDIS			Y			
P-b correlation for Omits O-NDIS			Y			
Flag for potential miskeying. Values can be: H, L, P, O, N, B					Y	
Mantel CHSQ Male-Female					Y	
Lower limit of 95% CI					Y	
Mantel-Haenszel Delta Male-Female					Y	
Upper limit of 95% CI					Y	
Mantel CHSQ White-Black					Y	
Lower limit of 95% CI					Y	
Mantel-Haenszel Delta White-Black					Y	
Upper limit of 95% CI					Y	
Mantel CHSQ White to Hispanic						
Mantel-Haenszel Delta White to Hispanic						
Mantel CHSQ White to Asian/Pacific Island						
Mantel-Haenszel Delta White to Asian/Pacific Island						

Mantel CHSQ Accommodated to Non-Accommodated						
Mantel-Haenszel Delta Accommodated to Non-Accommodated						
Mantel CHSQ Economically Disadvantage Non-Economically Disadvantage						
Mantel-Haenszel Delta Economically Disadvantage Non-Economically Disadvantage						
Mantel CHSQ English Language Learner to Non_English Language Learner						
Mantel-Haenszel Delta English Language Learner to Non_English Language Learner						
Mantel CHSQ Disabled Learner to Non_Disabled Learner						
Mantel-Haenszel Delta Disabled Learner to Non_Disabled Learner						
SMD M-F						
SMD W-B						
SMDS W to HIS						
SMDS W to API						
SMDS ACC to NACC						
SMDS ED to NED						
SMDS ELL to NELL						
SMDS DIS to NDIS						
DIF category for M-F (A, B, C, AA, BB, CC)			Y		Y	

DIF category for W-B (A, B, C, AA, BB, CC)			Y		Y	
DIF category for White to HIS (A, B, C, AA, BB, CC)						
DIF category for White to API (A, B, C, AA, BB, CC)						
DIF category for ACC to NACC (A, B, C, AA, BB, CC)			Y			
DIF category for ED to NED(A, B, C, AA, BB, CC)			Y			
DIF Category for ELL to NELL (A, B, C, AA, BB, CC)			Y			
DIF Category for DIS to NDis (A, B, C, AA, BB, CC)						
Favored group for M-F (Male, Female)					Y	
Favored group for W-B (White, Black)					Y	
Favored group for White to Hispanic						
Favored group for White to Asian/Pacific Island)						
Favored group for (Accommodated, Non-Accommodated)						
Favored group for (Economically Disadvantage, Non-Economically Disadvantage)						
Favored group for (English Language Learner , Non_English Language Learner)						
Favored group for (Disabled Learner , Non_Disabled Learner)						
A parameter (scaled)			Y		Y	

SE for A parameter (scaled)			Y		Y	
B parameter (scaled)			Y		Y	
SE for B parameter (scaled)			Y		Y	
C parameter (scaled)			Y		Y	
SE for C parameter (scaled)			Y		Y	
D1 category parameter (scaled)						
SE for D1 category parameter (scaled)						
D2 category parameter (scaled)						
SE for D2 category parameter (scaled)						
D3 category parameter (scaled)						
SE for D3 category parameter (scaled)						
D4 category parameter (scaled)						
SE for D4 category parameter (scaled)						
D5 category parameter (scaled)						
SE for D5 category parameter (scaled)						
D6 category parameter (scaled)						
SE for D6 category parameter (scaled)						
D7 category parameter (scaled)						
SE for D7 category parameter (scaled)						
D8 category parameter (scaled)						
SE for D8 category parameter (scaled)						

D9 category parameter (scaled)						
SE for D9 category parameter (scaled)						
D10 category parameter (scaled)						
SE for D10 category parameter (scaled)						
D11 category parameter (scaled)						
SE for D11 category parameter (scaled)						
D12 category parameter (scaled)						
SE for D12 category parameter (scaled)						
ZQ1 fit index			Y			
Fit Flag based on ZQ1 (blank, F)			Y			
Mean-square infit			Y		Y	
Mean-square infit flag (blank, F, FF)			Y			
Mean-square outfit			Y		Y	
Mean-square fit flag (blank, MM, MH, TP)			Y			
Mean-square outfit			Y		Y	
Fit level (0, 1, 2)			Y		Y	
Chi-Square statistics for 3PL and GPC fit index computed by PARSCALE					Y	
Degrees of freedom associated with the Chi-square fit index computed					Y	
P_value associated with the Chi-square fit index computed					Y	
IRT fit statistics for PARSCALE calibrated					Y	
Degrees of freedom associated with the sx2					Y	

p_value associated with the sx2					Y	
Fit flag based on sx2 statistics					Y	
(Theta cut for Basic)					Y	
(Theta cut for MET)					Y	
(Theta cut for EXCEED)					Y	
(ICC at cut for Basic)					Y	
(ICC at cut for MET)					Y	
(ICC at cut for Exceed)					Y	
Item information at PL cut point 1					Y	
Item information at PL cut point 2					Y	
Item information at PL cut point 3					Y	
Theta point 1			Y		Y	
Theta point 2			Y		Y	
Theta point 3			Y		Y	
Theta point 4			Y		Y	
Theta point 5			Y		Y	
Theta point 6			Y		Y	
Theta point 7			Y		Y	
Theta point 8			Y		Y	
Theta point 9			Y		Y	
Theta point 10			Y		Y	
Conditional Item Mean for Decile 1			Y		Y	
Conditional Item Mean for Decile 2			Y		Y	
Conditional Item Mean for Decile 3			Y		Y	
Conditional Item Mean for Decile 4			Y		Y	
Conditional Item Mean for Decile 5			Y		Y	
Conditional Item Mean for Decile 6			Y		Y	
Conditional Item Mean for Decile 7			Y		Y	

Conditional Item Mean for Decile 8			Y		Y	
Conditional Item Mean for Decile 9			Y		Y	
Conditional Item Mean for Decile 10			Y		Y	
Conditional Item Mean for Decile 1			Y		Y	
Conditional Item Mean for Decile 2			Y		Y	
Conditional Item Mean for Decile 3			Y		Y	
Conditional Item Mean for Decile 4			Y		Y	
Conditional Item Mean for Decile 5			Y		Y	
Conditional Item Mean for Decile 6			Y		Y	
Conditional Item Mean for Decile 7			Y		Y	
Conditional Item Mean for Decile 8			Y		Y	
Conditional Item Mean for Decile 9			Y		Y	
Conditional Item Mean for Decile 10			Y		Y	
Conditional Item Mean for Decile 1			Y		Y	
Conditional Item Mean for Decile 2			Y		Y	
Conditional Item Mean for Decile 3			Y		Y	
Conditional Item Mean for Decile 4			Y		Y	
Conditional Item Mean for Decile 5			Y		Y	
Conditional Item Mean for Decile 6			Y		Y	
Conditional Item Mean for Decile 7			Y		Y	

Conditional Item Mean for Decile 8			Y		Y	
Conditional Item Mean for Decile 9			Y		Y	
Conditional Item Mean for Decile 10			Y		Y	
Conditional Item Mean for Decile 1			Y		Y	
Conditional Item Mean for Decile 2			Y		Y	
Conditional Item Mean for Decile 3			Y		Y	
Conditional Item Mean for Decile 4			Y		Y	
Conditional Item Mean for Decile 5			Y		Y	
Conditional Item Mean for Decile 6			Y		Y	
Conditional Item Mean for Decile 7			Y		Y	
Conditional Item Mean for Decile 8			Y		Y	
Conditional Item Mean for Decile 9			Y		Y	
Conditional Item Mean for Decile 10			Y		Y	
Conditional Item Mean for Decile 1			Y		Y	
Conditional Item Mean for Decile 2			Y		Y	
Conditional Item Mean for Decile 3			Y		Y	
Conditional Item Mean for Decile 4			Y		Y	
Conditional Item Mean for Decile 5			Y		Y	
Conditional Item Mean for Decile 6			Y		Y	
Conditional Item Mean for Decile 7			Y		Y	

Conditional Item Mean for Decile 8			Y		Y	
Conditional Item Mean for Decile 9			Y		Y	
Conditional Item Mean for Decile 10			Y		Y	
Conditional Item Mean for Decile 1						
Conditional Item Mean for Decile 2						
Conditional Item Mean for Decile 3						
Conditional Item Mean for Decile 4						
Conditional Item Mean for Decile 5						
Conditional Item Mean for Decile 6						
Conditional Item Mean for Decile 7						
Conditional Item Mean for Decile 8						
Conditional Item Mean for Decile 9						
Conditional Item Mean for Decile 10						
Conditional Item Mean for Decile 1						
Conditional Item Mean for Decile 2						
Conditional Item Mean for Decile 3						
Conditional Item Mean for Decile 4						
Conditional Item Mean for Decile 5						
Conditional Item Mean for Decile 6						
Conditional Item Mean for Decile 7						
Conditional Item Mean for Decile 8						

Conditional Item Mean for Decile 9						
Conditional Item Mean for Decile 10						
Conditional Item Mean for Decile 1			Y		Y	In a separate file
Conditional Item Mean for Decile 2			Y		Y	In a separate file
Conditional Item Mean for Decile 3			Y		Y	In a separate file
Conditional Item Mean for Decile 4			Y		Y	In a separate file
Conditional Item Mean for Decile 5			Y		Y	In a separate file
Conditional Item Mean for Decile 6			Y		Y	In a separate file
Conditional Item Mean for Decile 7			Y		Y	In a separate file
Conditional Item Mean for Decile 8			Y		Y	In a separate file
Conditional Item Mean for Decile 9			Y		Y	In a separate file
Conditional Item Mean for Decile 10			Y		Y	In a separate file
Conditional Item Mean for Decile 1			Y			
Conditional Item Mean for Decile 2			Y			
Conditional Item Mean for Decile 3			Y			
Conditional Item Mean for Decile 4			Y			
Conditional Item Mean for Decile 5			Y			
Conditional Item Mean for Decile 6			Y			
Conditional Item Mean for Decile 7			Y			
Conditional Item Mean for Decile 8			Y			

Conditional Item Mean for Decile 9			Y			
Conditional Item Mean for Decile 10			Y			
Conditional Item Mean for Decile 1			Y			
Conditional Item Mean for Decile 2			Y			
Conditional Item Mean for Decile 3			Y			
Conditional Item Mean for Decile 4			Y			
Conditional Item Mean for Decile 5			Y			
Conditional Item Mean for Decile 6			Y			
Conditional Item Mean for Decile 7			Y			
Conditional Item Mean for Decile 8			Y			
Conditional Item Mean for Decile 9			Y			
Conditional Item Mean for Decile 10			Y			
Conditional Item Mean for Decile 1			Y			
Conditional Item Mean for Decile 2			Y			
Conditional Item Mean for Decile 3			Y			
Conditional Item Mean for Decile 4			Y			
Conditional Item Mean for Decile 5			Y			
Conditional Item Mean for Decile 6			Y			
Conditional Item Mean for Decile 7			Y			
Conditional Item Mean for Decile 8			Y			

Conditional Item Mean for Decile 9			Y			
Conditional Item Mean for Decile 10			Y			
Conditional Item Mean for Decile 1			Y			
Conditional Item Mean for Decile 2			Y			
Conditional Item Mean for Decile 3			Y			
Conditional Item Mean for Decile 4			Y			
Conditional Item Mean for Decile 5			Y			
Conditional Item Mean for Decile 6			Y			
Conditional Item Mean for Decile 7			Y			
Conditional Item Mean for Decile 8			Y			
Conditional Item Mean for Decile 9			Y			
Conditional Item Mean for Decile 10			Y			
Conditional Item Mean for Decile 1			Y			
Conditional Item Mean for Decile 2			Y			
Conditional Item Mean for Decile 3			Y			
Conditional Item Mean for Decile 4			Y			
Conditional Item Mean for Decile 5			Y			
Conditional Item Mean for Decile 6			Y			
Conditional Item Mean for Decile 7			Y			
Conditional Item Mean for Decile 8			Y			

Conditional Item Mean for Decile 9			Y			
Conditional Item Mean for Decile 10			Y			
Conditional Item Mean for Decile 1			Y			
Conditional Item Mean for Decile 2			Y			
Conditional Item Mean for Decile 3			Y			
Conditional Item Mean for Decile 4			Y			
Conditional Item Mean for Decile 5			Y			
Conditional Item Mean for Decile 6			Y			
Conditional Item Mean for Decile 7			Y			
Conditional Item Mean for Decile 8			Y			
Conditional Item Mean for Decile 9			Y			
Conditional Item Mean for Decile 10			Y			
Conditional Item Mean for Decile 1			Y			
Conditional Item Mean for Decile 2			Y			
Conditional Item Mean for Decile 3			Y			
Conditional Item Mean for Decile 4			Y			
Conditional Item Mean for Decile 5			Y			
Conditional Item Mean for Decile 6			Y			
Conditional Item Mean for Decile 7			Y			
Conditional Item Mean for Decile 8			Y			

Conditional Item Mean for Decile 9			Y			
Conditional Item Mean for Decile 10			Y			
			Y			
			Y			
95th percentile	Box & whisker plot: All	Option A			Y	95th percentile of theta for all students for Option A
75th percentile					Y	75th percentile of theta for all students for Option A
50th percentile					Y	50th percentile of theta for all students for Option A
25th percentile					Y	25th percentile of theta for all students for Option A
5th percentile					Y	5th percentile of theta for all students for Option A
95th percentile	Box & whisker plot: Males				Y	95th percentile of theta for male students for Option A
75th percentile					Y	75th percentile of theta for male students for Option A
50th percentile					Y	50th percentile of theta for male students for Option A
25th percentile					Y	25th percentile of theta for male students for Option A
5th percentile					Y	5th percentile of theta for male students for Option A

95th percentile	Box & whisker plot: Females				Y	95th percentile of theta for female students for Option A
75th percentile					Y	75th percentile of theta for female students for Option A
50th percentile					Y	50th percentile of theta for female students for Option A
25th percentile					Y	25th percentile of theta for female students for Option A
5th percentile					Y	5th percentile of theta for female students for Option A
95th percentile	Box & whisker plot: Whites				Y	95th percentile of theta for white students for Option A
75th percentile					Y	75th percentile of theta for white students for Option A
50th percentile					Y	50th percentile of theta for white students for Option A
25th percentile					Y	25th percentile of theta for white students for Option A
5th percentile					Y	5th percentile of theta for white students for Option A
95th percentile	Box & whisker plot: Blacks				Y	95th percentile of theta for black students for Option A
75th percentile					Y	75th percentile of theta for black students for Option A
50th percentile					Y	50th percentile of theta for black students for Option A
25th percentile					Y	25th percentile of theta for black students for Option A

5th percentile					Y	5th percentile of theta for black students for Option A
95th percentile	Box & whisker plot: All	Option B			Y	95th percentile of theta for all students for Option B
75th percentile					Y	75th percentile of theta for all students for Option B
50th percentile					Y	50th percentile of theta for all students for Option B
25th percentile					Y	25th percentile of theta for all students for Option B
5th percentile					Y	5th percentile of theta for all students for Option B
95th percentile	Box & whisker plot: Males				Y	95th percentile of theta for male students for Option B
75th percentile					Y	75th percentile of theta for male students for Option B
50th percentile					Y	50th percentile of theta for male students for Option B
25th percentile					Y	25th percentile of theta for male students for Option B
5th percentile					Y	5th percentile of theta for male students for Option B
95th percentile	Box & whisker plot: Females				Y	95th percentile of theta for female students for Option B
75th percentile					Y	75th percentile of theta for female students for Option B
50th percentile					Y	50th percentile of theta for female students for Option B

25th percentile					Y	25th percentile of theta for female students for Option B
5th percentile					Y	5th percentile of theta for female students for Option B
95th percentile	Box & whisker plot: Whites				Y	95th percentile of theta for white students for Option B
75th percentile					Y	75th percentile of theta for white students for Option B
50th percentile					Y	50th percentile of theta for white students for Option B
25th percentile					Y	25th percentile of theta for white students for Option B
5th percentile					Y	5th percentile of theta for white students for Option B
95th percentile	Box & whisker plot: Blacks				Y	95th percentile of theta for black students for Option B
75th percentile					Y	75th percentile of theta for black students for Option B
50th percentile					Y	50th percentile of theta for black students for Option B
25th percentile					Y	25th percentile of theta for black students for Option B
5th percentile					Y	5th percentile of theta for black students for Option B
95th percentile	Box & whisker plot: All	Option C			Y	95th percentile of theta for all students for Option C
75th percentile					Y	75th percentile of theta for all students for Option C

50th percentile					Y	50th percentile of theta for all students for Option C
25th percentile					Y	25th percentile of theta for all students for Option C
5th percentile					Y	5th percentile of theta for all students for Option C
95th percentile	Box & whisker plot: Males				Y	95th percentile of theta for male students for Option C
75th percentile					Y	75th percentile of theta for male students for Option C
50th percentile					Y	50th percentile of theta for male students for Option C
25th percentile					Y	25th percentile of theta for male students for Option C
5th percentile					Y	5th percentile of theta for male students for Option C
95th percentile	Box & whisker plot: Females				Y	95th percentile of theta for female students for Option C
75th percentile					Y	75th percentile of theta for female students for Option C
50th percentile					Y	50th percentile of theta for female students for Option C
25th percentile					Y	25th percentile of theta for female students for Option C
5th percentile					Y	5th percentile of theta for female students for Option C
95th percentile	Box & whisker plot: Whites				Y	95th percentile of theta for white students for Option C

75th percentile					Y	75th percentile of theta for white students for Option C
50th percentile					Y	50th percentile of theta for white students for Option C
25th percentile					Y	25th percentile of theta for white students for Option C
5th percentile					Y	5th percentile of theta for white students for Option C
95th percentile	Box & whisker plot: Blacks				Y	95th percentile of theta for black students for Option C
75th percentile					Y	75th percentile of theta for black students for Option C
50th percentile					Y	50th percentile of theta for black students for Option C
25th percentile					Y	25th percentile of theta for black students for Option C
5th percentile					Y	5th percentile of theta for black students for Option C
95th percentile	Box & whisker plot: All	Option D			Y	95th percentile of theta for all students for Option D
75th percentile					Y	75th percentile of theta for all students for Option D
50th percentile					Y	50th percentile of theta for all students for Option D
25th percentile					Y	25th percentile of theta for all students for Option D
5th percentile					Y	5th percentile of theta for all students for Option D

95th percentile	Box & whisker plot: Males				Y	95th percentile of theta for male students for Option D
75th percentile					Y	75th percentile of theta for male students for Option D
50th percentile					Y	50th percentile of theta for male students for Option D
25th percentile					Y	25th percentile of theta for male students for Option D
5th percentile					Y	5th percentile of theta for male students for Option D
95th percentile	Box & whisker plot: Females				Y	95th percentile of theta for female students for Option D
75th percentile					Y	75th percentile of theta for female students for Option D
50th percentile					Y	50th percentile of theta for female students for Option D
25th percentile					Y	25th percentile of theta for female students for Option D
5th percentile					Y	5th percentile of theta for female students for Option D
95th percentile	Box & whisker plot: Whites				Y	95th percentile of theta for white students for Option D
75th percentile					Y	75th percentile of theta for white students for Option D
50th percentile					Y	50th percentile of theta for white students for Option D
25th percentile					Y	25th percentile of theta for white students for Option D

5th percentile					Y	5th percentile of theta for white students for Option D
95th percentile	Box & whisker plot: Blacks				Y	95th percentile of theta for black students for Option D
75th percentile					Y	75th percentile of theta for black students for Option D
50th percentile					Y	50th percentile of theta for black students for Option D
25th percentile					Y	25th percentile of theta for black students for Option D
5th percentile					Y	5th percentile of theta for black students for Option D
95th percentile	Box & whisker plot: All	Omits			Y	95th percentile of theta for all students for omits
75th percentile					Y	75th percentile of theta for all students for omits
50th percentile					Y	50th percentile of theta for all students for omits
25th percentile					Y	25th percentile of theta for all students for omits
5th percentile					Y	5th percentile of theta for all students for omits
95th percentile	Box & whisker plot: Males				Y	95th percentile of theta for male students for omits
75th percentile					Y	75th percentile of theta for male students for omits
50th percentile					Y	50th percentile of theta for male students for omits
25th percentile					Y	25th percentile of theta for male students for omits

5th percentile					Y	5th percentile of theta for male students for omits
95th percentile	Box & whisker plot: Females				Y	95th percentile of theta for female students for omits
75th percentile					Y	75th percentile of theta for female students for omits
50th percentile					Y	50th percentile of theta for female students for omits
25th percentile					Y	25th percentile of theta for female students for omits
5th percentile					Y	5th percentile of theta for female students for omits
95th percentile	Box & whisker plot: Whites				Y	95th percentile of theta for white students for omits
75th percentile					Y	75th percentile of theta for white students for omits
50th percentile					Y	50th percentile of theta for white students for omits
25th percentile					Y	25th percentile of theta for white students for omits
5th percentile					Y	5th percentile of theta for white students for omits
95th percentile	Box & whisker plot: Blacks				Y	95th percentile of theta for black students for omits
75th percentile					Y	75th percentile of theta for black students for omits
50th percentile					Y	50th percentile of theta for black students for omits
25th percentile					Y	25th percentile of theta for black students for omits

5th percentile					Y	5th percentile of theta for black students for omits

Appendix C: Statistics and Terms 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 Item mean divided by the difference between minimum and maximum score points. 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.

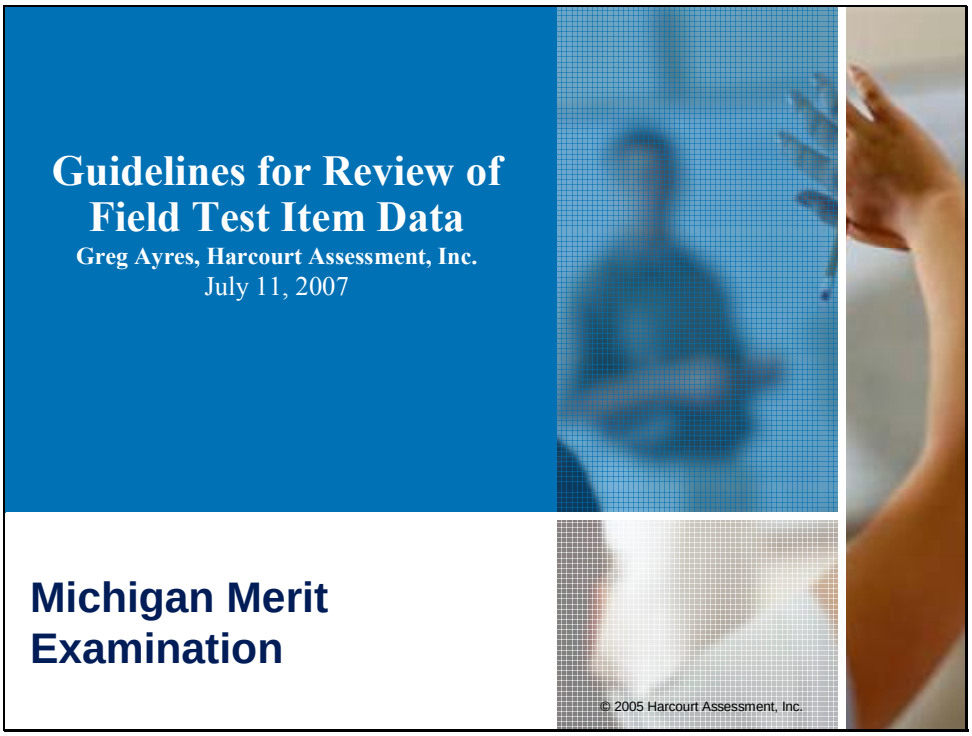
Item-Total Corr. Item-total correlation shows the relationship between a student's performance on the item and performance on the test as a whole. A high item-total correlation (e.g., above .50) indicates that students who earned more points on the item achieved higher total scores on the test than those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

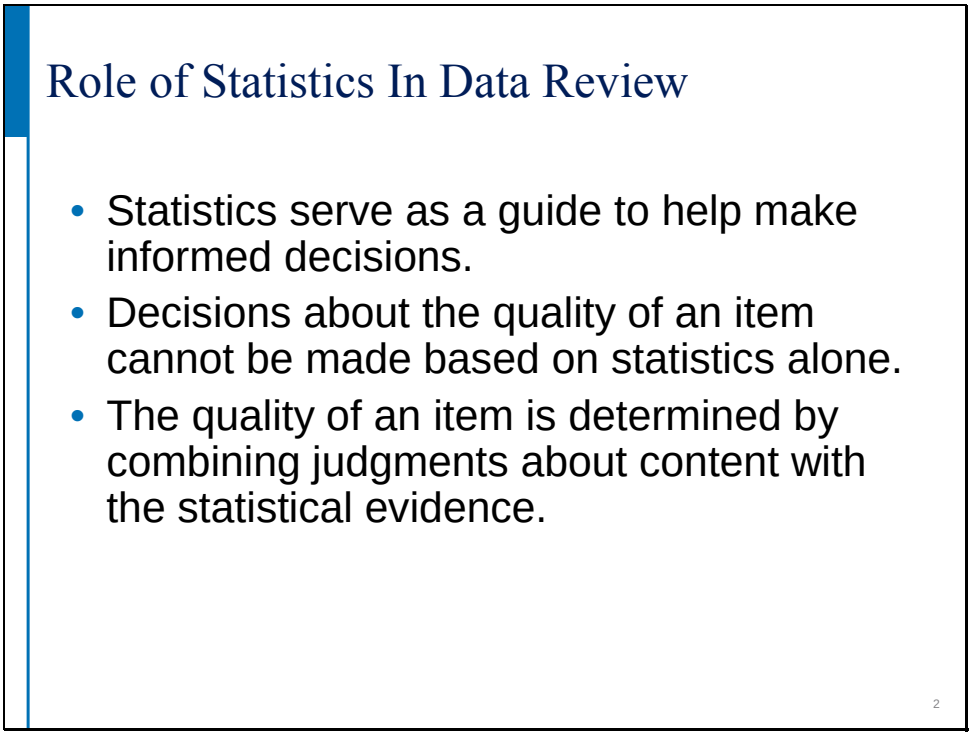
Difficulty Flag	This flag indicates that P-value, or adjusted p-value, or Rasch difficulty is out of the desired range.
PB Correlation Flag	This flag indicates that a MC item point-biserial correlation is smaller than the desired range of larger than 0.25.
Item-Total Corr. Flag	This flag indicates that 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

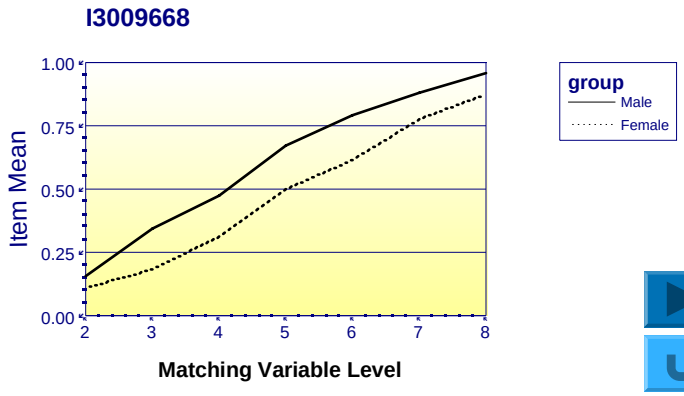
- Shows the percentage of students choosing each option on MC items, or earning a score point on CR items
- This percentage is given for all students and students grouped by ethnicity and gender.
- Option point-biserial correlation indicates the agreement between choosing each option (or earning a score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

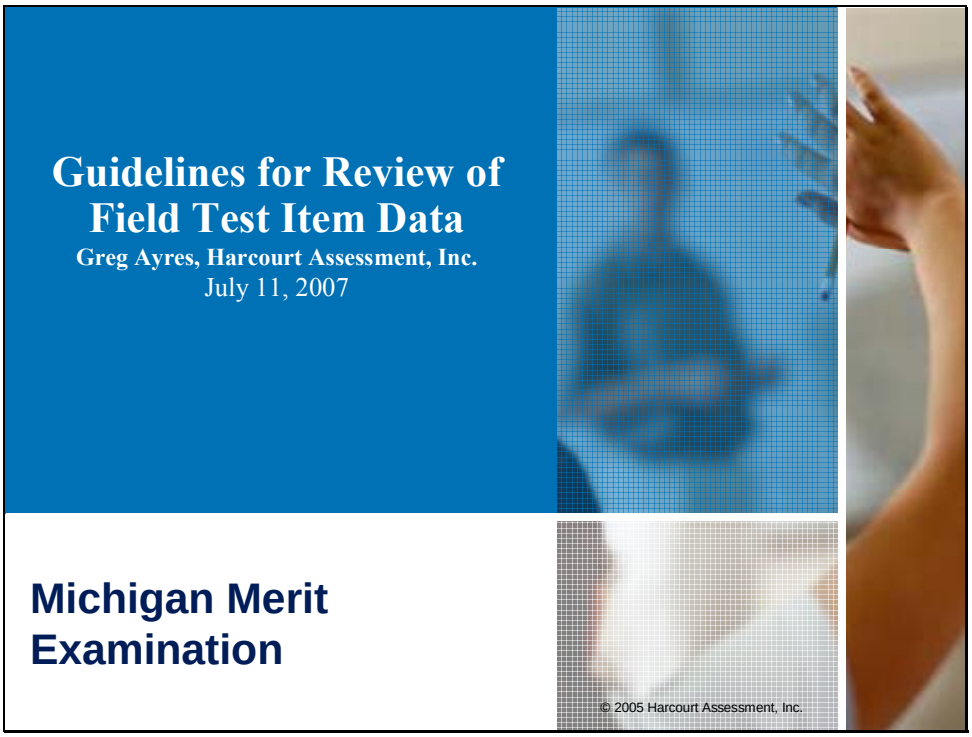
- Ask any questions that you may have
- Work in your respective subject area groups
- Enjoy the process

Thank you!

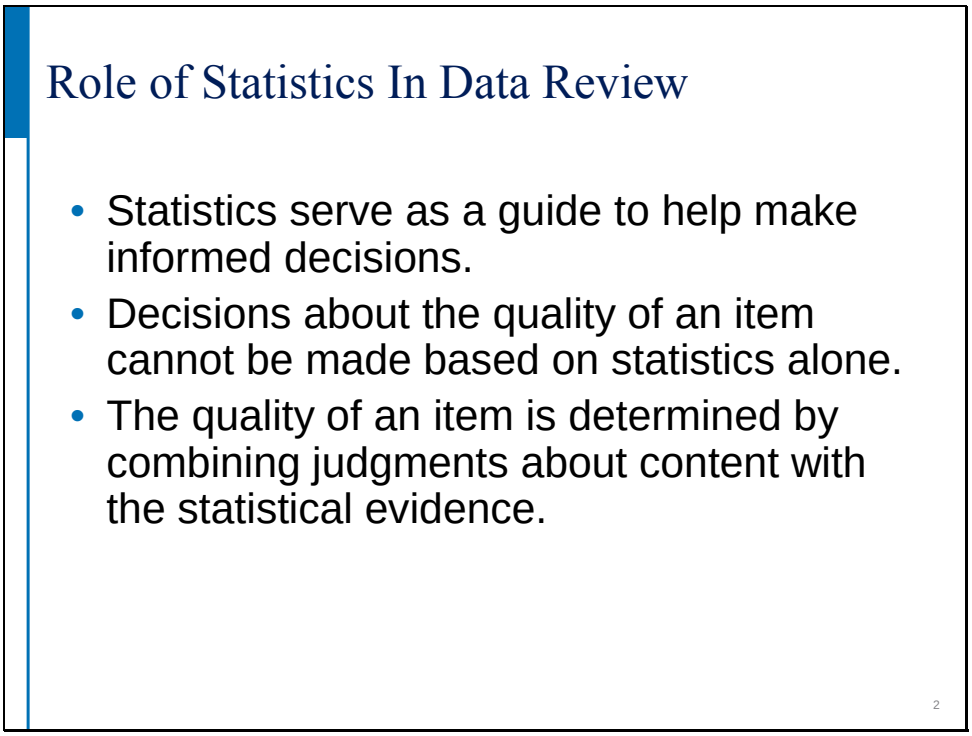
15

Appendix E: Guidelines for Content Review of Field Test Item Data

Slide 1



Slide 2



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