Maine Educational Assessment

Grades 5 and 8 Science

MeCAS Part I

2013–14 Technical Report

measured progress

100 EDUCATION WAY, DOVER, NH 38020 (800) 431-8901
WWW.MEASUREDPROGRESS.ORG

TABLE OF CONTENTS

CHAPTER 1 OVERVIEW OF MEA SCIENCE

The Maine Educational Assessment (MEA) currently assesses science and is administered to all students in grades 5 and 8 via standard administration, administration with accommodations, and/or alternate assessment. The test was administered to approximately 27,000 students during May 2014. The alternate assessment administration began in December 2013 and concluded at the end of April 2014.

The MEA is designed to be the measure of Maine's academic content standards in science, the *2007 Maine Learning Results* (MLRs), and to identify the knowledge and skills essential to prepare Maine students for work, higher education, citizenship, and personal fulfillment. These academic content standards express what students *should know* and *should be able to do* at various checkpoints during their education. They were developed using grade span standards in science (Pre-K–2, 3–5, 6–8, and 9–12).

1.1 PURPOSE OF THE ASSESSMENT SYSTEM

The purpose of Maine's Comprehensive Assessment System is to provide point-in-time information about the academic achievement and progress of Maine students. The MEA is one portion of this system and provides information in science. Student results are reported according to academic achievement descriptors utilizing cut scores established in standard setting for each of four achievement levels: Substantially Below Proficient, Partially Proficient, Proficient, and Proficient With Distinction. The results from this assessment and others provide educators and the public with information to guide future educational practices to meet the needs of students, while monitoring the continuous improvement efforts of schools, school administrative units (SAUs), and the state of Maine in achieving a world-class education system for all students.

1.2 CURRENT YEAR UPDATES

This year, the MEA Online Testing System was available to schools assessing their 8th graders in science. The paper-and-pencil forms were presented in an online version of the test. Thirty-five schools voluntarily participated; approximately 1,600 students participated in the online version of the test.

CHAPTER 2 TEST ADMINISTRATION

2.1 RESPONSIBILITY FOR ADMINISTRATION

As indicated in the *Principal/Test Coordinator Manual*, principals and/or their designated Maine Educational Assessment (MEA) coordinators were responsible for the proper administration of the MEA Science assessment. Manuals were used to ensure the uniformity of administration procedures from school to school. These manuals—the *Principal/Test Coordinator Manual*, the *Test Administrator Manual*, and the *Online Test Administration Manual*—stress the importance of test security and ethical administration while the tests are in the schools and contain explicit directions and scripts for test administrators to read aloud to test takers. These documents may be accessed on the Maine Department of Education's website at http://www.maine.gov/doe/mea/administration/index.html.

2.2 ADMINISTRATION PROCEDURES

MEA coordinators (principals or their designees) were instructed to read the *Principal/Test Coordinator Manual* prior to testing and to be familiar with the instructions given in the *Test Administrator Manual* and the *Online Test Administration Manual*. The *Principal/Test Coordinator Manual* included checklists to help each school prepare for testing. The checklists outlined tasks for the schools to perform before, during, and after test administration. Along with these checklists, the *Principal/Test Coordinator Manual* outlined the nature of the testing materials being sent to each school and discussed test security issues, such as how to inventory materials and track them at all times during administration and how to return the materials once testing was completed. The *Test Administrator Manual* and *Online Test Administration Manual* also included checklists to help administrators prepare themselves, their classrooms, and their students for the administration of the test. Both of these manuals contained sections that detailed the procedures to be followed for each test session. The *Principal/Test Coordinator Manual* and *Test Administrator Manual* gave instructions on preparing the material for its return to Measured Progress. Principals were required to complete an online Principal Certification of Proper Administration form at the conclusion of testing, certifying that all testing was administered according to MEA protocols.

2.3 PARTICIPATION REQUIREMENTS AND DOCUMENTATION

The intent is for *all* students in grades 5 and 8 to participate in the MEA Science assessment through standard administration, administration with accommodations, and/or alternate assessment. Any student who is absent during any session of the MEA is expected to take a makeup test within the testing window.

On those occasions where it was deemed necessary to exclude a student from sections of the assessment or from the assessment as a whole because of special considerations (e.g., hospitalization or a death in the family), schools were asked to seek the approval of the Maine Department of Education's (MDOE's) MEA review team. The names of the excluded students were forwarded to Measured Progress so these students would not be included in any reports or as part of the denominator representing the total number of students. Appendix C presents student participation in the MEA Science for all students by demographic group.

2.4 ADMINISTRATOR TRAINING

In addition to distributing the *Principal/Test Coordinator Manual*, the *Test Administrator Manual*, and the *Online Test Administration Manual*, the MDOE, along with Measured Progress, broadcasted a statewide test administration workshop to train and inform school personnel about the MEA testing procedures. This training was posted on the MDOE website at http://www.maine.gov/doe/mea/administration/index.html.

Three recorded training modules were also provided for the online test administration training for each of the three levels of administrators— technology coordinators, principal/test coordinators, and test administrators. Additional training resources for online testing included a Quick Start Guide, a Quick Reference Guide, a User Manual for the iTest System, a Student Training Module, and a Proctor Test Workstation (PTW)/Student Test Workstation (STW) Interaction video. The online trainings and materials were posted on both the Measured Progress website, http://iservices.measuredprogress.org, and the MDOE website at http://www.maine.gov/doe/mea/administration/index.html.

2.5 DOCUMENTATION OF ACCOMMODATIONS

The approved accommodations for eligible students were listed on Page 2 of the student answer booklet. This information was coded in by the appropriate staff after testing was completed. The *Principal/Test Coordinator Manual* and the *Test Administrator Manual* provided directions for coding the information related to accommodations.

All students who were considered for accommodations on the MEA should have had their individual situations reviewed by a team within the school prior to the time of testing. For every student with an identified exceptionality requiring an individual education plan (IEP), schools were required to hold an IEP team meeting that addressed that student's needs for accommodations. For other students needing test accommodations who did not have an identified exceptionality, a meeting was required that included one of the student's teachers, the building principal, related services personnel, and, whenever possible, the student's parents. If it was not possible for the parents to attend the meeting, they were notified of the committee's recommendations for accommodations prior to the time of testing.

Recommended accommodations were to be consistent with those accommodations already being used in the student's instructional program. Any such accommodations were reflected either in the minutes of the IEP team meeting (for students requiring an IEP) or in a statement prepared for the cumulative folders of students not requiring IEPs. Schools were given the following statement as a "model": *The student will participate in the [__]th grade Maine Educational Assessment as scheduled during the month of May 2014 with the following accommodations*.

Table 2-1. 2013–14 MEA Science: Numbers of Students Tested With and Without Accommodations by Subject and Grade

Subject	*Grade*	*Number of Students Tested*	
		Without Accommodations	*With Accommodations*
Science	5	10,514	2,466
	8	11,692	1,781

Appendix D also shows the accommodation frequencies observed for the May 2014 MEA administration. The MDOE *Accommodations Guide*, which includes detailed descriptions of approved accommodations and their proper application, is presented in Appendix E.

During the online version of the test, all students had the following tools available to them for multiple-choice questions: increase and decrease font size, change background color, change font color, highlight text, cross out text, erase highlighted and crossed-out text. For constructed-response questions, typed text could be cut, copied, pasted, bulleted, bolded, italicized, underlined, enlarged, minimized, or indented.

2.6 TEST SECURITY

Maintaining test security is critical to the success of the MEA. The *Principal/Test Coordinator Manual*, the *Test Administrator Manual*, and the *Online Test Administration Manual* explain in detail all test security measures and test administration procedures. School personnel were informed that any concerns about breaches in test security were to be reported to the school's test coordinator and/or principal immediately. The test coordinator and/or principal was responsible for immediately reporting the concern to the district superintendent and the state assessment director at the Department of Education. Test security was also strongly emphasized at the test administration workshops. Principals were required to log on to a secure website to complete the Principal's Certification of Proper Test Administration form for each grade level tested at their school; they also had to provide the number of students tested online, the number of secure tests received from Measured Progress, the number of paper-and-pencil tests administered to students, and the number of secure test materials that they were returning to Measured Progress. Principals were instructed to submit the form by entering a unique password, which acted as their digital signature. By signing and submitting the form, the principal certified that the tests were administered according to the test

administration procedures outlined in the *Principal/Test Coordinator Manual*, the *Test Administrator Manual*, and the *Online Test Administration Manual*; that the security of the tests was maintained; that no secure material was duplicated or in any way retained in the school; and that all test materials had been accounted for and returned to Measured Progress.

2.7 TEST AND ADMINISTRATION IRREGULARITIES

There were no test irregularities in the May 2014 administration.

2.8 TEST ADMINISTRATION WINDOW

The test administration window was May 5–16, 2014.

2.9 SERVICE CENTERS

To provide additional support to schools before, during, and after testing, Measured Progress established the MeCAS Service Center and the Measured Progress Technology Product Support HelpDesk. The support of these service centers is essential to the successful administration of any statewide test program. These service centers provide a centralized location that individuals in the field can call using a toll-free number to ask specific questions or report any problems they may be experiencing. Representatives are responsible for receiving, responding to and tracking calls, and then routing issues to the appropriate person(s) for resolution. All calls are logged into a database that includes notes regarding the issue and resolution of each call.

The MeCAS Service Center was open to receive calls from 7:30 a.m. to 4:30 p.m., Monday through Friday, beginning two weeks before the start of testing and ending two weeks after the conclusion of testing. The Measured Progress Technology Product Support HelpDesk was available from 8:00 a.m. to 5:00 p.m., Monday through Friday to answer technical questions related to online testing.

CHAPTER 3 TEST DESIGN AND DEVELOPMENT

3.1 TEST SPECIFICATIONS

3.1.1 Criterion-Referenced Test

The MEA is a criterion-referenced science test. Items on the MEA are developed specifically for Maine and are directly linked to Maine's science content standards. These content standards are the basis for the reporting categories and are used to help guide the development of test items.

3.1.2 Item Types

Maine educators and students are familiar with the types of items used in the assessment program. The types of items and their functions are described below:

- **Multiple-choice (MC)** items are used to provide breadth of coverage within a content area. Because they require no more than a minute for most students to answer, MC items make efficient use of limited testing time and allow for coverage of a wide range of knowledge and skills.
- **Constructed-response (CR)** items typically require students to use higher-order thinking skills—evaluation, analysis, summarization, and so on—to construct satisfactory responses. CR items take most students approximately 5 to 10 minutes to complete.

Note that the use of released MEA items to prepare students to respond to multiple-choice and constructed-response items is appropriate and encouraged.

3.1.3 Description of Test Design

The MEA is structured using both *common* and *field-test* items. Common items are taken by all students in a given grade level. Student scores are based only on common items. Field-test items are divided among the forms of the test for each grade level. Each student takes only one form of the test and therefore answers a fraction of the field-test items. Field-test items are not identifiable to test takers and have a negligible impact on testing time. Because all students participate in the field test, it provides the minimum sample size (750–1,500 students per item) needed to produce reliable data that can be used to inform item selection for future tests.

3.2 SCIENCE TEST SPECIFICATIONS

3.2.1 Standards

The 2013–14 MEA Science test items are aligned to the content standards of "D: The Physical Setting" and "E: The Living Environment," as described in the Science and Technology section of Maine's *Learning Results: Parameters for Essential Instruction*. No other science content standards are subject to statewide assessment. Content specialists use the content standards, performance indicators, and descriptors to help guide the development of test questions, which may address one or more of the performance indicators listed below.

D. The Physical Setting

- D1: Universe and Solar System—Students explain the physical formation and changing nature of our universe and solar system, and how our past and present knowledge of the universe and solar system developed.
- D2: Earth—Students describe and analyze the biological, physical, energy, and human influences that shape and alter Earth systems.
- D3: Matter and Energy—Students describe the structure, behavior, and interaction of matter at the atomic level and the relationship between matter and energy.
- D4: Force and Motion—Students understand that the laws of force and motion are the same across the universe.

E. The Living Environment

- E1: Biodiversity—Students describe and analyze the evidence for relatedness among and within diverse populations of organisms and the importance of biodiversity.
- E2: Ecosystem—Students describe and analyze the interactions, cycles, and factors that affect short-term and long-term ecosystem stability and change.
- E3: Cells—Students describe the structure and function of cells at the intracellular and molecular levels, including differentiation to form systems, interactions between cells and their environment, and the impact of cellular processes and changes on individuals.
- E4: Heredity and Reproduction—Students examine the role of DNA in transferring traits from generation to generation, in differentiating cells, and in evolving new species.
- E5: Evolution—Students describe the interactions between and among species, populations, and environments that lead to natural selections and evolution.

3.2.2 Item Types

The science test includes multiple-choice and constructed-response items. Each multiple-choice item requires students to select the correct response from four choices. Each type of item is worth a specific number of points in the student's total science score, as shown in Table 3-1.

Table 3-1. 2013–14 MEA Science: Item Types

Item Type	*Possible Score Points*
MC	0 or 1
CR	0, 1, 2, 3, or 4

MC = multiple-choice
CR = constructed-response

Consistent with the annual release policy, 50% of the items were released from the 2012–13 MEA Science test. A practice test composed of released science items is available on the MDOE website at http://www.maine.gov/doe/mea/resources/index.html. Schools are encouraged to incorporate the use of the released items in their instructional activities so that students will be familiar with them.

3.2.3 Test Design

Table 3-2 summarizes the numbers and types of items that were used to compute student scores on the 2013–14 MEA Science test. Additionally, each test form had eight multiple-choice field-test items and one constructed-response field-test item that did not affect student scores.

Table 3-2. 2013–14 MEA Science: Science Items

Grade	*Session 1*	*Session 2*	*Session 3*	*TOTAL*	
				MC	*CR*
5	8 MC, 2 CR	6 MC, 1 CR	18 MC, 1 CR	32	4
8	11 MC, 2 CR	11 MC, 1 CR	18 MC, 1 CR	40	4

MC = multiple-choice; CR = constructed-response

3.2.4 Blueprints

Table 3-3 shows the distribution of points across the science standards. For grade 5, D1–D2 had one constructed-response item; D3–D4 had one constructed-response item; and E1–E5 had two constructed-response items. For grade 8, D1–D2 had one constructed-response item; D3–D4 had one constructed-response item; and E1–E5 had two constructed-response items.

Table 3-3. 2013–14 MEA Science: Distribution of Score Points

Science Standards	*Grade 5*	*Grade 8*
D1-D2 (Earth & Space)	12	17
D3-D4 (Physical)	12	17
E1-E5 (Life)	24	22
Total Score Points	48	56

3.2.5 Depth of Knowledge

Each item on the MEA Science test is assigned a depth of knowledge (DOK) level. The DOK level reflects the complexity of mental processing students use to answer an item. DOK is not synonymous with difficulty. Each of the four DOK levels is described below.

- **Level 1 (Recall).** This level requires the recall of information such as a fact, definition, term, or simple procedure. These items require students only to demonstrate a rote response, use a well-known formula, or follow a set procedure.
- **Level 2 (Skill/Concept).** This level requires mental processing beyond that of recalling or reproducing a response. These items require students to make some decisions about how to approach the item.
- **Level 3 (Strategic Thinking).** This level requires reasoning, planning, and using evidence. These items require students to handle more complexity and abstraction than items at the previous two levels.
- **Level 4 (Extended Thinking).** This level requires planning, investigating, and complex reasoning over an extended period of time. Students are required to make several connections within and across content areas. This level may require students to design and conduct experiments. Because of the nature of this level, there are no Level 4 items on the MEA.

It is important that the MEA measures a range of depths of knowledge. Table 3-4 shows the distribution of score points across the DOK levels used on the MEA.

Table 3-4. 2013–14 MEA Science: Distribution of Score Points Across Depth of Knowledge (DOK)

DOK Level	*Grade 5*	*Grade 8*
1	8	10
2	30	39
3	10	7
Total	48	56

3.2.6 Use of Calculators and Reference Sheets

Calculators are not used or needed when taking the science tests. There are no science reference sheets.

3.3 TEST DEVELOPMENT PROCESS

3.3.1 Item Development

Items used on the MEA are developed and customized specifically for use on the MEA and are consistent with Maine content standards and performance indicators. Measured Progress test developers work with the Maine state science specialist and with Maine educators to verify the alignment of items to the appropriate Maine content standards.

The development process combined the expertise of Measured Progress test developers, the Maine state science specialist, and committees of Maine educators to help ensure that items met the needs of the MEA program. All items used on the common portions of the MEA were reviewed by a committee of Maine content experts, a committee of Maine bias experts, and three external content experts.

3.3.2 Item Reviews at Measured Progress

The test developers at Measured Progress reviewed newly developed items for

- alignment to the intended content standard;
- item integrity, including science content and structure, format, clarity, possible ambiguity, and single correct answer;
- appropriateness and quality of graphics;
- appropriateness of scoring guide descriptions and distinctions;
- completeness of associated item documentation (e.g., scoring guide, content codes, key, grade level, depth of knowledge, and contract identified); and
- appropriateness for the designated grade level.

3.3.3 Item Reviews at State Level

A committee of Maine classroom teachers from across the state reviewed the items before field testing. Teacher participants were selected based on their content-area expertise and grade-level familiarity. The purpose of the review was to evaluate new items for the embedded field test and determine their suitability for the assessment by answering the following four questions:

- Does the item align with the assigned content standard and performance indicator?
- Is the science content accurate?
- Is the science content grade-level appropriate?
- Does the item provide maximum accessibility for all students?

3.3.4 Bias and Sensitivity Review

Bias review is an essential component of the development process. During the bias review process, items were reviewed by a committee of Maine educators who represented various student subgroups, including the visually or hearing impaired and students for whom English is a second language. Items were examined for content that might cause the test to be inaccessible for these students or that might generally offend or dismay students, teachers, or parents. Being aware of these considerations in the development of assessment items and materials can avoid many unduly controversial issues, and unfounded concerns can be allayed before the test forms are produced.

3.3.5 External Expert Review

The test items were classified into three groups based on science content. Three science experts (one in earth/space science, one in life science, one in physical science) reviewed the group of items corresponding to their area of expertise. The expert reviewers primarily evaluated each item for correct science content. For the MC items, the experts also indicated whether the keyed answer was correct and whether it was the only correct answer among the options given. The MDOE state science specialist and Measured Progress test developers reviewed the experts' evaluations and made appropriate adjustments to the items as necessary.

3.3.6 Reviewing and Refining

Recommended changes from the Item Review and Bias and Sensitivity meetings, as well as the comments from the three external science experts, were reviewed and considered by the Maine state science specialist. Measured Progress test developers made the edits that were approved by the Maine state science specialist.

3.3.7 Item Editing

Measured Progress editors reviewed and edited the items to ensure adherence to sound testing principles and to style guidelines in the *Chicago Manual of Style*, 16th edition. These principles include the stipulations that items

- demonstrate correct grammar, punctuation, usage, and spelling;
- are written in a clear, concise style;
- contain unambiguous explanations that tell students what is required to attain a maximum score;
- are written at a reading level that allows students to demonstrate their knowledge of the subject matter being tested regardless of reading ability;

- exhibit high technical quality regarding psychometric characteristics;
- have appropriate answer options or score point descriptors; and
- are free of potentially insensitive content.

3.3.8 Item Selection and Operational Test Assembly

Measured Progress test developers met with the Maine state science specialist to select the common items. In preparation for the meeting, the test developers and psychometricians at Measured Progress considered the following in selecting sets of items to propose for the common test:

- **Content coverage/match to test design and blueprints.** The test designs and blueprints stipulate a specific number of multiple-choice and constructed-response items. Item selection for the embedded field test was based on the number of items in the existing pool of items that are eligible for the common test.
- **Item difficulty and complexity.** Item statistics drawn from the data analysis of previously field-tested items were used to ensure quality psychometric characteristics as well as similar levels of difficulty and complexity from year to year.
- **"Cueing" items.** Items were reviewed for any information that might "cue" or provide information that would help to answer another item.

At the meeting, the Maine state science specialist reviewed the proposed sets of items and made the final selection of items for the common test.

The test developers then sorted and laid out the items into test forms. During assembly of the test forms, the following criteria were considered:

- **Key patterns.** The sequence of keys (correct answers) was reviewed to ensure that their order appeared random.
- **Option balance.** Items were balanced across forms so that each form contained a roughly equivalent number of key options (As, Bs, Cs, and Ds).
- **Page fit.** Item placement was modified to ensure the best fit and arrangement of items on any given page.
- **Relationships among forms.** Although field-test items differ from form to form, these items must take up the same number of pages in all forms so that sessions begin on the same page in every form. Therefore, the number of pages needed for the longest form often determines the layout of each form.
- **Visual appeal.** The visual accessibility of each page of the form was always taken into consideration, including such aspects as the amount of "white space," the density of the test, and the number of graphics.

3.3.9 Operational Test Draft Review

After the forms were laid out as they would appear in the final test booklets, the forms were again thoroughly reviewed by Measured Progress editors to ensure that the items appeared exactly as intended. Any changes made during test construction were reviewed and approved by the test developer. The Maine state science specialist then read the forms for final approval.

3.3.10 Alternative Presentations

The common test for each grade was translated into Braille by the American Printing House for the Blind, a subcontractor that specializes in test materials for blind and visually impaired students. In addition, Form 1 for each grade was adapted into a large-print version.

CHAPTER 4 SCORING

4.1 MACHINE-SCORED ITEMS

Multiple choice item responses were compared to scoring keys using item analysis software. Correct answers were assigned a score of one point and incorrect answers were assigned zero points. Student responses with multiple marks and blank responses were also assigned zero points.

The hardware elements of the scanners monitor themselves continuously for correct data reads, as does the software that drives these scanners. Standard checks include recognition of a sheet that does not belong or is upside down or backwards and identification of critical data that are missing (e.g., a student ID number), test forms that are out of range or missing, and page or document sequence errors. When a problem is detected, the scanner stops and displays an error message directing the operator to investigate and correct the situation.

4.2 PERSON-SCORED ITEMS

The images of student responses to constructed-response items were hand-scored through Measured Progress's electronic scoring system, iScore. Use of iScore minimized the need for readers to physically handle answer booklets and related scoring materials. Student confidentiality was easily maintained, since all MEA scoring was "blind" (i.e., district, school, and student names were not visible to readers). The iScore system maintained the linkage between the student response images and their associated test booklet numbers.

Through iScore, qualified readers at computer terminals accessed electronically scanned images of student responses. Readers evaluated each response and recorded each score via keypad or mouse entry through the iScore system. When a reader finished one response, the next response appeared immediately on the computer screen.

Imaged responses from all answer booklets were sorted into item-specific groups for scoring purposes. Readers reviewed responses from only one item at a time; however, imaged responses from a student's entire booklet were always available for viewing when necessary, and the physical booklet was also available to the chief reader onsite. (Chief reader and other scoring roles are described in the section that follows.)

The use of iScore also helped ensure that access to student response images was limited to only those who were scoring or working for Measured Progress in a scoring management capacity.

4.2.1 Scoring Location and Staff

The iScore database, its operation, and its administrative controls are all based in Dover, New Hampshire. Table 4-1 presents the locations where 2013–14 MEA test item responses by grade were scored.

Table 4-1. 2013–14 MEA Science: Operational Scoring Locations by Grade

Grade	Dover, NH	Menands, NY	Longmont, CO
5	X		
8	X		

The iScore system monitored accuracy, reliability, and consistency across. Constant daily communication and coordination were accomplished through email, telephone, faxes, and secure websites to ensure that critical information and scoring modifications were shared and implemented across the scoring site.

Staff Positions

The following staff members were involved with scoring the 2013–14 MEA responses:

- The MEA scoring project manager, an employee of Measured Progress, was located in Dover, New Hampshire, and oversaw communication and coordination of scoring.
- The iScore operational manager and iScore administrators, employees of Measured Progress, were located in Dover, New Hampshire, and coordinated technical communication.
- A chief reader in the science content area ensured consistency of scoring for all grades tested in that content area. Chief readers also provided read-behind activities (defined in a later section) for quality assurance coordinators. Chief readers are employees of Measured Progress.
- Quality assurance coordinators (QACs), selected from a pool of experienced senior readers for their ability to score accurately and their ability to instruct and train readers, participated in benchmarking activities for each specific grade of the science content area. QACs provided read-behind activities (defined in a later section) for senior readers at the scoring site. The ratio of QACs and senior readers to readers was approximately 1:11.
- Senior readers (SRs), selected from a pool of skilled and experienced readers, provided read-behind activities (defined in a later section) for the readers at their scoring tables (2 to 12 readers at each table). The ratio of QACs and SRs to readers was approximately 1:11.
- Readers at the Dover, New Hampshire, scoring site scored operational and field test MEA 2013–14 student responses. Recruitment of readers is described in Section 4.2.3.

4.2.2 Benchmarking Meetings

In preparation for implementing MEA scoring guidelines, Measured Progress scoring staff prepared and facilitated benchmarking meetings held with the MEA state science specialist representing the department of education. The purpose of these meetings was to establish guidelines for scoring MEA items during the current field-test scoring session and for future operational scoring sessions.

Chief readers selected several dozen student responses for each item that were identified as illustrative midrange examples of the respective score points. Chief readers presented these responses to the MEA science content specialist during benchmarking meetings and worked collaboratively with them to finalize an authoritative set of score point exemplars for each field-test item. As a matter of practice, these sets are included in the scoring training materials each time an item is administered.

This repeated use of MEA-approved sets of midrange score point exemplars helps ensure that readers follow established guidelines each time a particular MEA item is scored.

4.2.3 Reader Recruitment and Qualifications

For scoring the 2013–14 MEA, Measured Progress actively sought a diverse scoring pool. The broad range of reader backgrounds typically includes scientists, editors, business professionals, authors, teachers, graduate school students, and retired educators. Demographic information about readers (e.g., gender, race, educational background) was electronically captured for reporting.

Readers were required to have successfully attained a four-year college degree or higher. In all cases, potential readers were required to submit documentation (e.g., resume and/or transcripts) of their qualifications.

Table 4-2 summarizes the qualifications of the 2013–14 MEA scoring leadership and readers.

Table 4-2. 2013–14 MEA Science: Qualifications of Scoring Leadership and Readers—Spring Administration

Scoring Responsibility	*Educational Credentials*				*Total*
	Doctorate	Master's	Bachelor's	Other	
Scoring Leadership	0.0%	43.8%	56.2%	0.0%	100.0%
Readers	4.2%	32.6%	63.2%	0.0%	100.0%

Scoring Leadership = chief readers, quality assurance coordinators, and senior readers

Readers were either temporary Measured Progress employees or were secured through temporary employment agencies. All readers were required to sign a nondisclosure/confidentiality agreement.

4.2.4 Methodology for Scoring Polytomous Items

Possible Score Points

The ranges of possible score points for the different polytomous items are shown in Table 4-3.

Table 4-3. 2013–14 MEA Science: Possible Score Points for Polytomous Types

Polytomous Item Type	*Possible Score Point Range*
Constructed-response	0–4
Nonscorable items	0

Nonscorable Items

Readers could designate a response as nonscorable for any of the following reasons:

- response was blank (no attempt to respond to the question)
- response was unreadable (illegible, too faint to see, or only partially legible/visible)—see note below
- response was written in the wrong location (seemed to be a legitimate answer to a different question)—see note below

Note: "Unreadable" and "wrong location" responses were eventually resolved by researching the actual answer document (electronic copy or hard copy, as needed) to identify the correct location (in the answer document) or to more closely examine the response and then assign a score.

Scoring Procedures

Scoring procedures for polytomous items included both single scoring and double-blind scoring. Single scored items were scored by one reader. Double-blind scored items were scored independently by two readers whose scores were tracked for "interrater agreement" (for further discussion of double-blind scoring and interrater agreement, see Section 4.2.7 and Appendix S).

4.2.5 Reader Training

Reader training began with an introduction of the onsite scoring staff and an overview of the MEA program's purpose and goals (including discussion about the security, confidentiality, and proprietary nature of testing materials, scoring materials, and procedures).

Next, readers thoroughly reviewed and discussed the scoring guide for each item to be scored. Each item-specific scoring guide included the item itself and score point descriptions.

Following review of an item's scoring guide, readers reviewed the particular response set organized for that training: Anchor Sets, Training Sets, and Qualifying Sets. (These are defined below.)

During training, readers could highlight or mark hard copies of the Anchor and Training Sets (as well as first Qualifying Sets after the qualification round), even if all or part of the set was also presented online via computer.

Anchor Set

Readers first reviewed an Anchor Set of exemplary responses for an item. This is a set approved by the science content specialist representing the Maine Department of Education. Responses in Anchor Sets are typical, rather than unusual or uncommon; solid, rather than controversial or borderline; and true, meaning that they had scores that could not be changed by anyone other than the MEA client and Measured Progress scoring services staff. Each Anchor Set contains one client-approved sample response per score point, which is considered to be a midrange exemplar. Each sample response has, where necessary, the MEA science content specialist's rationale for choosing that response as a score point anchor. The set includes a second sample response if there is more than one plausible way to illustrate the merits and intent of a score point.

Responses were read aloud to the room of readers in descending score order. Announcing the true score of each anchor response, trainers facilitated group discussion of responses in relation to score point descriptions to help readers internalize the typical characteristics of score points.

This Anchor Set continued to serve as a reference for readers as they went on to calibration, scoring, and recalibration activities for that item.

Training Set

Next, readers practiced applying the scoring guide and anchors to responses in the Training Set. The Training Set typically included 10 to 15 student responses designed to help establish both the full score point range and the range of possible responses within each score point. The Training Set often included unusual responses that were less clear or solid (e.g., shorter than normal, employing atypical approaches, simultaneously containing very low and very high attributes, and written in ways difficult to decipher). Responses in the Training Set were presented in randomized score point order.

After readers independently read and scored a Training Set response, trainers would poll readers or use online training system reports to record the initial range of scores. Trainers then led group discussion of one or two responses, directing reader attention to difficult scoring issues (e.g., the borderline between two score points). Throughout the process, trainers modeled how to discuss scores by referring to the Anchor Set and to scoring guides.

Qualifying Set

After the Training Set was completed, readers were required to score responses accurately and reliably in Qualifying Sets assembled for constructed-response items. The 10 responses in each Qualifying Set were selected from an array of responses that clearly illustrated the range of score points for that item as

reviewed and approved by the state specialist. Hard copies of the responses were also made available to readers after the qualification round so that they could make notes and refer back during the post-qualifying discussion.

To be eligible to live score one of the above items, readers were required to demonstrate scoring accuracy rates of at least 80% exact agreement (i.e., to exactly match the predetermined score on at least 8 of the 10 responses) and at least 90% exact or adjacent agreement (i.e., to exactly match or be within one score point of the predetermined score on 9 or 10 of the 10 responses). In other words, readers were allowed one discrepant score (i.e., 1 score of 10 that was more than one score point from the predetermined score) provided they had at least eight exact scores.

Retraining

Readers who did not pass the first Qualifying Set were retrained as a group by reviewing their performance with scoring leadership and then scoring a second Qualifying Set of responses. If they achieved the required accuracy rate on the second Qualifying Set, they were allowed to score operational responses.

Readers who did not achieve the required scoring accuracy rates on the second Qualifying Set were not allowed to score responses for that item. Instead, they either began training on a different item or were dismissed from scoring for that day.

4.2.6 Leadership Training

QACs and select SRs were trained in a separate training session immediately prior to reader training. In addition to discussing the items and their responses, QAC and SR training included greater detail on the client's rationale behind the score points than that covered with regular readers in order to better equip QACs and SRs to handle questions from the latter.

4.2.7 Monitoring of Scoring Quality Control

Readers were monitored for continued accuracy and consistency throughout the scoring process using the following methods and tools (which are defined in this section):

- Embedded Committee-Reviewed Responses (CRRs)
- Read-Behind Procedures
- Double-Blind Scoring
- Recalibration Sets
- Scoring Reports

It should be noted that any reader whose accuracy rate fell below the expected rate for a particular item and monitoring method was retrained on that item. Upon approval by the QAC or chief reader as

appropriate (see below), the reader was allowed to resume scoring. Readers who met or exceeded the expected accuracy rates continued scoring.

Furthermore, the accuracy rate required of a reader to qualify to score live was higher than that required to continue to score responses live. The reason for the difference is that an "exact score" in double-blind scoring requires that two readers choose the same score for potentially borderline responses (in other words, an exact score is dependent upon two peers agreeing on responses that often do not sit neatly in the middle of their score point spectrum), whereas an "exact score" in qualification requires only that a single reader match a score pre-established as sitting in the middle of their respective score point by scoring leadership. The use of multiple monitoring techniques is critical toward monitoring reader accuracy during the process of live scoring.

Embedded Committee-Reviewed Responses (CRRs)

Committee-Reviewed Responses (CRRs) are previously scored responses that are loaded ("embedded") by scoring leadership into iScore and distributed "blindly" to readers during scoring. Embedded CRRs may be chosen either before or during scoring and are inserted into the scoring queue so that they appear the same as all other live student responses.

Between 5 and 30 embedded CRRs were distributed at random points throughout the first full day of scoring to ensure that readers were sufficiently calibrated at the beginning of the scoring period. Individual readers often received up to 20 embedded CRRs within the first 100 responses scored and up to 10 additional responses within the next 100 responses scored on that first day.

Any reader who fell below the required scoring accuracy rate was retrained before being allowed by the QAC to continue scoring. Once allowed to resume scoring, scoring leadership carefully monitored these readers by increasing the number of read-behinds (defined in the next section).

Embedded CRRs were employed for all constructed-response items.

Read-Behind Procedures

Read-behind scoring refers to scoring leadership (usually a SR) scoring a response after a reader has already scored the response. The practice was applied to all constructed-response item types.

Responses placed into the read-behind queue were randomly selected by scoring leadership; readers were not aware which of their responses would be reviewed by their SR. The iScore system allowed one, two, or three responses per reader to be placed into the read-behind queue at a time.

The SR entered his or her score into iScore before being allowed to see the reader's score. The SR then compared the two scores and the score of record (i.e., the reported score) was determined as follows:

- If there was exact agreement between the scores, no action was necessary; the regular reader's score remained.
- If the scores were adjacent (i.e., differed by one point), the SR's score became the score of record. (A significant number of adjacent scores for a reader triggered an individual scoring consultation with the SR, after which the QAC determined whether or when the reader could resume scoring.)
- If the scores were discrepant (i.e., differed by more than one point), the SR's score became the score of record. (This triggered an individual consultation with the SR, after which the QAC determined whether or when the reader could resume scoring on that item.)

Table 4-4 illustrates how scores were resolved by read-behind.

Table 4-4. 2013–14 MEA Science: Examples of Read-Behind Scoring Resolutions

Reader Score	*QAC/SR Score*	*Score of Record*
4	4	4
4	3	3*
4	2	2*

* QAC/SR's score.

SRs were tasked with conducting, on average, five read-behinds per reader throughout each half scoring day; however, senior readers conducted a proportionally greater number of read-behinds for readers who seemed to be struggling to maintain, or who fell below, accuracy standards.

In addition to regular read-behinds, scoring leadership could choose to do read-behinds on any reader at any point during the scoring process to gain an immediate, real time "snapshot" of a reader's accuracy.

Double-Blind Scoring

Double-blind scoring refers to two readers independently scoring a response without knowing whether the response was to be double-blind scored. The practice was applied to all constructed-response item types. Table 4-5 shows by which method(s) both common and equating constructed-response item responses for each operational test were scored.

Table 4-5. 2013–14 MEA Science: Frequency of Double-Blind Scoring by Grade

Grade	*Content Area*	*Responses Double-Blind Scored*
5, 8	Science	10%
5, 8	Unreadable responses	100%
5, 8	Blank responses	100%

If there was a discrepancy (a difference greater than one score point) between double-blind scores, the response was placed into an arbitration queue. Arbitration responses were reviewed by scoring leadership (SR

or QAC) without knowledge of the two readers' scores. Scoring leadership assigned the final score. Appendix S provides the MEA 2013–14 percentages of agreement between readers for each common item for each grade.

Scoring leadership consulted individually with any reader whose scoring rate fell below the required accuracy rate, and the QAC determined whether or when the reader could resume scoring on that item. Once the reader was allowed to resume scoring, scoring leadership carefully monitored the reader's accuracy by increasing the number of read-behinds.

Recalibration Sets

To determine whether readers were still calibrated to the scoring standard, readers were required to take an online Recalibration Set at the start and midpoint of the shift upon their resumption of scoring (daytime shifts are typically 7.5 hours and evening shifts 5.5 hours in duration).

Each Recalibration Set consisted of five responses representing the entire range of possible scores, including some with a score point of 0.

- Readers who were discrepant on two of five responses of the first Recalibration Set, or were exact on two or fewer, were not permitted to score on that item that day and were either assigned to a different item or dismissed for the day.
- Readers who were discrepant on only one of five responses of the first Recalibration Set, and/or exact on three, were retrained by their SR by discussing the Recalibration Set responses in terms of the score point descriptions and the original Anchor Set. After this retraining, such readers began scoring operational responses under the proviso that the reader's scores for that day and that item would be kept only if the reader was exact on all five of five responses of the second Recalibration Set administered at the shift midpoint. The QAC determined whether or when these readers had received enough retraining to resume scoring operational responses. Scoring leadership also carefully monitored the accuracy of such readers by significantly increasing the number of their read-behinds.
- Readers who were not discrepant on any response of the first Recalibration Set, and exact on at least four, were allowed to begin scoring operational responses immediately, under the proviso that this Recalibration performance would be combined with that of the second Recalibration Set administered at the shift midpoint.

The results of both Recalibration Sets were combined with the expectation that readers would have achieved an overall 80% exact and 90% adjacent standard for that item for that day.

The scoring project manager voided all scores posted on that item for that day by readers who did not meet the accuracy requirement. Responses associated with voided scores were reset and redistributed to readers with demonstrated accuracy for that item.

Recalibration Sets were employed for all constructed-response items and were first administered at the start of the second day of scoring on an item since the first day of scoring an item is monitored using the item's initial qualification set and set of Embedded CRRs. In the event an item was scored during a third day,

newly assembled Recalibration Sets were administered similarly to how the sets were administered on the second day.

Scoring Reports

Measured Progress's electronic scoring software, iScore, generated multiple reports that were used by scoring leadership to measure and monitor readers for scoring accuracy, consistency, and productivity. These reports are further discussed in the following section.

4.2.7.1 Reports Generated During Scoring

Because of the complexity of scoring a large-scale assessment project such as that for MEA, computer-generated reports were necessary to ensure that:

- overall group-level accuracy, consistency, and reliability of scoring were maintained at acceptable levels
- immediate, real-time individual reader data were available to allow early intervention when necessary
- scoring schedules were maintained

The following reports were produced by iScore for internal use throughout each scoring day by scoring leadership (including Senior Readers, QACs, Chief Readers, and the Scoring Project Manager, where applicable):

- **The Read-Behind Summary** showed the total number of read-behind responses for each reader and noted the number and percentages of exact, adjacent, and discrepant scores with the SR/QAC. Scoring leadership could choose to generate this report by choosing options (such as "Today," "Past Week," and "Cumulative") from a pull-down menu. The report could also be filtered to select data for a particular item or across all items. This report was used in conjunction with other reports to determine whether a reader's scores would be voided (i.e., sent back out to the floor to be rescored by other readers). The benefit of this report is that it can reveal the degree to which an individual reader agrees with their QAC or SR on how best to score live responses.
- **The Double-Blind Summary** showed the total number of double-scored responses of each reader and noted the number and percentages of exact, adjacent, and discrepant scores with second readers. This report was used in conjunction with other reports to determine whether a reader's scores should be voided (i.e., sent back out to the floor to be rescored by other readers). The benefit of this report is that it can reveal the degree to which readers are in agreement with each other about how best to score live responses.
- **The Accuracy Summary** combined read-behind and double-blind data, showing the total number for the readers, their accuracy rates, and their score point distributions.
- **The Embedded CRR Summary** showed, for each reader (by item or across all items), the total number of responses scored, the number of embedded CRRs scored, and the numbers and percentages of exact, adjacent, and discrepant scores with the chief reader. This report

was used in conjunction with other reports to determine whether a reader's scores should be voided (i.e., sent back out to the floor to be rescored by other readers). The benefit of this report is that it can reveal the degree to which an individual reader agrees with his or her chief reader on how to best score live responses. Also, since embedded CRRs are administered during the first hours of scoring, this report can provide an early illustration of agreement between readers and chief readers.

- **The Qualification Statistics Summary** listed each reader by name and ID number, identified which Qualifying Set(s) they did and did not take and, for the ones taken, their pass rate. In addition to the pass rates of individuals, the report also showed numbers of readers passing or failing a particular Qualifying Set. The QAC could use this report to determine how readers within their scoring group performed on specific Qualifying Sets.
- **The Summary Statistics Report** showed the total number of student responses for an item, and identified, for the time at which the report was generated, the following:
  - the number of single and double-blind scorings that had been performed; and
  - the number of single and double-blind scorings yet to be performed.

CHAPTER 5 SCORE REPORTING

5.1 PRIMARY REPORTS

Measured Progress created the following primary reports for MEA Science:

- Individual Student Report (ISR)
- Student score label
- School, SAU, and State Grade Level Summary Report
- SAU All Grades Summary Report
- Interactive reporting

ISRs, Grade Level Summary Reports, and All Grades Summary Reports were posted online via a secure website on August 14, 2014. Interactive reporting was available online via a secure website on August 14, 2014. ISRs were printed and shipped to schools on September 5, 2014 for distribution to parents and guardians. Student score labels were also shipped on this date for schools to keep with student records. Sample reports are included in Appendices F and G.

5.2 INDIVIDUAL STUDENT REPORT AND STUDENT LABELS

An ISR was sent to the student's school to be given to parent(s)/guardians(s). The report was also posted online via a secure website for school, SAU, and state personnel. The front cover contained a letter from the Maine commissioner of education, an explanation of what the MEA is, who should participate, what the test looks like, how the test was developed, and how results should be used. Science results included an achievement level, a scaled score, and an error estimate on the scaled score. Student answers were provided for released items. The report also included a sample released multiple-choice item with suggestions for how parents can access more information on the released items. A graphic that compared the student's performance with school, SAU, and state results was incorporated in the ISR. Also included is a graphic displaying the student's performance compared to how Proficient students performed. Suggestions for helping students in science were provided. A student label with the student's achievement level was also provided to schools for placement in the student's file.

Schools are instructed to keep all student data, including copies of the ISR and student score labels, secure within the school and SAU. The Family Educational Rights and Privacy Act (FERPA) requires that access to individual student results be restricted to the student, the student's parent(s)/guardian(s), and authorized school personnel.

5.3 GRADE LEVEL SUMMARY RESULTS REPORTS

The *MEA Science Grade Level Summary Results* consists of four parts: the Grade Level Summary, the Science Results, the Disaggregated Science Results, and the Questionnaire Results.

The Grade Level Summary Report provides participation and performance summaries of the MEA Science test. The participation section on the top half of the page shows the number and percentage of students who were enrolled, tested, and not tested. The number and percentage of students tested with an approved accommodation, Limited English Proficient (LEP) students tested, LEP students tested with an approved accommodation, Individualized Education Plan (IEP) students tested, and IEP students tested with an approved accommodation are reported. The number and percent of students not tested is broken up into state-approved reasons and other reasons. The total number of students enrolled is defined as the number of students tested plus the number of students not tested. The performance summary on the bottom half of the report displays the achievement level distribution and average scaled score for the tested students in the school, SAU, and state.

The science results page provides information on historical test results and performance in specific domains. The purpose of these sections is to help schools and SAUs determine the extent to which their curricula are effective in helping students to achieve the particular standards. Information for school, SAU, and state includes:

- the total number of students enrolled, not tested (state-approved reason), not tested (other reason), and tested;
- the total number and percentage of students at each achievement level (based on the number in the tested column); and
- the mean scaled score.

Information about each science domain includes the following:

- the total possible points for that category
- a graphic display of the percent of total possible points for the school, SAU, and state. In this graphic display, there are symbols representing school, SAU, and state performance. In addition, there is a line representing the standard error of measurement. This statistic indicates how much a student's score could vary if the student were examined repeatedly with the same test (assuming that no learning were to occur between test administrations).

The disaggregated results pages present the relationship between performance and student reporting variables (see list on the following page) across school, SAU, and state levels. Each page shows the number of students categorized as enrolled, not tested (state-approved reason), not tested (other reason), and tested. The tables also provide the number and percentage of students within each of the four achievement levels and the mean scaled score by each reporting category.

The list of student reporting categories is as follows:

- All Students
- Gender
- Race/Ethnicity
- LEP Status (Limited English Proficiency)
- IEP (Individualized Education Plan)
- SES (Socioeconomic Status)
- Migrant
- Title I
- 504 Plan

The data for achievement levels and mean scaled score are based on the number shown in the tested column. The data for the reporting categories were provided by a data file download from the MDOE's student data system, Infinite Campus State Edition. Because performance is being reported by categories that can contain relatively low numbers of students, school personnel are advised, under FERPA guidelines, to treat these pages confidentially.

The questionnaire results page presents the relationship between performance and student response to selected questionnaire questions across school, SAU, and state levels. The table provides the percentage of students tested, the number and percentage of students within each of the four achievement levels, and the mean scaled score for each possible response to the selected question. Because performance is being reported by categories that can contain relatively low numbers of students, school personnel are advised, under FERPA guidelines, to treat these pages confidentially.

5.4 INTERACTIVE REPORTING

Four interactive reports were available via a secure website: Item Analysis Report, Achievement Level Summary, Released Items Summary Data, and Longitudinal Data. Each of these interactive reports is described in the following sections. Sample interactive reports are provided in Appendix G. To access these four interactive reports, the user clicked the interactive tab on the home page of the system and selected the report desired from the drop-down menu. Next, the user applied basic filtering options, such as the name of the SAU or school and the grade level test, to open the specific report. At this point, the user had the option of printing the report for the entire grade level or applying advanced filtering options to select a subgroup of students to analyze. Advanced filtering options include gender, ethnicity, LEP, IEP, and SES. All interactive reports, with the exception of the Longitudinal Data Report, allowed the user to provide a custom title for the report.

5.4.1 Item Analysis Report

The MEA Science Item Analysis Report provides a roster of all students in a school and provides performance on the common items that are released to the public. The student names and identification numbers are listed as row headers down the left side of the report. The items are listed as column headers in the same order they appear in released item documentation.

For each item, the following are shown:

- the content strand
- the depth of knowledge (DOK) code
- the item type
- the correct response key for multiple-choice items
- the possible score points

For each student, multiple-choice items are marked either with a plus sign (+), indicating that the student chose the correct multiple-choice response, or a letter (from A to D), indicating the incorrect response chosen by the student. For constructed-response items, the number of points earned is shown. All responses to released items are shown in the report, regardless of the student's participation status. The columns on the right side of the report show the Total Test results, broken into several categories. The Subcategory Points Earned columns show points earned by the student in each content area subcategory relative to total possible points. A Total Points Earned column is a summary of all points earned and total possible points in the content area. The last two columns show the student's scaled score and achievement level. Students reported as Not Tested are given a code in the achievement level column to indicate the reason the student did not test. It is important to note that not all items used to compute student scores are included in this report; only released items are included. At the bottom of the report, the average percentage correct for each multiple-choice item and average scores for the short answer and constructed-response items are shown for the school, SAU, and state. When the user applies advanced filtering criteria, the School and SAU Percent Correct/Average Score rows at the bottom of the report are blanked out and only the Group row and the State row for the group selected contain data. This report can be saved, printed, or exported as a PDF, XLS, or CSV file.

The Item Analysis Roster is marked as confidential to ensure it is kept secure within the school and SAU. FERPA requires that access to individual student results be restricted to the student, the student's parent(s)/guardian(s), and authorized school personnel.

5.4.2 Achievement Level Summary

The Achievement Level Summary provides a visual display of the percentages of students in each achievement level for a selected grade. The four achievement levels are represented by various colors in a pie

chart. A separate table is also included below the chart that shows the number and percentage of students in each achievement level. This report can be saved, printed, or exported as a PDF or JPG file.

5.4.3 Item Analysis Data

The Released Items Summary Data report is a school-level report that provides a summary of student responses to the released items for a selected grade. The report is divided into two sections by item type (multiple-choice and constructed-response). For multiple-choice items, the total number/percent of students who answered the item correctly and the number of students who chose each incorrect option or provided an invalid response are reported. An invalid response on a multiple-choice item is defined as "the item was left blank" or "the student selected more than one option for the item." For constructed-response items, point value and average score for the item are reported. Users are also able to view the actual released items within this report. If a user clicks on a particular magnifying glass icon next to a released item number, a pop-up box will open displaying the released item.

5.4.4 Longitudinal Data Reports

The Longitudinal Data Report is a confidential student level report that provides individual student performance data for multiple test administrations. The state-assigned student identification number is used to link students across test administrations. Student performance on future test administrations will be included on this report over time. This report can be saved, printed, or exported as a PDF file for a single student or for all students within a group. As with all other student data, this report is confidential and, per FERPA, requires that access is restricted.

5.5 ALL GRADE SUMMARY REPORTS

The All Grades Summary Report provides details on student performance by grade level tested at the SAU and state levels only.

Reported information includes:

- the total number of students enrolled, not tested (state-approved reason), not tested (other reason), and tested;
- the total number and percentage of students at each achievement level (based on the number in the tested column); and
- the mean scaled score.

5.6 DECISION RULES

To ensure that 2013–14 MEA Science reported results are accurate relative to collected data and other pertinent information, a document that delineates analysis and reporting rules was created. These decision

rules were observed in the analyses of MEA Science test data and in reporting the test results. Moreover, these rules are the main reference for quality assurance checks. The decision rules document is found in Appendix H.

The first set of rules pertains to general issues in reporting scores. Each issue is described, and pertinent variables are identified. The actual rules applied are described by the way they impact analyses and aggregations and their specific impact on each of the reports. The general rules are further grouped into issues pertaining to test items, school type, student exclusions, and number of students for aggregations.

The second set of rules pertains to reporting student participation. These rules describe which students were counted and reported for each subgroup in the student participation report.

5.7 QUALITY ASSURANCE

Quality assurance measures are embedded throughout the entire process of analysis and reporting. The data processor, data analyst, and psychometrician assigned to work on MEA implement quality control checks of their respective computer programs and intermediate products. Moreover, when data are handed off to different functions within the Data and Reporting Services (DRS) and Psychometrics and Research (P&R) departments, the sending function verifies that the data are accurate before handoff. Additionally, when a function receives a data set, the first step is to verify the data for accuracy.

Another type of quality assurance measure is parallel processing. Students' scaled scores are assigned by a psychometrician through a process of equating and scaling. The scaled scores are also computed by a data analyst to verify that scaled scores and corresponding achievement levels are assigned accurately. Respective scaled scores and assigned achievement levels are compared across all students for 100% agreement. Different exclusions that determine whether each student receives scaled scores and/or is included in different levels of aggregation are also parallel processed. Using the decision rules document, two data analysts independently write a computer program that assigns students' exclusions. For each grade and content area combination, the exclusions assigned by each data analyst are compared across all students. Only when 100% agreement is achieved can the rest of the data analysis be completed.

The third aspect of quality control involves the procedures implemented by the quality assurance group to check the accuracy of reported data. Using a sample of schools and SAUs, the quality assurance group verifies that reported information is correct. The step is conducted in two parts: (1) verify that the computed information was obtained correctly through appropriate application of different decision rules, and (2) verify that the correct data points populate each cell in the MEA reports. The selection of sample schools and SAUs for this purpose is very specific and can affect the success of the quality control efforts. There are two sets of samples selected that may not be mutually exclusive.

The first set includes those that satisfy the following criteria:

- One-school SAU

- Two-school SAU
- Multi-school SAU

The second set of samples includes SAUs or schools that have unique reporting situations as indicated by decision rules. This second set is necessary to ensure that each rule is applied correctly. The second set includes the following criteria:

- Private school (those that receive a stipulated percentage of federal and/or state funds)
- Small school that receives no school report
- Small SAU that receives no SAU report
- SAU that receives a report but with schools that are too small to receive a school report
- School with excluded (not tested) students
- School with homeschooled students

The quality assurance group uses a checklist to implement its procedures. After the checklist is completed, sample reports are circulated for psychometric checks and program management review. The appropriate sample reports are then presented to the MDOE for review and sign-off.

CHAPTER 6 CLASSICAL ITEM ANALYSIS

As noted in Brown (1983), "A test is only as good as the items it contains." A complete evaluation of a test's quality must include an evaluation of each item. Both *Standards for Educational and Psychological Testing* (AERA et al., 2014) and *Code of Fair Testing Practices in Education* (2004) include standards for identifying quality items. Items should assess only knowledge or skills that are identified as part of the domain being tested and should avoid assessing irrelevant factors. Items should also be unambiguous and free of grammatical errors, potentially insensitive content or language, and other confounding characteristics. In addition, items must not unfairly disadvantage students, particularly racial, ethnic, or gender groups. Both qualitative and quantitative analyses are conducted to ensure that MEA Science items meet these standards. Qualitative analyses are described in earlier chapters of this report; this chapter focuses on quantitative evaluations. Statistical evaluations are presented in four parts: 1) difficulty indices, 2) item-test correlations, 3) differential item functioning (DIF) statistics, and 4) dimensionality analyses. The item analyses presented here are based on the statewide administration of the MEA Science in spring 2014.

6.1 CLASSICAL DIFFICULTY AND DISCRIMINATION INDICES

All multiple-choice and constructed-response items are evaluated in terms of item difficulty according to standard classical test theory practices. Difficulty is defined as the average proportion of points achieved on an item and is measured by obtaining the average score on an item and dividing it by the maximum possible score for the item. Multiple-choice items are scored dichotomously (correct vs. incorrect) so, for these items, the difficulty index is simply the proportion of students who correctly answered the item. Constructed-response items are scored polytomously, meaning that a student can achieve a score of 0, 1, 2, 3, or 4. By computing the difficulty index as the average proportion of points achieved, the indices for the different item types are placed on a similar scale, ranging from 0.0 to 1.0 regardless of the item type. Although this index is traditionally described as a measure of difficulty, it is properly interpreted as an easiness index, because larger values indicate easier items. An index of 0.0 indicates that all students received no credit for the item, and an index of 1.0 indicates that all students received full credit for the item.

Items that are answered correctly by almost all students provide little information about differences in student abilities, but they do indicate knowledge or skills that have been mastered by most students. Similarly, items that are correctly answered by very few students provide little information about differences in student abilities, but may indicate knowledge or skills that have not yet been mastered by most students. In general, to provide the best measurement, difficulty indices should range from near-chance performance (0.25 for four-option multiple-choice items or essentially zero for constructed-response items) to 0.90, with the majority of items generally falling between around 0.4 and 0.7. However, on a standards-referenced assessment such as

the MEA Science, it may be appropriate to include some items with very low or very high item difficulty values to ensure sufficient content coverage.

A desirable characteristic of an item is for higher ability students to perform better on the item than lower ability students do. The correlation between student performance on a single item and total test score is a commonly used measure of this characteristic of the item. Within classical test theory, the item-test correlation is referred to as the item's discrimination because it indicates the extent to which successful performance on an item discriminates between high and low scores on the test. For constructed-response items, the item discrimination index used was the Pearson product-moment correlation; for multiple-choice items, the corresponding statistic is commonly referred to as a point-biserial correlation. The theoretical range of these statistics is -1.0 to 1.0, with a typical observed range from 0.2 to 0.6.

Discrimination indices can be thought of as measures of how closely an item assesses the same knowledge and skills assessed by other items contributing to the criterion total score. That is, the discrimination index can be thought of as a measure of construct consistency.

A summary of the item difficulty and item discrimination statistics for each grade is presented in Table 6-1. Note that the statistics are presented for all items as well as by item type (multiple-choice and constructed-response). The mean difficulty and discrimination values shown in the table are within generally acceptable and expected ranges.

Table 6-1. 2013–14 MeCAS Part I: Summary of Item Difficulty and Discrimination Statistics by Grade

Grade	*Item Type*	*Number of Items*	p-*Value*		*Discrimination*	
			Mean	*Standard Deviation*	*Mean*	*Standard Deviation*
5	ALL	36	0.68	0.18	0.28	0.08
	MC	32	0.71	0.16	0.27	0.07
	CR	4	0.41	0.08	0.41	0.05
8	ALL	44	0.63	0.17	0.32	0.08
	MC	40	0.65	0.16	0.30	0.07
	CR	4	0.48	0.14	0.46	0.09

A comparison of indices across grade levels is complicated because these indices are population dependent. Direct comparisons would require that either the items or students were common across groups. Since that is not the case, it cannot be determined whether differences in performance across grade levels are because of differences in student abilities, differences in item difficulties, or both. With this caveat in mind, it appears that generally students in grade 8 found their items more difficult than students in grade 5 found theirs, although the difference was slight.

Comparing the difficulty indices of multiple-choice items and constructed-response items is inappropriate because multiple-choice items can be answered correctly by guessing. Thus, it is not surprising that the difficulty indices for multiple-choice items tend to be higher (indicating that students performed better

on these items) than the difficulty indices for constructed-response items. Similarly, discrimination indices for the four-point constructed-response items were larger than those for the dichotomous items because of the greater variability of the former (i.e., the partial credit these items allow) and the tendency for correlation coefficients to be higher given greater variances of the correlates.

In addition to the item difficulty and discrimination summaries presented above, item-level classical statistics and item-level score point distributions were also calculated. Item level classical statistics are provided in Appendix I; item difficulty and discrimination values are presented for each item. The item difficulty and discrimination indices are within generally acceptable and expected ranges. Very few items were answered correctly at near-chance or near-perfect rates. Similarly, the positive discrimination indices indicate that students who performed well on individual items tended to perform well overall. There were a small number of items with low discrimination indices, but none were negative. While it is not inappropriate to include items with low discrimination values or with very high or very low item difficulty values to ensure that content is appropriately covered, there were very few such cases on the MEA Science. Item-level score point distributions are provided for constructed-response items in Appendix I; for each item, the percentage of students who received each score point is presented.

6.2 DIFFERENTIAL ITEM FUNCTIONING

Code of Fair Testing Practices in Education (2004) explicitly states that subgroup differences in performance should be examined when sample sizes permit and that actions should be taken to ensure that differences in performance are because of construct-relevant, rather than irrelevant, factors. *Standards for Educational and Psychological Testing* (AERA et al., 2014) includes similar guidelines. As part of the effort to identify such problems, MEA Science items were evaluated in terms of differential item functioning (DIF) statistics.

For the MEA Science, the standardization DIF procedure (Dorans & Kulick, 1986) was employed to evaluate subgroup differences. The standardization DIF procedure is designed to identify items for which subgroups of interest perform differently, beyond the impact of differences in overall achievement. The DIF procedure calculates the difference in item performance for two groups of students (at a time) matched for achievement on the total test. Specifically, average item performance is calculated for students at every total score. Then an overall average is calculated, weighting the total score distribution so that it is the same for the two groups.

When differential performance between two groups occurs on an item (i.e., a DIF index in the "low" or "high" categories, explained below), it may or may not be indicative of item bias. Course-taking patterns or differences in school curricula can lead to DIF, but for construct-relevant reasons. On the other hand, if subgroup differences in performance could be traced to differential experience (such as geographical living conditions or access to technology), the inclusion of such items should be reconsidered.

Computed DIF indices have a theoretical range from -1.0 to 1.0 for multiple-choice items, and the index is adjusted to the same scale for constructed-response items. Dorans and Holland (1993) suggested that index values between -0.05 and 0.05 should be considered negligible. The preponderance of MEA Science items fell within this range. Dorans and Holland further stated that items with values between -0.10 and -0.05 and between 0.05 and 0.10 (i.e., "low" DIF) should be inspected to ensure that no possible effect is overlooked, and that items with values outside the [-0.10, 0.10] range (i.e., "high" DIF) are more unusual and should be examined very carefully.[1]

For the 2013–14 MEA Science, the following subgroup comparisons were evaluated for DIF:

- Male versus Female
- No Disability versus Disability
- Not Economically Disadvantaged versus Economically Disadvantaged
- Non-LEP versus LEP
- White (non-Hispanic) versus Black or African American or Hispanic

The tables in Appendix J present the number of items classified as either "low" or "high" DIF, overall and by group favored.

6.3 DIMENSIONALITY ANALYSIS

The MEA Science tests were each designed to measure and report a single score on science achievement using a unidimensional scale. Thus, each of these tests is said to be measuring a single dimension, and the term "unidimensionality" is used to describe it.

Because each test is constructed with multiple content area subcategories and item types, and their associated knowledge and skills, the subtests associated with each of these could potentially result in a large number of secondary dimensions being invoked beyond the primary dimension that all the items on a test have in common. Generally, the scores on such subtests are highly correlated with each other; therefore, the primary dimension they share typically explains an overwhelming majority of variance in test scores. In fact, the presence of just such a dominant primary dimension is the psychometric assumption that provides the foundation for the unidimensional IRT models that were used for calibrating, linking, scaling, and equating the 2013–14 MEA Science test forms.

The purpose of dimensionality analysis is to investigate whether violation of the assumption of test unidimensionality is statistically detectable and, if so, (a) the degree to which unidimensionality is violated and (b) the nature of the multidimensionality. Findings from dimensionality analyses performed on the 2013–

[1] It should be pointed out here that DIF for items is evaluated initially at the time of field testing. If an item displays high DIF, it is flagged for review by a Measured Progress content specialist. The content specialist consults with the Department of Education to determine whether to include the flagged item in a future operational test administration.

14 MEA Science common items for grades 5 and 8 are reported on the following page. (Note: only common items were analyzed since they are used for score reporting.)

Dimensionality analyses were conducted using the nonparametric IRT-based methods DIMTEST (Stout, 1987; Stout, Froelich, & Gao, 2001) and DETECT (Zhang & Stout, 1999). Nonparametric techniques were preferred for this analysis because such techniques avoid strong parametric modeling assumptions while still adhering to the fundamental principles of item response theory. Parametric techniques, such as nonlinear factor analysis, make strong assumptions that are often inappropriate for real data, such as assuming a normal distribution for ability and lower asymptotes of zero for the item characteristic curves.

Both DIMTEST and DETECT use as their basic statistical building block the estimated average conditional covariances for item pairs. A conditional covariance is the covariance between two items conditioned on expected total score for the rest of the test, and the average conditional covariance is obtained by averaging over all possible conditioning scores. When a test is strictly unidimensional, all conditional covariances are expected to take on values within random noise of zero, indicating statistically independent item responses for examinees with equal expected scores. Non-zero conditional covariances are essentially violations of the principle of local independence, and local dependence implies multidimensionality. Thus, non-random patterns of positive and negative conditional covariances are indicative of multidimensionality. In particular, when multiple dimensions are present, items measuring the same dimension will have positive conditional covariance with each other, whereas items measuring different dimensions will have negative conditional covariances with each other. For example, if multiple-choice items measure a different dimension from constructed-response items, we would expect multiple-choice items to have positive conditional covariances with each other, constructed-response items to have positive conditional covariances with each other, and multiple-choice items to have negative conditional covariances with constructed-response items.

DIMTEST is a hypothesis-testing procedure for detecting violations of local independence. For the exploratory analyses conducted for the MEA Science tests, the data were first divided into a training sample and a cross-validation sample. Then an analysis of the conditional covariances was conducted on the training sample data to find the cluster of items that displays the greatest evidence of local dependence. The cross-validation sample was then used to test whether the conditional covariances of the selected cluster of items displays local dependence, conditioning on total score on the non-clustered items. The DIMTEST statistic follows a standard normal distribution under the null hypothesis of unidimensionality.

DETECT is an effect-size measure of multidimensionality. For the exploratory analyses conducted for the MEA Science tests, as with DIMTEST, the data were first randomly divided into a training sample and a cross-validation sample. (Note: The training and cross-validation samples used for the DETECT analyses were randomly drawn independently of the samples used for the DIMTEST analyses.) The training sample was then used to find a set of mutually exclusive and collectively exhaustive clusters of items that best fit a systematic pattern of positive conditional covariances for pairs of items from the same cluster and negative conditional covariances from different clusters. Next, the clusters from the training sample were used with the

cross-validation sample data to average the conditional covariances: within-cluster conditional covariances were summed, from this sum the between-cluster conditional covariances were subtracted, this difference was divided by the total number of item pairs, and this average was multiplied by 100 to yield an index of the average violation of local independence for an item pair. DETECT values less than 0.2 indicate very weak multidimensionality (or near unidimensionality), values of 0.2 to 0.4 weak to moderate multidimensionality; values of 0.4 to 1.0 moderate to strong multidimensionality, and values greater than 1.0 very strong multidimensionality.

As mentioned above, in applying DIMTEST and DETECT to the 2013–14 MEA Science tests, the data for each grade were split into a training sample and a cross-validation sample. Each grade had at least 11,800 student examinees, so the training sample and cross-validation sample for each grade had at least 5,900 students. DIMTEST was then applied to each grade. DETECT was applied to each dataset for which the DIMTEST null hypothesis was rejected in order to estimate the effect size of the multidimensionality.

Because of the large sample sizes of both MEA Science tests, DIMTEST would be sensitive to even quite small violations of unidimensionality, and the null hypothesis was rejected at level 0.01 for both datasets. These results were not surprising because strict unidimensionality is an idealization that almost never holds exactly for a given dataset. Thus, it was important to use DETECT to estimate the effect size of the violations of local independence found by DIMTEST. Table 6-2 displays the multidimensional effect size estimates from DETECT.

As shown in Table 6-2, both of the DETECT values indicate multidimensionality is very weak (DETECT values less than 0.20). Also shown in Table 6-2 are the DETECT values from last year's dimensionality analysis. This year's results are seen to be very similar to last year's in that both sets of results indicated very weak multidimensionality for both tests.

Table 6-2. 2013–14 MEA Science: Multidimensionality Effect Sizes

Grade	*Multidimensionality Effect Size*	
	2013–2014	*2012–2013*
5	.13	.14
8	.10	.14

We also investigated how DETECT divided the tests into clusters to see if there were any discernable patterns with respect to the multiple-choice and open-response item types. Focusing only on multiple-choice and open-response separation that occurs in both the DETECT clusters and in the pattern of the signs of the conditional covariances (i.e., strong and clear separation), this year's analysis indicates no strong separation in either Grade 5 or Grade 8. This result agreed with the findings from most past years for the science tests.

The combined results of DETECT and DIMTEST indicate that multidimensionality is present in both MEA Science tests but the magnitude is small in both cases. Investigating the patterns of signs in the conditional covariance matrices, there were no indications of strong non-random patterns. We also

investigated the DETECT clusters with respect to multiple-choice and open-response separation, and found no strong evidence in either Grade 5 or Grade 8. Investigating the DETECT effect size, the violations of local independence were very small and do not warrant any changes in test design or scoring. In addition, we only investigated the separation of items by multiple-choice and open-response. A more thorough investigation by substantive content experts may result in cogent interpretations of other clusters in regard to skills and knowledge areas measured by the items.

CHAPTER 7 ITEM RESPONSE THEORY SCALING AND EQUATING

This chapter describes the procedures used to calibrate, equate, and scale the MEA Science test. During the course of these psychometric analyses, a number of quality control procedures and checks on the processes were implemented. These procedures included evaluation of the calibration processes (e.g., checking the number of Newton cycles required for convergence for reasonableness, checking item parameters and their standard errors for reasonableness, examining test characteristic curves [TCCs] and test information functions [TIFs] for reasonableness), evaluation of model fit, evaluation of equating items (e.g., delta analyses, rescore analyses, examination of a-plots and b-plots for reasonableness), and evaluation of the scaling results (e.g., parallel processing by the Data Services and Static Reporting and Psychometrics and Research departments, comparison of lookup tables to the previous year's lookup tables). An equating report, which provided complete documentation of the quality control procedures and results, was submitted to the MDOE for their approval prior to production of student reports.

Table 7-1 lists items that required intervention either during item calibration or as a result of the evaluations of the equating items. For each flagged item, the table shows the reason it was flagged and what action was taken. Descriptions of the evaluations and results are included in the Item Response Theory Results and Equating Results sections found later in this chapter.

Table 7-1. 2013–14 MEA Science: Items that Required Intervention During IRT Calibration and Equating

Grade	*Item*	*Reason*	*Action*
5	228136	b/b analysis	removed from equating
8	158374	c-parameter	set c = 0
	187347	b/b analysis	removed from equating

7.1 ITEM RESPONSE THEORY

All MEA Science items were calibrated using item response theory (IRT). IRT uses mathematical models to define a relationship between an unobserved measure of student performance, usually referred to as theta (θ), and the probability (p) of getting a dichotomous item correct or of getting a particular score on a polytomous item. In IRT, it is assumed that all items are independent measures of the same construct (i.e., of the same θ). Another way to think of θ is as a mathematical representation of the latent trait of interest. Several common IRT models are used to specify the relationship between θ and p (Hambleton & van der Linden, 1997; Hambleton & Swaminathan, 1985). The process of determining the specific mathematical relationship between θ and p is called item calibration. After items are calibrated, they are defined by a set of

parameters that specify a nonlinear, monotonically increasing relationship between θ and p. Once the item parameters are known, an estimate of θ for each student can be calculated. This estimate, $\hat{\theta}$, is considered to be an estimate of the student's true score or a general representation of student performance. It has characteristics that may be preferable to those of raw scores for equating purposes.

For the 2013–14 MEA Science, the three-parameter logistic (3PL) model was used for dichotomous items and the graded response model (GRM) was used for polytomous items. The 3PL model for dichotomous items can be defined as

$$P_i\left(1\middle|\theta_j,\xi_i\right) = c_i + (1 - c_i)\frac{\exp[Da_i(\theta_j - b_i)]}{1+\exp[Da_i(\theta_j - b_i)]}$$

where
i indexes the items,
j indexes students,
a represents item discrimination,
b represents item difficulty,
c is the pseudo guessing parameter,
ξ_i represents the set of item parameters (a, b, and c), and
D is a normalizing constant equal to 1.701.

In the GRM for polytomous items, an item is scored in $k + 1$ graded categories that can be viewed as a set of k dichotomies. At each point of dichotomization (i.e., at each threshold), a two-parameter model can be used. This implies that a polytomous item with $k + 1$ categories can be characterized by k item category threshold curves (ICTC) of the two-parameter logistic form

$$P_{ik}^{*}\left(1\middle|\theta_j, a_i, b_i, d_{ik}\right) = \frac{\exp[Da_i(\theta_j - b_i + d_{ik})]}{1+\exp[Da_i(\theta_j - b_i + d_{ik})]}$$

where
i indexes the items,
j indexes students,
k indexes threshold,
a represents item discrimination,
b represents item difficulty,
d represents threshold, and
D is a normalizing constant equal to 1.701.

After computing k ICTCs in the GRM, $k + 1$ item category characteristic curves (ICCCs) are derived by subtracting adjacent ICTCs

$$P_{ik}(1|\theta_j) = P^*_{i(k-1)}(1|\theta_j) - P^*_{ik}(1|\theta_j)$$

where
P_{ik} represents the probability that the score on item *i* falls in category *k*, and
P^*_{ik} represents the probability that the score on item *i* falls above the threshold *k* ($P^*_{i0} = 1$ and $P^*_{i(m+1)} = 0$).

The GRM is also commonly expressed as

$$P_{ik}(k|\theta_j, \xi_i) = \frac{\exp[Da_i(\theta_j - b_i + d_k)]}{1+\exp[Da_i(\theta_j - b_i + d_k)]} - \frac{\exp[Da_i(\theta_j - b_i + d_{k+1})]}{1+\exp[Da_i(\theta_j - b_i + d_{k+1})]}$$

where
ξ_i represents the set of item parameters for item *i*.

Finally, the item characteristic curve (ICC) for polytomous items is computed as a weighted sum of ICCCs, where each ICCC is weighted by a score assigned to a corresponding category.

$$P_i(1|\theta_j) = \sum_k^{m+1} w_{ik} P_{ik}(1|\theta_j)$$

For more information about item calibration and determination, the reader is referred to Lord and Novick (1968), Hambleton and Swaminathan (1985), or Baker and Kim (2004).

7.2 ITEM RESPONSE THEORY RESULTS

The tables in Appendix K give the IRT item parameters of all common items on the 2013–14 MEA Science tests by grade. In addition, Appendix L shows graphs of the TCCs and TIFs, which are defined below.

TCCs display the expected (average) raw score associated with each θ_j value between -4.0 and 4.0. Mathematically, the TCC is computed by summing the ICCs of all items that contribute to the raw score. Using the notation introduced in Section 7.1, the expected raw score at a given value of θ_j is

$$E(X|\theta_j) = \sum_{i=1}^{n} P_i(1|\theta_j)$$

where
i indexes the items (and n is the number of items contributing to the raw score),
j indexes students (here, θ_j runs from -4.0 to 4.0), and
$E(X|\theta_j)$ is the expected raw score for a student of ability θ_j.

The expected raw score monotonically increases with θ_j, consistent with the notion that students of high ability tend to earn higher raw scores than do students of low ability. Most TCCs are "S-shaped"—flatter at the ends of the distribution and steeper in the middle.

The TIF displays the amount of statistical information that the test provides at each value of θ_j. Information functions depict test precision across the entire latent trait continuum. There is an inverse

relationship between the information of a test and its standard error of measurement (SEM). For long tests, the SEM at a given θ_j is approximately equal to the inverse of the square root of the statistical information at θ_j (Hambleton, Swaminathan, & Rogers, 1991), as follows:

$$SEM(\theta_j) = \frac{1}{\sqrt{i(\theta_j)}}$$

Compared to the tails, TIFs are often higher near the middle of the distribution where most students are located and where most items are sensitive by design.

Table 7-1, presented at the beginning of this chapter, shows that three items were flagged based on the quality control checks implemented during the calibration process. One item in grade 8 was flagged because the guessing parameter (c-parameter) was poorly estimated. Difficulty in estimating the c-parameter is not at all unusual and is well documented in psychometric literature, especially when an item's discrimination is below 0.50. Fixing the c-parameter to zero resulted in reasonable and stable item parameter estimates and improved model fit. One item in grade 5 and one item in grade 8 were also flagged because the difficulty level of each item showed statistical evidence of having changed significantly from previous usage. Both items were dropped from the equating set.

The number of Newton cycles required for convergence for each grade during the IRT analysis can be found in Table 7-2. For both grades, the number of cycles required fell within acceptable ranges.

Table 7-2. 2013–14 MEA Science: Number of Newton Cycles Required for Convergence

Grade	*Cycles*
5	37
8	30

7.3 EQUATING

The purpose of equating is to ensure that scores obtained from different forms of a test are equivalent to each other. Equating may be used if multiple test forms are administered in the same year, as well as to equate one year's forms to those given in the previous year. Equating ensures that students are not advantaged or disadvantaged because the test form they took is easier or harder than those taken by other students.

The 2013–14 administration of MEA Science used a raw score-to-theta equating procedure in which test forms were equated to the theta scale established on the reference form (i.e., the form used in the most recent standard setting). This is accomplished through the chained linking design, in which every new form is equated back to the theta scale of the previous year's test form. It can therefore be assumed that the theta scale of every new test form is the same as the theta scale of the reference form, since this is where the chain originated.

The groups of students who took the equating items on the 2013–14 MEA Science tests are not equivalent to the groups who took them in the reference year. IRT is particularly useful for equating scenarios that involve nonequivalent groups (Allen & Yen, 1979). Equating for MEA Science uses the anchor-test-nonequivalent-groups design described by Petersen, Kolen, and Hoover (1989). In this equating design, no assumption is made about the equivalence of the examinee groups taking different test forms (that is, naturally occurring groups are assumed). Comparability is instead evaluated by utilizing a set of anchor items (also called equating items). However, the equating items are designed to mirror the common test in terms of item types and distribution of emphasis.

Item parameter estimates for 2013–14 were placed on the 2012–13 scale by using the method of Stocking and Lord (1983), which is based on the IRT principle of item parameter invariance. According to this principle, the equating items for both the 2012–13 and 2013–14 MEA Science tests should have the same item parameters. After the item parameters for each 2013–14 test were estimated using PARSCALE (Muraki & Bock, 2003), the Stocking and Lord method was employed to find the linear transformation (slope and intercept) that adjusted the equating items' parameter estimates such that the 2013–14 TCC was as close as possible to that of 2012–13.

7.4 EQUATING RESULTS

Prior to calculating the Stocking and Lord transformation constants, a variety of evaluations of the equating items were conducted. Three items were flagged for evaluation as a result of these procedures.

Appendix M presents the results from the delta analysis. This procedure was used to evaluate adequacy of equating items; no items had "true" in the discard status, indicating that no items was flagged as inappropriate for use in equating.

Also presented in Appendix M are the results from the rescore analysis. With this analysis, 200 random papers from the previous year were interspersed with this year's papers to evaluate scorer consistency from one year to the next. All effect sizes were well below the criterion value for excluding an item as an equating item, 0.80 in absolute value.

Finally, a-plots and b-plots, which show IRT parameters for 2013–14 plotted against the values for 2012–13, are presented in Appendix N. One item in grade 5 and one item in grade 8 were identified as outliers in the b/b plot and were removed as equating items.

Once all evaluations of the equating items were complete, the Stocking and Lord method of equating was used to place the item parameters onto the previous year's scale, as described above. The Stocking and Lord transformation constants are presented in Table 7-3.

Table 7-3. 2013–14 MEA Science: Stocking and Lord Transformation Constants

Grade	α-*slope*	b-*intercept*
5	0.998	0.232
8	0.968	0.197

The next administration of MEA Science (2014–15) will be scaled to the 2013–14 administration using the same equating method described above.

7.5 ACHIEVEMENT STANDARDS

MEA Science standards to establish achievement level cut scores in grades 5 and 8 were set in May 2009. The standard setting meeting and results were discussed in the 2009 technical report and standard setting report provided at that time. As alluded to in the discussion of equating above, the scale was established during that base year, and the forms serve as the reference for subsequent equating. The θ metric cut scores that emerged from the standard setting meeting will remain fixed throughout the assessment program unless standards are reset for any reason.

7.6 REPORTED SCALED SCORES

7.6.1 Description of Scale

Because the θ scale used in IRT calibrations is not readily understood by most stakeholders, reporting scales were developed for the MEA Science tests. The reporting scale is a simple linear transformation of the underlying θ scale used in the IRT calibrations. The scales were developed such that the scaled scores for grade 5 range from 500 to 580, and for grade 8 from 800 through 880. Likewise, for each grade level, the Partially Proficient/Proficient and the Proficient/Proficient With Distinction cuts were placed at "x42" and "x62" respectively, where x denotes the grade level. (At the student level, scaled scores were reported as even numbers only.)

By providing information that is more specific about the position of a student's results, scaled scores supplement achievement level scores. School and SAU level scaled scores are calculated by computing the average of student level scaled scores. Students' raw scores (i.e., total number of points) on the MEA Science tests were translated to scaled scores using the data analytic process known as *scaling*. Scaling simply converts from one scale to another. In the same way that a given temperature can be expressed on either Fahrenheit or Celsius scales, or the same distance can be expressed in either miles or kilometers, student scores on the 2013–14 MEA Science tests can be expressed in raw or scaled scores.

It is important to note that converting from raw scores to scaled scores does not change students' achievement level classifications. Given the relative simplicity of raw scores, it is fair to question why scaled scores for MEA are reported instead of raw scores. Scaled scores make the reporting of results consistent. To illustrate, standard setting typically results in different *raw* cut scores across grades. The raw cut score between Partially Proficient and Proficient could be, for example, 35 in grade 5 but 33 in grade 8, yet both of

these raw scores would be transformed to scaled scores of x42. It is this uniformity across *scaled scores* that facilitates the understanding of student performance. The psychometric advantage of scaled scores over raw scores comes from there being linear transformations of θ. Since the θ scale is used for equating, scaled scores are comparable from one year to the next. Raw scores are not.

7.6.2 Calculations

The scaled scores are obtained by a simple translation of ability estimates ($\hat{\theta}$) using the linear relationship between threshold values on the θ metric and their equivalent values on the scaled score metric. Students' ability estimates are based on their raw scores and are found by mapping through the TCC. Scaled scores are calculated using the linear equation

$$SS = m\hat{\theta} + b$$

where
m is the slope and
b is the intercept.

A separate linear transformation was used for each grade. Each line is determined by fixing the "x42" and "x62" values (e.g., 542 and 562 for grade 5). Because only two points within the θ scaled score space were fixed, the cut scores between Substantially Below Proficient and Partially Proficient varied across grades. Table 7-4 presents the cuts on the theta scale and the reporting scale for each grade (i.e., the minimum scaled score for getting into the next achievement level). It is important to repeat that the values in Table 7-4 do not change from year to year, because the cut scores along the θ scale do not change unless standards are reset. Also, in a given year it may not be possible to attain a particular scaled score, but the scaled score cuts will remain the same.

Table 7-4. 2013–14 MEA Science: Cut Scores on the Theta and Reporting Scales by Grade

Grade	Θ *Cuts*			*Minimum*	*Scaled Score Cuts*			*Maximum*
	SBP/PP	*PP/P*	*P/PWD*		*SBP/PP*	*PP/P*	*P/PWD*	
5	-1.2750	-0.1229	1.7747	500	530	542	562	580
8	-1.3906	-0.3913	1.0164	800	828	842	862	880

SBP = Substantially Below Proficient; PP = Partially Proficient; P = Proficient; PWD = Proficient With Distinction

Table 7-5 shows the slope and intercept terms used to calculate the scaled scores. Note that the values in Table 7-5 will not change unless the standards are reset.

Table 7-5. 2013–14 MEA Science: Scaled Score Transformation Constants by Grade

Grade	α-*slope*	b-*intercept*
5	10.53963	543.2953
8	14.20757	847.5594

Appendix O contains raw score to scaled score lookup tables for this year and last year. These are the actual tables that were used to determine student scaled scores, error bands, and achievement levels.

7.6.3 Distributions

Graphs of the scaled score cumulative frequency distributions for the last three years are presented in Appendix P. Note that the graphs show the percentage of students at or below each scaled score; thus, the lowest line in a given graph depicts the highest performing group. For example, in the graph for grade 8, the line showing the cumulative distribution for 2013–14 is consistently higher than that for 2012–13. This pattern indicates that student performance on the grade 5 science test declined from 2011–12 to 2012–13.

Appendix P also shows, in Table P-1, achievement level distributions by grade. Results are shown for the 2011–12, 2012–13, and 2013–14 administrations.

CHAPTER 8 RELIABILITY

Although an individual item's performance is an important focus for evaluation, a complete evaluation of an assessment must also address the way items function together and complement one another. Tests that function well provide a dependable assessment of the student's level of ability. Unfortunately, no test can do this perfectly. A variety of factors can contribute to a given student's score being either higher or lower than his or her true ability. For example, a student may misread an item or mistakenly fill in the wrong bubble when he or she knew the answer. Collectively, extraneous factors that impact a student's score are referred to as measurement error. Any assessment includes some amount of measurement error; that is, no measurement is perfect. This is true of all academic assessments—some students will receive scores that underestimate their true ability, and other students will receive scores that overestimate their true ability. When tests have a high amount of measurement error, student scores are very unstable. Students with high ability may get low scores or vice versa. Consequently, one cannot reliably measure a student's true level of ability with such a test. Assessments that have less measurement error (i.e., errors made are small on average and student scores on such a test will consistently represent their ability) are described as reliable.

There are a number of ways to estimate an assessment's reliability. One possible approach is to give the same test to the same students at two different points in time. If students receive the same scores on each test, then the extraneous factors affecting performance are small and the test is reliable. (This is referred to as "test-retest reliability.") A potential problem with this approach is that students may remember items from the first administration or may have gained (or lost) knowledge or skills in the interim between the two administrations. A solution to the "remembering items" problem is to give a different, but parallel test at the second administration. If student scores on each test correlate highly, the test is considered reliable. (This is known as "alternate forms reliability" because an alternate form of the test is used in each administration.) This approach, however, does not address the problem that students may have gained (or lost) knowledge or skills in the interim between the two administrations. In addition, the practical challenges of developing and administering parallel forms generally preclude the use of parallel forms reliability indices. One way to address the latter two problems is to split the test in half and then correlate students' scores on the two half-tests; this, in effect, treats each half-test as a complete test. By doing this, the problems associated with an intervening time interval and with creating and administering two parallel forms of the test are alleviated. This is known as a "split-half estimate of reliability." If the two half-test scores correlate highly, items on the two half-tests must be measuring very similar knowledge or skills. This is evidence that the items complement one another and function well as a group. This also suggests that measurement error will be minimal.

The split-half method requires psychometricians to select items that contribute to each half-test score. This decision may have an impact on the resulting correlation, since each different possible split of the test into halves will result in a different correlation. Another problem with the split-half method of calculating

reliability is that it underestimates reliability, because test length is cut in half. All else being equal, a shorter test is less reliable than a longer test. Cronbach (1951) provided a statistic, α (alpha), that eliminates the problem of the split-half method by comparing individual item variances to total test variance. Cronbach's α was used to assess the reliability of the 2013–14 MEA Science

$$\alpha \equiv \frac{n}{n-1}\left[1 - \frac{\sum_{i=1}^{n} \sigma^2_{(Y_i)}}{\sigma^2_x}\right]$$

where
i indexes the item,
n is the total number of items,
$\sigma^2_{(Y_i)}$ represents individual item variance, and
σ^2_x represents the total test variance.

8.1 Reliability and Standard Errors of Measurement

Table 8-1 presents descriptive statistics, Cronbach's α coefficient, and raw score standard errors of measurement (SEMs) for each grade.

Table 8-1. 2013–14 MEA Science: Raw Score Descriptive Statistics, Cronbach's Alpha, and Standard Errors of Measurement (SEM) by Grade

Grade	*Number of Students*	*Raw Score*			*Alpha*	*SEM*
		Maximum	*Mean*	*Standard Deviation*		
5	12,979	48	29.30	6.56	0.79	2.99
8	13,473	56	33.65	8.45	0.85	3.27

Because different grades have different test designs (e.g., the number of items varies by test), it is inappropriate to make inferences about the quality of one test by comparing its reliability to that of another test from a different grade.

8.2 Subgroup Reliability

The reliability coefficients discussed in the previous section were based on the overall population of students who took the 2013–14 MEA Science tests. Appendix Q presents reliabilities for various subgroups of interest. Subgroup Cronbach's α's were calculated using the formula defined above based only on the members of the subgroup in question in the computations; values are only calculated for subgroups with 10 or more students.

For several reasons, the results of this section should be interpreted with caution. First, inherent differences between tests preclude making valid inferences about the quality of a test based on statistical comparisons with other tests. Second, reliabilities are dependent not only on the measurement properties of a

test, but on the statistical distribution of the studied subgroup. For example, it can readily be seen in Appendix Q that subgroup sample sizes may vary considerably, which results in natural variation in reliability coefficients. Alternatively α, which is a type of correlation coefficient, may be artificially depressed for subgroups with little variability (Draper & Smith, 1998). Third, there is no industry standard to interpret the strength of a reliability coefficient, and this is particularly true when the population of interest is a single subgroup.

8.3 SUBCATEGORY RELIABILITY

Of even more interest are reliabilities for the reporting subcategories within MEA Science, described in Chapter 5. Cronbach's α coefficients for subcategories were calculated via the same formula defined previously using just the items of a given subcategory in the computations. Results are presented in Appendix Q. Once again as expected, because they are based on a subset of items, rather than the full test, computed subcategory reliabilities were lower (sometimes substantially so) than were overall test reliabilities, and interpretations should take this into account. The subcategory reliabilities were lower than those based on the total test and approximately to the degree one would expect based on classical test theory. Qualitative differences between subtests once again preclude valid inferences about the quality of the full test based on statistical comparisons among subtests.

8.4 RELIABILITY OF ACHIEVEMENT LEVEL CATEGORIZATION

While related to reliability, the accuracy and consistency of classifying students into achievement categories are even more important statistics in a standards-based reporting framework (Livingston & Lewis, 1995). After the achievement levels were specified and students were classified into those levels, empirical analyses were conducted to determine the statistical accuracy and consistency of the classifications. For the MEA Science tests, students are classified into one of four achievement levels: Substantially Below Proficient, Partially Proficient, Proficient, or Proficient With Distinction. This section of the report explains the methodologies used to assess the reliability of classification decisions, and results are given.

Accuracy refers to the extent to which decisions based on test scores match decisions that would have been made if the scores did not contain any measurement error. Accuracy must be estimated, because errorless test scores do not exist. Consistency measures the extent to which classification decisions based on test scores match the decisions based on scores from a second, parallel form of the same test. Consistency can be evaluated directly from actual responses to test items if two complete and parallel forms of the test are given to the same group of students. In operational test programs, however, such a design is usually impractical. Instead, techniques have been developed to estimate both the accuracy and consistency of classification decisions based on a single administration of a test. The Livingston and Lewis (1995) technique

was used for the 2013–14 MEA Science tests because it is easily adaptable to all types of testing formats, including mixed-format tests.

The accuracy and consistency estimates reported in Appendix R make use of "true scores" in the classical test theory sense. A true score is the score that would be obtained if a test had no measurement error. Of course, true scores cannot be observed and so must be estimated. In the Livingston and Lewis method, estimated true scores are used to categorize students into their "true" classifications.

For the 2013–14 MEA Science tests, after various technical adjustments (described in Livingston & Lewis, 1995), a four-by-four contingency table of accuracy was created for each grade, where cell [*i, j*] represented the estimated proportion of students whose true score fell into classification *i* (where *i* = 1 to 4) and observed score into classification *j* (where *j* = 1 to 4). The sum of the diagonal entries (i.e., the proportion of students whose true and observed classifications matched) signified overall accuracy.

To calculate consistency, true scores were used to estimate the joint distribution of classifications on two independent, parallel test forms. Following statistical adjustments per Livingston and Lewis (1995), a new four-by-four contingency table was created for each grade and populated by the proportion of students who would be categorized into each combination of classifications according to the two (hypothetical) parallel test forms. Cell [*i, j*] of this table represented the estimated proportion of students whose observed score on the first form would fall into classification *i* (where *i* = 1 to 4) and whose observed score on the second form would fall into classification *j* (where *j* = 1 to 4). The sum of the diagonal entries (i.e., the proportion of students categorized by the two forms into exactly the same classification) signified overall consistency.

Another way to measure consistency is to use Cohen's (1960) coefficient κ (kappa), which assesses the proportion of consistent classifications after removing the proportion of consistent classifications that would be expected by chance. It is calculated using the following formula:

$$\kappa = \frac{(\text{Observed agreement}) - (\text{Chance agreement})}{1 - (\text{Chance agreement})} = \frac{\sum_i C_{ii} - \sum_i C_{i.} C_{.i}}{1 - \sum_i C_{i.} C_{.i}}$$

where

$C_{i.}$ is the proportion of students whose observed achievement level would be Level *i* (where *i* = 1–4) on the first hypothetical parallel form of the test;

$C_{.i}$ is the proportion of students whose observed achievement level would be Level *i* (where *i* = 1–4) on the second hypothetical parallel form of the test;

C_{ii} is the proportion of students whose observed achievement level would be Level *i* (where *i* = 1–4) on both hypothetical parallel forms of the test.

Because κ is corrected for chance, its values are lower than other consistency estimates.

8.4.1 Accuracy and Consistency

The accuracy and consistency analyses described above are provided in Table R-1 of Appendix R. The table includes overall accuracy and consistency indices, including kappa. Accuracy and consistency

values conditional upon achievement level are also given. For these calculations, the denominator is the proportion of students associated with a given achievement level. For example, the conditional accuracy value is 0.74 for Substantially Below Proficient for grade 5. This figure indicates that among the students whose true scores placed them in this classification, 74% would be expected to be in this classification when categorized according to their observed scores. Similarly, a consistency value of 0.53 indicates that 53% of students with observed scores in the Substantially Below Proficient level would be expected to score in this classification again if a second, parallel test form were used.

For some testing situations, the greatest concern may be decisions around level thresholds. For example, in testing done for No Child Left Behind (NCLB) accountability purposes, the primary concern is distinguishing between students who are proficient and those who are not yet proficient. In this case, the accuracy of the Partially Proficient/Proficient threshold is of greatest interest. Table R-2 in Appendix R provides accuracy and consistency estimates at each cut, as well as false positive and false negative decision rates for the 2013–14 MEA Science tests. (A false positive is the proportion of students whose observed scores were above the cut and whose true scores were below the cut. A false negative is the proportion of students whose observed scores were below the cut and whose true scores were above the cut.)

The above indices are derived from Livingston and Lewis's (1995) method of estimating the accuracy and consistency of classifications. It should be noted that Livingston and Lewis discuss two versions of the accuracy and consistency tables. A standard version performs calculations for forms parallel to the form taken. An "adjusted" version adjusts the results of one form to match the observed score distribution obtained in the data. The tables use the standard version for two reasons: (1) this "unadjusted" version can be considered a smoothing of the data, thereby decreasing the variability of the results; and (2) for results dealing with the consistency of two parallel forms, the unadjusted tables are symmetrical, indicating that the two parallel forms have the same statistical properties. This second reason is consistent with the notion of forms that are parallel; that is, it is more intuitive and interpretable for two parallel forms to have the same statistical distribution.

Note that, as with other methods of evaluating reliability, Decision Accuracy and Consistency (DAC) statistics calculated based on small groups can be expected to be lower than those calculated based on larger groups. For this reason, the values presented in Appendix R should be interpreted with caution. In addition, it is important to remember that it is inappropriate to compare DAC statistics between grades.

8.5 INTERRATER CONSISTENCY

Chapter 4 of this report describes in detail the processes that were implemented to monitor the quality of the hand-scoring of student responses for constructed-response items. One of these processes was double-blind scoring: approximately 10% of student responses were randomly selected and scored independently by two different scorers. Results of the double-blind scoring were used during the scoring process to identify

scorers that required retraining or other intervention and are presented here as evidence of the reliability of the MEA Science tests. A summary of the interrater consistency results is presented in Table 8-2. Results in the table are collapsed across the hand-scored items by grade. This same information is provided at the item level in Appendix S.

Table 8-2. 2013–14 MEA Science: Summary of Interrater Consistency Statistics Collapsed Across Items by Grade

Grade	*Number of*		*Percent*		*Correlation*	*Percent of Third Scores*
	Score Categories	*Included Scores*	*Exact*	*Adjacent*		
5	5	5510	67.82	29.29	0.81	2.45
8	5	5316	73.44	24.02	0.85	2.26

CHAPTER 9 VALIDITY

Because interpretations of test scores, and not a test itself, are evaluated for validity, the purpose of the *2013–14 MEA Science Technical Report* is to describe several technical aspects of the MEA Science tests in support of score interpretations (AERA, 2014). Each chapter contributes an important component in the investigation of score validation: test development and design; test administration; scoring, scaling, and equating; item analyses; reliability; and score reporting.

Standards for Educational and Psychological Testing (AERA, et al., 2014) provides a framework for describing sources of evidence that should be considered when constructing a validity argument. The evidence around test content, response processes, internal structure, relationship to other variables, and consequences of testing speaks to different aspects of validity but are not distinct *types* of validity. Instead, each contributes to a body of evidence about the comprehensive validity of score interpretations.

Evidence on test content validity is meant to determine how well the assessment tasks represent the curriculum and standards for each content area and grade level. Content validation is informed by the item development process, including how the test blueprints and test items align to the curriculum and standards. Viewed through the lens provided by the standards, evidence based on test content was extensively described in Chapters 2 and 3. Item alignment with MEA Science content standards; item bias, sensitivity, and content appropriateness review processes; adherence to the test blueprint; use of multiple item types; use of standardized administration procedures, with accommodated options for participation; and appropriate test administration training are all components of validity evidence based on test content. As discussed earlier, all questions are aligned by MEA Science educators to Maine's 2007 *Learning Results* (MLRs) and undergo several rounds of review for content fidelity and appropriateness. Items are presented to students in multiple formats (constructed-response and multiple-choice). Finally, tests are administered according to state-mandated standardized procedures with allowable accommodations.

The scoring information in Chapter 4 describes the steps taken to train and monitor hand-scorers, as well as quality control procedures related to scanning and machine scoring. To speak to student response processes, however, additional studies would be helpful and might include an investigation of students' cognitive methods using think-aloud protocols.

Evidence based on internal structure is presented in Chapters 6 through 8 in great detail in the discussions of item analyses, scaling and equating, and reliability. Technical characteristics of the internal structure of the assessments are presented in terms of classical item statistics (item difficulty, item-test correlation), differential item functioning analyses, dimensionality analyses, reliability, standard errors of measurement, and item response theory parameters and procedures. Each test is equated to the same grade and content area test from the prior year in order to preserve the meaning of scores over time. In general, item difficulty and discrimination indices were in acceptable and expected ranges. Very few items were answered

correctly at near-chance or near-perfect rates. Similarly, the positive discrimination indices indicate that most items were assessing consistent constructs, and students who performed well on individual items tended to perform well overall.

Evidence based on the consequences of testing is addressed in the scaled score information in Chapter 7 and the reporting information in Chapter 5. Each of these chapters speaks to the efforts undertaken to provide accurate and clear information to the public regarding test scores. Scaled scores offer the advantage of simplifying the reporting of results across grade levels and subsequent years. Achievement levels provide users with reference points for mastery at each grade level, which is another useful and simple way to interpret scores. Several different standard reports are provided to stakeholders. Additional evidence of the consequences of testing could be supplemented with broader investigation of the impact of testing on student learning.

9.1 QUESTIONNAIRE DATA

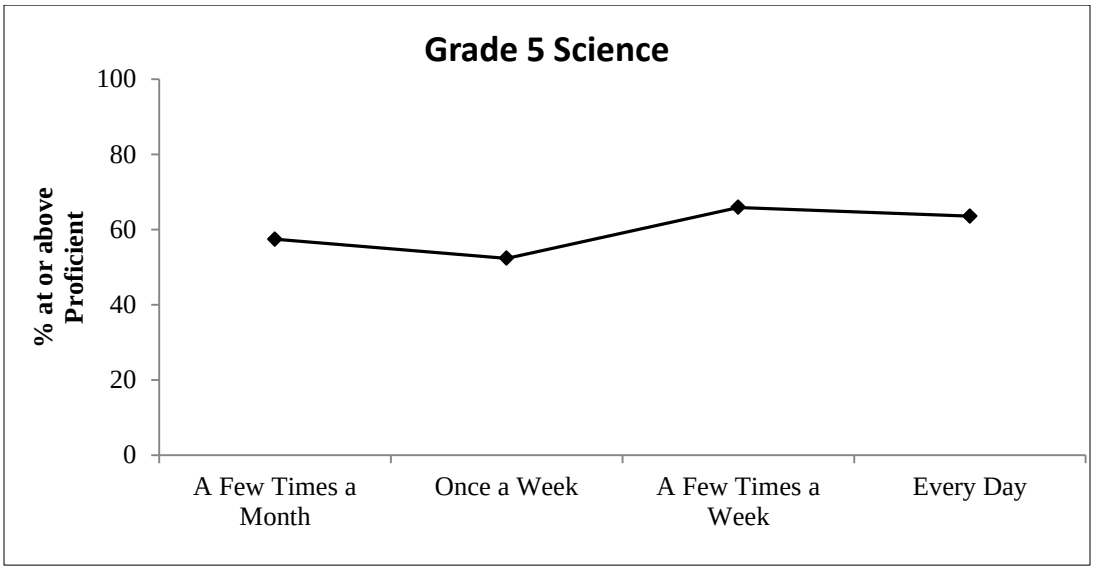
External validity of the 2013–14 MEA Science assessment is conveyed by the relationship of test scores and situational variables such as time spent patterns, self-image, and attitude toward content matter. These situational variables were all based on student questionnaire data collected during the administration of the 2013–14 MEA Science test. Note that no inferential statistics are included in the results presented below; however, because the numbers of students are quite large, differences in average scores may be statistically significant.

9.1.1 Time Spent

Examinees in grade 5 were asked how often they had science classes. Figure 9-1 shows that students who had science classes every day or a few times a week did better than those who had science class once a week or a few times a month, but the difference was slight.

Question: How often do you do science in class?

Figure 9-1. 2013–14 MEA Science: Questionnaire Responses—Time Spent Grade 5



9.1.2 Self-Image

Examinees in grades 5 and 8 were asked how they would rate themselves as a student in science. Figures 9-2 and 9-3 indicate a positive relationship between self-image as a student and MEA Science scores.

Question: Which of the following best describes how you rate yourself as a student in science?

Figure 9-2. 2013–14 MEA Science: Questionnaire Responses—Self-Image Grade 5

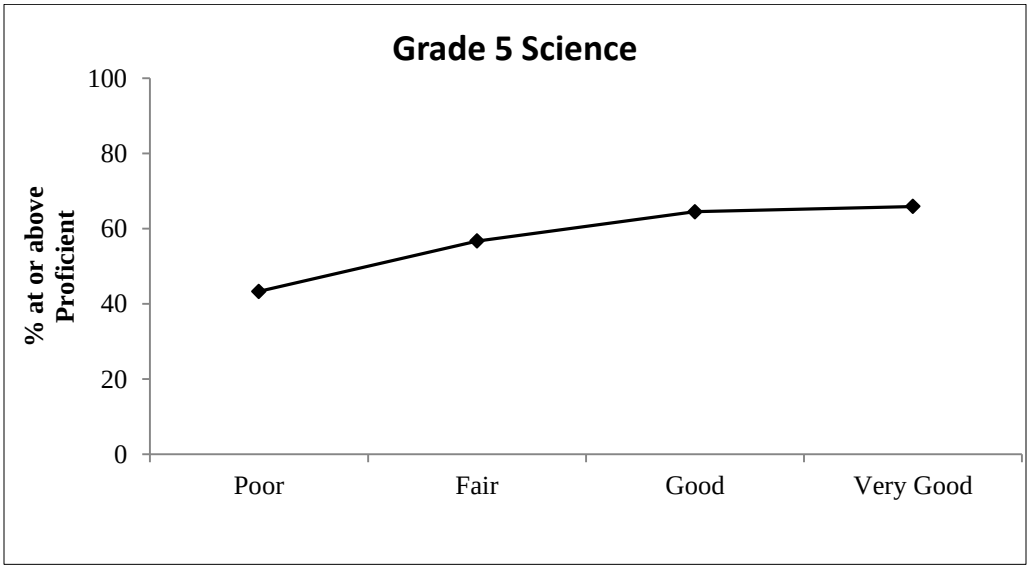
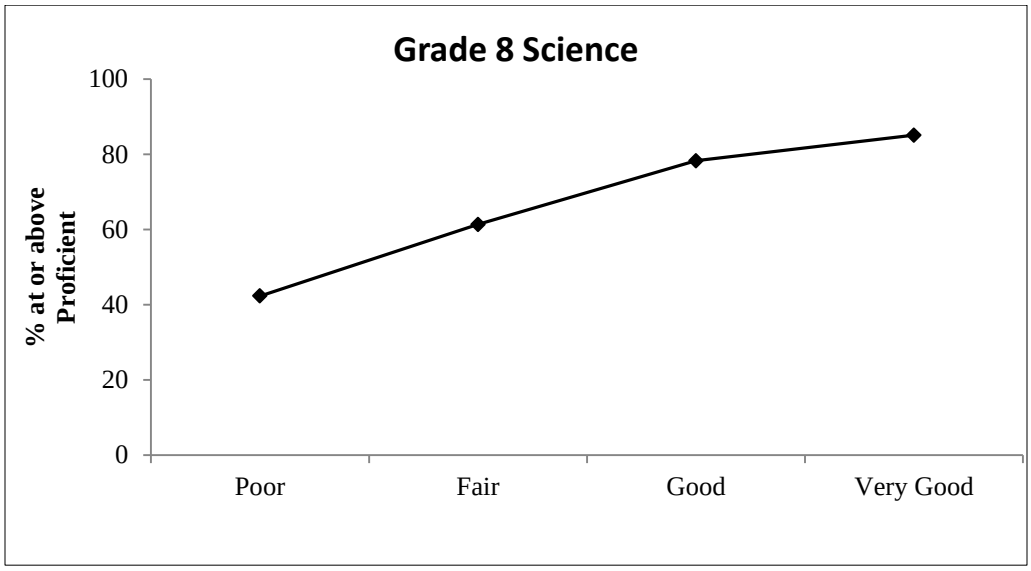


Figure 9-3. 2013–14 MEA Science: Questionnaire Responses—Self-Image Grade 8

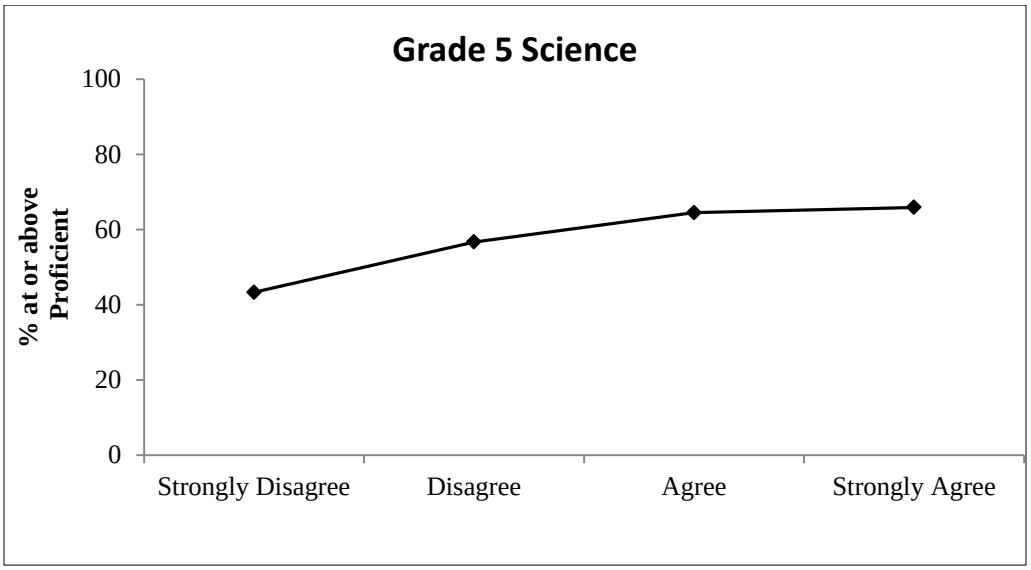


9.1.3 Attitude Toward Content Area

Questionnaire items related to examinees' attitudes toward science were administered to students in grades 5 and 8. For grade 5, students were asked how they felt about the statement "Science is interesting and fun." For grade 8, students were asked how they felt about the statement "My knowledge of science will be useful to me as an adult." Figures 9-4 and 9-5 indicate that students' attitudes toward science are related positively to MEA scores.

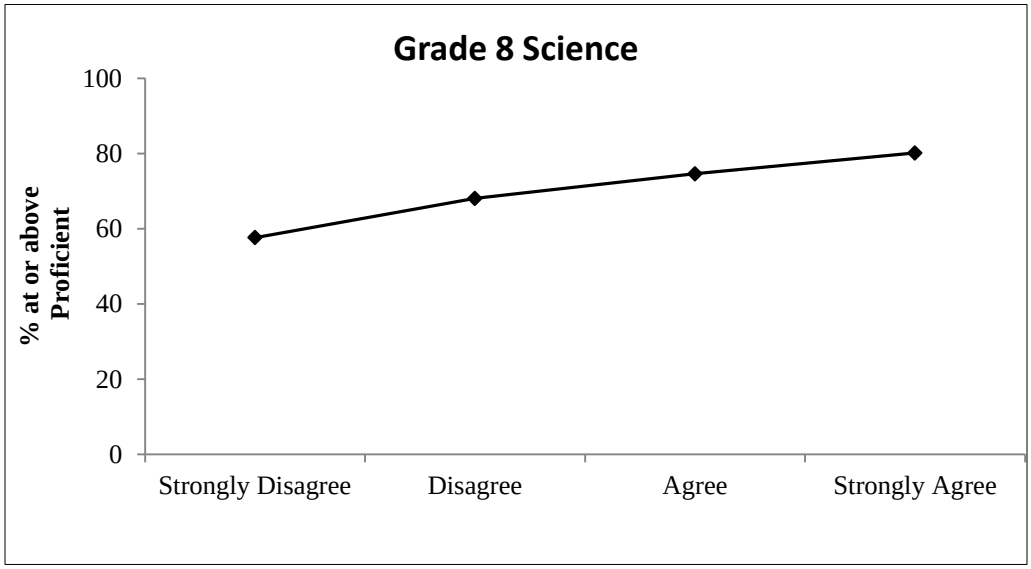
Question: How do you feel about the following statement? *Science is interesting and fun.*

Figure 9-4. 2013–14 MEA Science: Questionnaire Responses—Attitude Toward Content Grade 5



Question: How do you feel about the following statement? *My knowledge of science will be useful to me as an adult.*

Figure 9-5. 2013–14 MEA Science: Questionnaire Responses—Attitude Toward Content Grade 8



9.1.4 Match of Questions to What Is Learned in School

Students in grades 5 and 8 were asked how well the questions on the MEA Science test matched what they had learned in school about science. Figures 9-6 and 9-7 indicate that there is a positive relationship

between how well students feel the questions match what they have learned in science and MEA Science scores.

Question: How well do the questions that you have just been given on this MEA test match what you have learned in school about science?

Figure 9-6. 2013–14 MEA Science: Questionnaire Responses—Match of Questions to What Is Learned in School Grade 5

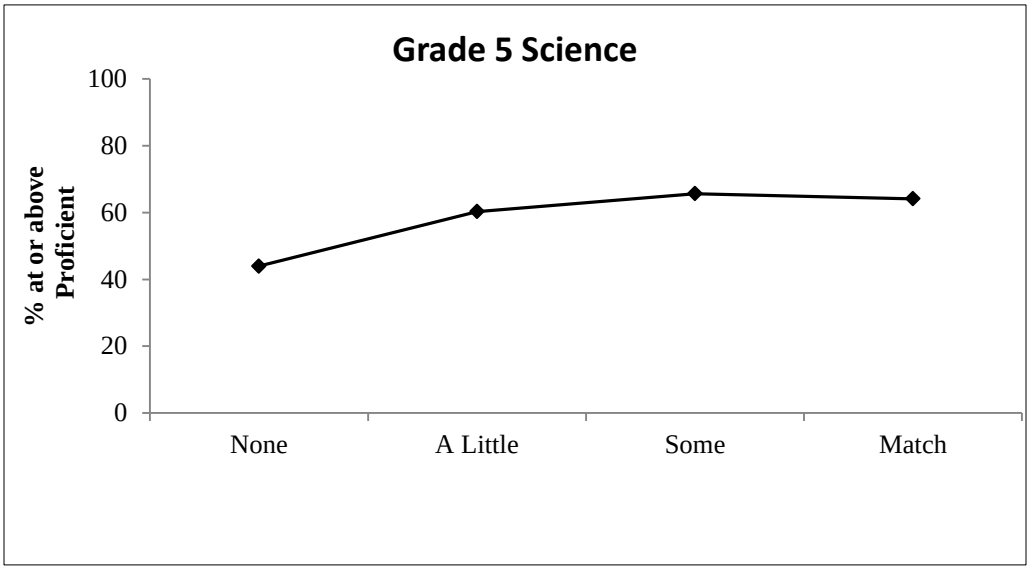
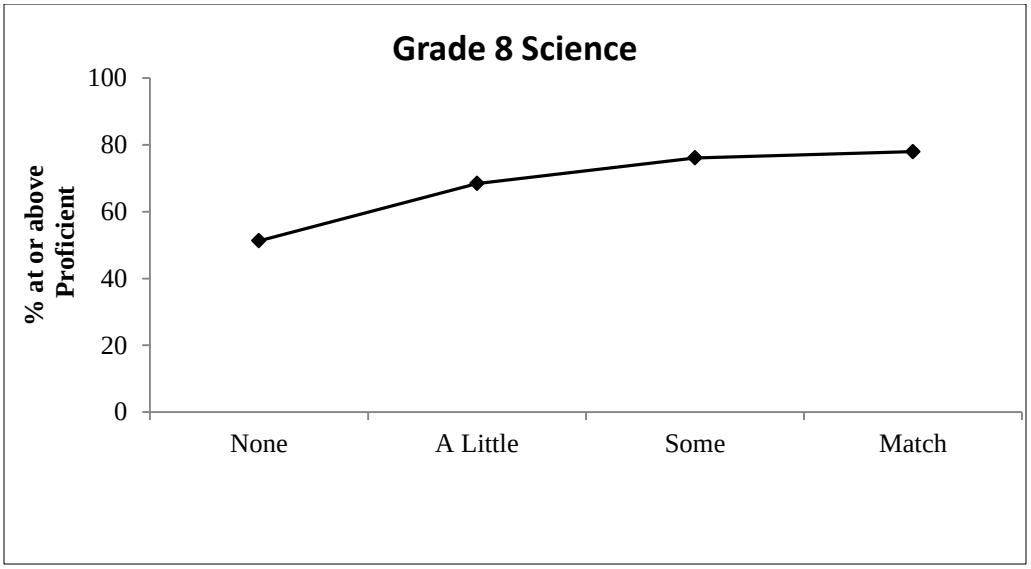


Figure 9-7. 2013–14 MEA Science: Questionnaire Responses—Match of Questions to What Is Learned in School Grade 8



9.1.5 Difficulty of Assessment

For grades 5 and 8, students were asked how difficult they found the test. Figures 9-8 and 9-9 indicate that there is a modest negative relationship between how difficult the students felt the items were and overall MEA Science scores (i.e., students who found the test more difficult received lower scores than students who found the test easier). The pattern is more consistent for grade 8 than for grade 5.

Question: How difficult was this science test?

Figure 9-8. 2013–14 MEA Science: Questionnaire Responses—Difficulty of Assessment Grade 5

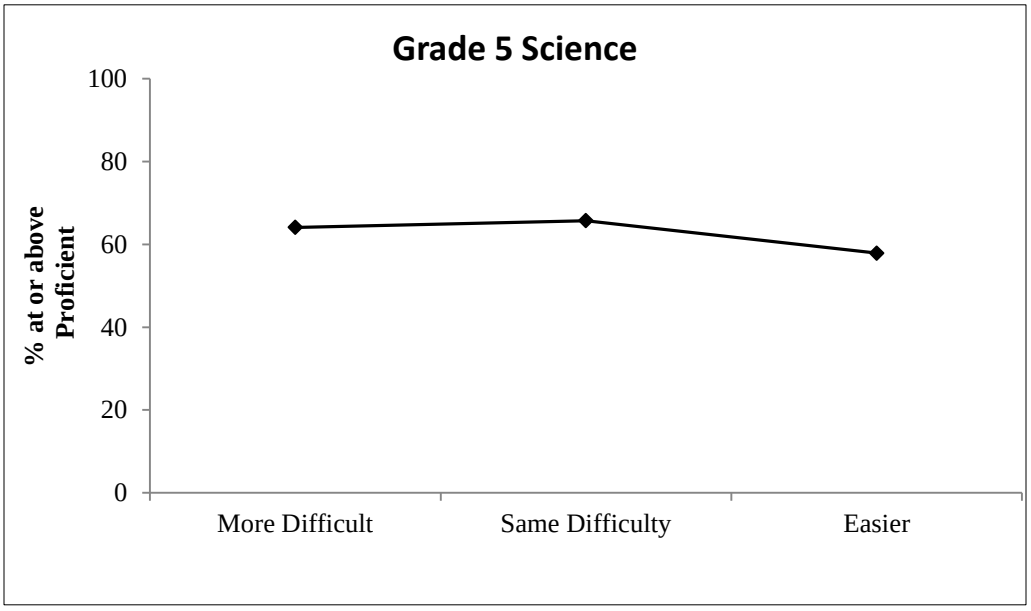
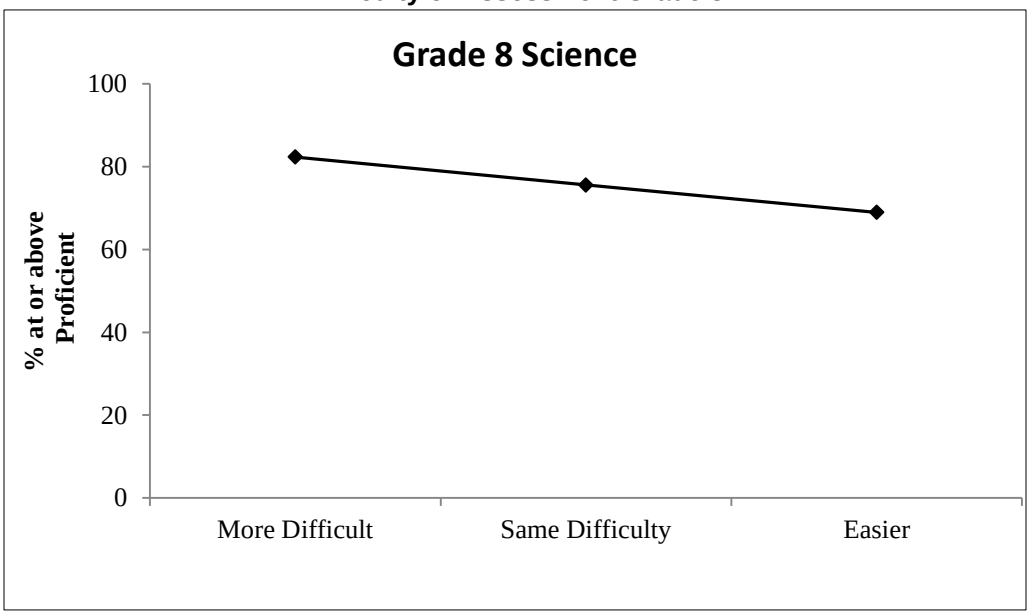


Figure 9-9. 2013–14 MEA Science: Questionnaire Responses—Difficulty of Assessment Grade 8



The evidence presented in this report supports inferences of student achievement on the content represented in *Maine Learning Results* and grade level expectations for science for the purposes of program and instructional improvement and as a component of school accountability.

REFERENCES

Allen, M. J., & Yen, W. M. (1979). *Introduction to measurement theory.* Belmont, CA: Wadsworth, Inc.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. Washington, D.C.: American Educational Research Association.

Baker, F. B., & Kim, S-H. (2004). *Item response theory: Parameter estimation techniques* (2nd ed.). New York: Marcel Dekker, Inc.

Brown, F. G. (1983). *Principles of educational and psychological testing* (3rd ed.). Fort Worth: Holt, Rinehart, and Winston.

Chicago Manual of Style (15th ed.). (2003). Chicago: University of Chicago Press.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, *20*, 37–46.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*, 297–334.

Dorans, N. J., & Holland, P. W. (1993). DIF detection and description. In P. W. Holland & H. Wainer (Eds.) *Differential item functioning* (pp. 35–66). Hillsdale, NJ: Lawrence Erlbaum Associates.

Dorans, N. J., & Kulick, E. (1986). Demonstrating the utility of the standardization approach to assessing unexpected differential item performance on the Scholastic Aptitude Test. *Journal of Educational Measurement*, *23*, 355–368.

Draper, N. R. & Smith, H. (1998). *Applied regression analysis* (3rd ed.). New York: John Wiley and Sons, Inc.

Hambleton, R. K., & Swaminathan, H. (1985). *Item response theory: Principles and applications*. Boston, MA: Kluwer Academic Publishers.

Hambleton, R. K., Swaminathan, H., & Rogers, J. H. (1991). *Fundamentals of item response theory*. Newbury Park, CA: Sage Publications, Inc.

Hambleton, R. K., & van der Linden, W. J. (1997). *Handbook of modern item response theory*. New York, NY: Springer-Verlag.

Joint Committee on Testing Practices. (2004). *Code of fair testing practices in education*. Washington, D.C.: National Council on Measurement in Education.

Livingston, S. A., & Lewis, C. (1995). Estimating the consistency and accuracy of classifications based on test scores. *Journal of Educational Measurement*, *32*, 179–197.

Lord, F. M., & Novick, M. R. (1968). Statistical theories of meta test scores. Reading, MA: Addison-Wesley.

Muraki, E. & Bock, R. D. (2003). *PARSCALE 4.1*. Lincolnwood, IL: Scientific Software International.

Petersen, N. S., Kolen, M. J., & Hoover, H. D. (1989). Scaling, norming, and equating. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 221–262).

Stocking, M. L., & Lord, F. M. (1983). Developing a common metric in item response theory. *Applied Psychological Measurement, 7,* 201-210.

Stout, W. F. (1987). A nonparametric approach for assessing latent trait dimensionality. *Psychometrika, 52,* 589–617.

Stout, W. F., Froelich, A. G., & Gao, F. (2001). Using resampling methods to produce an improved DIMTEST procedure. In A. Boomsma, M. A. J. van Duign, & T. A. B. Snijders (Eds.), *Essays on item response theory* (pp. 357–375). New York: Springer-Verlag.

Zhang, J., & Stout, W. F. (1999). The theoretical DETECT index of dimensionality and its application to approximate simple structure. *Psychometrika, 64,* 213–249.

APPENDICES

APPENDIX A—MAINE TECHNICAL ADVISORY COMMITTEE MEMBERS

Table A-1. 2013–14 MEA Science: Technical Advisory Committee Members

Member Name	*Member Affiliation*
Brian Gong	Executive Director, National Center for Improvement of Educational Assessment
Lenora Murray	Assistant Superintendent, MSAD #49
Stephen Slater	Assistant Director of Assessment, Oregon Department of Education
Betsy Webb	Superintendent of Schools, Bangor Public Schools
Martha Thurlow	Director, NCEO/University of Minnesota

APPENDIX B – ITEM REVIEW AND BIAS AND SENSITIVITY REVIEW COMMITTEE MEMBERS

MEA 2013–14 Science Bias and Sensitivity Committee Members

Name	*Department*
Lynne Adams	Augusta School Department (SPED)
Judy Carey	Catholic Charities (Ed Services for Blind/Visually Impaired Children)
Melvin Curtis	Retired (SPED)
Julia O'Brien-Merrill	Retired (ESL)
Rebecca Perez	Mountain Valley Middle School (ESL Teacher)

MEA 2013-14 Item Review Committee Science Grade 5

Name	*School*
Jim Chandler	Auburn Land Lab
Laurette Darling	Albert S Hall School
Sheree Granger	The School at Sweetser, Saco
Cecilia Joyce	East End Community School
Bethann Montpetit	MSAD #32 Ashland
Sarah Otterson	Hebron Station School
Maria Pololi	Harriet Beecher Stowe Elementary School
Nancy Philbrick	MSAD #17 Oxford
Pamela Thompson	MSTA Board

MEA 2013-14 Item Review Committee Science Grade 8

Name	*School*
Diana Allen	Sanford Junior High School
Barbara Benjamin-McManus	Lewiston Middle School
Patricia Bernhardt	James F. Doughty School
Jim Chandler	Auburn Land Lab
Denise Friant	Woolwich Central School
Ricia Hyde	Waterville Junior High
Elizabeth Ladner	Hall-Dale Middle School
Tracy Vassiliev	James F. Doughty School

APPENDIX C—PARTICIPATION RATES

Table C-1. 2013–14 MEA Science: Summary of Participation by Demographic Category—Science

Description	*Number Tested*	*Percent*
All Students	26,453	100.00
Male	13,547	51.21
Female	12,905	48.78
Gender not reported	1	0.00
Hispanic or Latino	458	1.73
American Indian or Alaskan Native	200	0.76
Asian	393	1.49
Black or African American	788	2.98
Native Hawaiian or Pacific Islander	29	0.11
White (non-Hispanic)	24,193	91.46
Two or More Races (non-Hispanic)	391	1.48
Race not reported	1	0.00
Currently receiving LEP services	733	2.77
Former LEP student – monitoring year 1	73	0.28
Former LEP student – monitoring year 2	52	0.20
LEP: All Other Students	25,595	96.76
Students with an IEP	4,381	16.56
IEP: All Other Students	22,072	83.44
Economically Disadvantaged Students	11,843	44.77
SES: All Other Students	14,610	55.23
Migrant Students	8	0.03
Migrant: All Other Students	26,445	99.97
Students receiving Title 1 Services	2,984	11.28
Title 1: All Other Students	23,469	88.72
Plan 504	1,104	4.17
Plan 504: All Other Students	25,349	95.83

APPENDIX D—ACCOMMODATION FREQUENCIES

Table D-1. 2013–14 MEA Science: Numbers of Students Tested with Accommodations by Accommodation Type and Grade

Accommodation Code	*Grade 5*	*Grade 8*
sciaccomT1	1,384	838
sciaccomT2	63	48
sciaccomT3	708	391
sciaccomT4	75	34
sciaccomS1	1,670	905
sciaccomS2	28	13
sciaccomP1	341	166
sciaccomP2	1,950	1,424
sciaccomP3	1,279	505
sciaccomP4	172	176
sciaccomP5	1,204	900
sciaccomP6	6	7
sciaccomP7	895	464
sciaccomP8	13	11
sciaccomP9	1	2
sciaccomP10	9	17
sciaccomP11	4	4
sciaccomR1	594	219
sciaccomR2	54	15
sciaccomR3	43	15
sciaccomR4	24	71
sciaccomR5	42	8
sciaccomR6	4	3
sciaccomR7	9	14
sciaccomO1	0	2
sciaccomM3	0	1

APPENDIX E—ACCOMMODATIONS GUIDE

Maine Educational Assessment - Science

Accommodations Guide

Table of Contents

Foreword

The No Child Left Behind (NCLB) Act mandates that all students be included in state science assessments. In Maine, all students in grades 5 and 8 will participate in the MEA science assessment through one of three avenues: standard administration, administration with accommodations, or through alternate assessment.

This *MEA-Science Accommodations Guide* is a version of the *NECAP Accommodations Guide* which has been edited for use in the MEA-Science test and is provided to ensure consistent accommodations information for testing in grades 3-8 in Maine. It is intended to supplement test administration information contained in the *MEA Principal Test Coordinator Manual* and grade-specific *Test Administrator Manuals* produced for the MEA science test. Legal requirements for students identified for federally funded programs have been taken into account in the development of this document.

Please note that any references to science in the *NECAP Accommodations Guide* do not necessarily apply to MEA science testing. This MEA science-specific Accommodations Guide should be used for MEA testing.

The Maine Department of Education appreciates the collaboration among all the New England Common Assessment Program states as well as the permission of the other NECAP states and the Center for Assessment (NCIEA) in allowing the editing of this document for use in Maine's MEA science test.

Contact Information

If you have any questions regarding the accommodations allowed in the Maine Educational Assessment in science, contact the following staff:

Maine Department of Education

Susan Fossett, Assessment Coordinator, 207-624-6775 susan.fossett@maine.gov

Measured Progress
If you have any questions regarding materials or administering the MEA, contact the Measured Progress MEA Test Administration Helpdesk at (866) 615-2745.

Section 1: Purpose of the Accommodations Guide

This guide is supplementary to the Prinicipal/Test Coordinator Manual and the Test Administrator Manuals. It is to be used in conjunction with these manuals, which include overall administration information for all students.

It is important to remember that this guide also contains information that is relevant for all students. On the MEA tests, standard test accommodations are allowable for all students if approved by a local team. In order to make sure that all students are given access to test accommodations that they may need during state testing, it is vital for test administrators to become familiar with allowable accommodations and the appropriate ways to administer them.

The MEA Accommodations Guide has been created to…

- ensure a standardized accommodation administration process is followed.
- help schools make appropriate accommodation choices for students and understand what, if any, consequences are attached to their decisions.
- help schools provide students with the best opportunity to show what they know within the state testing environment.
- help building administrators and test coordinators train school personnel involved in the administration of accommodations.

REMINDER
Testing accommodations should be made for individual students; they are not designed for use with entire classrooms. They should not give students unfair advantages. Rather they are meant to remove barriers that may exist due to a student's learning style or disability. Students should have had experience using an accommodation during routine instruction and/or test-taking, as appropriate, in the classroom prior to its use on the MEA.

Section 2: Introduction to MEA-Science Accommodations

All students are eligible to utilize appropriate assessment accommodations listed in the Table of Standard Test Accommodations when participating in the MEA. Any accommodation(s) utilized for the assessment of individual students must be:

- the result of a decision made by the IEP, Section 504, or other school team that includes, whenever possible, the student's parent(s) or guardian(s). Under most circumstances accommodation decisions should be made prior to testing. There always will be emergencies and unforeseen cases that require an accommodation decision to be made during the test administration window. These cases should be rare and should not preclude the accommodation decision being made by an appropriate school team,
- based on the individual student's needs,
- consistent with those accommodations used during the student's regular classroom instruction, including test-taking and, if applicable, consistent with the student's IEP or 504 Plan, and
- documented at the appropriate local level.

Test accommodations are changes in setting, timing (including scheduling), presentation format, or response format that do not alter in any significant way what the test measures or the comparability of results. When used properly, appropriate test accommodations remove barriers to participation in the assessment and provide students with diverse learning needs an equitable opportunity to demonstrate their knowledge and skills.

Most students who need accommodations can be supported successfully by using one or more of the standard accommodations listed in the timing (T), setting (S), presentation (P), and response (R) categories on the MEA Table of Standard Test Accommodations.

The accommodations included in the *Table of Standard Test Accommodations* are based on research, best practice, and educators' experiences administering the MEA tests over several years. However, the *Table of Standard Test Accommodations* is not an exhaustive list of allowable accommodations. Individual students may have particular needs that require the use of an accommodation not included in the *Table of Standard Test Accommodations*. In such cases, school personnel may contact the Department of Education to request verification of the comparability of a proposed accommodation using the procedures described under Other Accommodations on page 18.

Section 3: Making Accommodation Decisions

Decisions about the use of accommodations should **not** be made by an individual. Decisions should be made by the school team responsible for planning the student's academic program and should include parent participation. It is recommended that every school identify a process to determine how accommodation decisions are made for students who do not have IEPs or 504 plans. Many schools already have Student Support or Child Study Teams in place for the purpose of addressing the individual needs of students in general education.

The role of the team is to discuss the accommodations that a student may need for MEA testing, decide which accommodations will be used by the student, and document the process.

These teams would typically involve the following individuals:

For students in general education:

- Educators involved in supporting the student
- Parent(s) and/or guardian(s)
- Student (as appropriate)

For students with 504 Plans, the student's existing 504 Team, including:

- Educators involved in supporting the student
- Parent(s) and/or guardian(s)
- Student (as appropriate)

For students with disabilities, the student's existing IEP Team, including:

- Special and general educators involved in supporting the student
- Parent(s) and/or guardian(s)
- Student (as appropriate)

For students who are English language learners (ELL):

- Teachers of English language learners or bilingual students and general educators involved in supporting the student
- Interpreter (as appropriate)
- Parent(s) and/or guardian(s)
- Student (as appropriate)

Assessment Supports that are not Accommodations

When making decisions about accommodation use for an individual student, the team should be aware of the many assessment supports allowable for **all students** during MEA testing that are not considered accommodations. Although these supports are not considered accommodations and their use does not need to be recorded on the Student Answer Booklet, thoughtful planning and preparation for the student's test-taking experience should include consideration of these available supports including how and when they can be used appropriately. General test supports include the basic minimum conditions that should be provided to create an appropriate testing environment. Some limited individualization is permitted within these general supports.

The following are examples of assessment supports that are not accommodations:

Before Testing, the Test Administrator:

- Provides students with MEA released items/practice test experience that is appropriate to their tested grade level if school personnel believe this will help the students to become familiar with the test format and test taking procedures.
- Encourages and reminds students to get a good night's rest and a healthful breakfast prior to test administration to help them focus and give their best effort.
- Is familiar with (and to) the students who are being tested.
- Schedules group test sessions during the normal school day with distractions minimized.

During Testing, the Test Administrator:

- Provides a quiet, appropriate group testing space equipped with testing materials students will need. As needed, provides appropriate furniture and placement, including as appropriate, preferential seating for certain students, or assigns all student seating. Study carrels may be assigned to minimize testing distractions, if needed.
- As needed, provides motor and balance supports, special individualized pencil grips, positioning or balance equipment, wheelchairs, walkers, or occupational or physical therapy supports that <u>permit motor movement</u> enabling a student to interact with the test are allowed and are not treated as a test accommodation - provided that they do not provide any academic function for the student.
    - Note: In general, assistive communication technology devices ARE treated as accommodation supports. For more information, refer to the *Table of Standard Test Accommodations* or contact the Department of Education.
- Allows individually prescribed corrective lenses, glasses, or hearing aids without accommodation. For other special purpose auditory or visual supports and aids, see *Table of Standard Test Accommodations*.
- Provides the scripted general test directions (from the test administrator's manual) to all students,
    - **During the Introduction to the Session, Test Administrators may:**
        A. <u>repeat</u> <u>scripted directions</u> for students as needed, and
        B. <u>clarify only the scripted directions (general procedures)</u> that are read to the <u>entire class.</u>

    - **During Testing, Test Administrators may:**
        answer questions about the <u>*very few* test navigation directions</u> found inside the test booklet (such as: 'Mark your answer to number 15 on page 4'; 'Go on to the next page'; or 'Stop'.)

- Provide active, 'walk-around' proctoring during testing, to ensure that students remain productively on task and focused
- May, upon student request, pronounce single words
    - Students may ask the test administrator to pronounce single words they do not recognize. Test administrators may pronounce single written words in English
    - ***Note**:* Test administrators should not read entire sentences as a general support. This level of support should be treated as an accommodation. Support that involves reading entire sentences to the student must be determined and planned by the educational team before testing, consistent with routine practice, and documented appropriately as either an accommodation or a modification, depending upon tested content area.

During Testing, a Student may:

- Circle, underline, or mark text in the Student Test Booklet that he or she finds to be important to him or her during the test. This is a test taking strategy and not an accommodation *if the student is doing this independently*
- Use post-it notes to flag test session stop signs they must not go beyond
    - ***Note:** It is very important that any p*ost-it notes or similar inserted material be *removed before* the booklet is returned for scoring.
- Use a plain (non-ruler) straight edge (e.g., scrap paper) to assist with visual tracking while they read.

After Testing, a Test Administrator may:

- Permit appropriate reading for those students who finish early while others are still working. This must be planned before testing.

❖ **General Note:** Many supports that exceed the limits of the procedures described above are considered to be formal accommodations and are included in the *Table of Standard Accommodations*. However, if a procedure provides so *much* support that it prevents the student from demonstrating the construct being tested, then it changes the very academic content the test is attempting to measure, and for this reason it will be treated as a modification. It is very important to understand the difference between accommodations and modifications. Please read this manual completely and, if you need more information, contact the Department of Education assessment office.

The Appropriate Use of an Accommodation: Finding the Balance

There is an important distinction between instructional accommodations and testing accommodations. Supports provided at the beginning of the instructional process are designed to help students' first experience, learn, and practice a new skill. The long term purpose of **instructional accommodations** or other early supports is to ultimately help the student *learn to become as fluent and as independent as possible* in performing that skill. For this reason, instructional accommodations should incorporate a scaffolded *fading process* that provides much more support early in the learning process as skill acquisition is just beginning. Later in the instructional process the need for early levels of support should be challenged or tested to see how much control can be assumed by the student. The intensive supports used very early in instruction may at times greatly simplify or may even *modify* the skill the student is learning helping to guide, shape, and successively approximate the student's behavior to ensure that he or she experiences some early success while moving closer to real skill performance. When

planning instructional supports, the path to student independence must always be kept in mind. Plan with the end in mind, always move toward independence.

As effective instruction continues, early intensive supports (or modifications) are faded, allowing the student to demonstrate the academic skill with increasing independence. As higher levels of skill independence are achieved, supports are faded back further still until **the least intrusive accommodation** or, perhaps even full independence is achieved. The least intrusive accommodation is the level of support that will allow the student to demonstrate the skill in the most independent manner possible for that student.

For example, if a student has a certain type of visual processing difficulty, he or she may need (for some years) to use a straight edge to guide visual tracking while reading, but eventually learns to perform the actual reading task with full independence to the extent of his or her capability. At the point of testing, this student no longer has a person holding the tracking tool or reading the passage to him or her. This has become the independent responsibility of the student; yet remaining student needs for support are still being met. Independent use of the visual tracking tool has become the least intrusive accommodation for the student at this point.

Testing accommodations should be those accommodations that *are the least intrusive accommodations possible* to meet the needs of the student while *allowing the maximum level of independence possible* for that student. They represent the current balance point the instructional fading process has achieved. Testing accommodations, therefore, represent the highest point of independent skill acquisition that has been achieved with that student *to date* through the instructional process. Testing accommodations do not necessarily represent the instructional end point, but they do represent a point in time that lies beyond the earliest phases of skill acquisition. Some skill independence should be seen if instruction has been effective. Teams must remember to carefully consider *long term independence* and thoughtfully design the process of fading supports when choosing and planning instructional methods.

The key is finding the right balance of supports for a given student and actively, consistently, and constructively supporting the growth of student independence.

Section 4: Standard Test Accommodations

This section of the Accommodations Guide contains a copy of the Table of Standard Test Accommodations and a discussion of each of the six categories of accommodations in the table: Timing (T), Setting (S), Presentation (P), Response (R), Other Accommodations (O), and Modifications (M). For each category, there is a list of the standard accommodations, an overview of the category, explanations of the use of selected accommodations as needed, and examples and procedures to be followed as needed.

The Table of Standard Test Accommodations is a list of accommodations that are available to all students on an individual basis, regardless of disability status. Following procedures previously outlined in Section 3 (Making Accommodation Decisions), school teams will refer to the Tables of Standard Test Accommodations when making decisions for MEA testing.

All accommodations must be recorded by the test administrator on the Student Answer Booklet. Please be sure to bubble in only those accommodations that the student actually used for the MEA test for each content area in which they were used.

It is important to note that if the team believes a student needs an accommodation that is not listed, the school must contact appropriate personnel at the Department of Education to discuss the proposed accommodation (see contact information on page ii). The approval process and any consequences that result from the use of that accommodation will be discussed.

MEA SCIENCE ACCOMMODATIONS TABLE

Revised December 2010

NOTE: To ensure consistency across Maine's grades 3-8 testing program, the same accommodation codes are used in both the MEA-Science and NECAP tests (as applicable).The chart below contains accommodations and codes to be used for the MEA-Science test and should be bubbled on the student's answer booklet after testing.

Any accommodation(s) used for the assessment of an individual student will be the result of a team decision made at the local level. All decisions regarding the use of accommodations must be made on an individual student basis – not for a large group, entire class, or grade level. Accommodations are available to all students on the basis of individual need regardless of disability status and should be consistent with the student's normal routine during instruction and assessment. This table is not intended to be used as a stand-alone document and should always be used in conjunction with the MEA *Accommodations Guide* and/or Test Administrator manuals.

T. Timing		
Code	Tests were administered	Details on Delivery of Accommodations
T1	with time to complete a session extended beyond the scheduled administration time within the same day.	The scheduled administration time already includes additional time and the vast majority of students complete the test session within that time period. Extended time within a single sitting may be needed by students who are unable to meet time constraints. A test session may be extended until the student can no longer sustain the activity.
T2	so that only a portion of the test session was administered on a particular day.	In rare and severe cases, the extended time accommodation (T1) may not be adequate for a student not able to complete a test session within a single day. A test session may be administered to a student as two or more "mini-sessions" if procedures are followed to maintain test security and ensure that the student only has access to the items administered on that day (see the *MEA Accommodations Guide* for details).
T3	with short, supervised breaks.	Multiple or frequent breaks may be required by a student whose attention span, distractibility, or physical condition, requires shorter working periods.
T4	at the time of day or day of week that takes into account the student's medical needs or learning style.	Individual scheduling may be used for a student whose school performance is noticeably affected by the time of day or day of the school week on which it is done. This accommodation may not be used specifically to change the order of administration of test sessions. This accommodation must not result in the administration of a test session to an individual student prior to the regularly scheduled administration time for that session for all students.

S. Setting		
Code	Tests were administered	Details on Delivery of Accommodations
S1	in a separate location within the school by trained school personnel.	A student or students may be tested individually or in small groups in an alternative site within the school to reduce distractions for themselves or others, or to increase physical access to special equipment.

S2	in an out-of-school setting by trained school personnel.	Out-of-school testing may be used for a student who is hospitalized or tutored because they are unable to attend school. The test must be administered by trained school personnel familiar with test administration procedures and guidelines. Relatives/guardians of the student may not be used as the test administrator.

P. Presentation		
Code	Tests were administered	Details on Delivery of Accommodations
P1	individually.	Individual or small group testing may be used to minimize distractions for a student or students whose test is administered out of the classroom or so that others will not be distracted by other accommodations being used (e.g., dictation)
P2	in a small group.	
P3	with test and directions read aloud in English or signed to the student.	A reader may be used for a student whose inability to read would hinder performance on the Science test. Words must be read as written. No translations (with the exception of signed language) or explanations are allowed. Trained personnel may use sign language to administer the test.
P4	with only test directions read aloud or signed to the student.	A reader may be used for a student whose inability to read or locate directions would hinder performance on the test. Note that most directions on the MEA test occur at the beginning of the test session and are already read aloud by the test administrator. Guidelines for what are and are not "test directions" must be followed. With the exception of sign language and the case of students enrolled in a program where the test administrator routinely presents information in a foreign language, directions may not be translated.
P5	with administrator verification of student understanding following the reading of test directions.	After test directions have been read, the test administrator may ask the student to explain what he/she has been asked to do. If directions have been misunderstood by the student, the test directions may be paraphrased or demonstrated. Test items MUST NOT be paraphrased or explained.
P6	using alternative or assistive technology that is part of the student's communication system.	The test may be presented through his/her regular communication system to a student who uses alternative or assistive technology on a daily basis.
P7	by trained school personnel known to the student other than the student's classroom teacher.	A student may be more comfortable with a test administrator who works with the student on a regular basis, but is not the student's regular teacher for the general curriculum or other staff assigned as test administrator. All test administrators must be trained school personnel familiar with test administration and accommodations procedures and guidelines.
P8	using a large-print version of assessment.	Both large-print and Braille versions of the assessment require special preparation and processing and must be pre-ordered. Directions for ordering these materials are included in communications sent to school principals prior to the test.
P9	using Braille version of assessment.	
P10	using a word-to-word translation dictionary for ELL students.	A student with limited English proficiency may have a word-to-word dictionary available for individual use as needed. A word-to-word dictionary is one that does not include any definitions. Information on acceptable dictionaries is provided on the departments' websites.
P11	using visual or auditory supports.	The test may be presented using visual aids such as visual magnification devices, reduction of visual print by blocking or other techniques, or acetate shields; or auditory devices such as special acoustics, amplification, noise buffers, whisper phones, or calming music.

R. Response		
Code	Tests were administered	Details on Delivery of Accommodations
R1	with a student dictating responses to school personnel.	A student may dictate answers to constructed-response or short-answer questions to locally trained personnel or record oral answers in an individual setting so that other students will not benefit by hearing answers or be otherwise disturbed. Policies regarding recorded answers must be followed prior to returning test materials.
R2	with a student dictating responses using alternative or assistive technology/devices that are part of the student's communication system.	Technology is used to permit a student to respond to the test. Policies regarding recorded answers must be followed prior to returning test materials.
R3	with a student using approved tools or devices to minimize distractions.	Noise buffers, place markers, carrels, etc. may be used to minimize distractions for the student. This accommodation does NOT include assistive devices such as templates, graphic organizers, or other devices intended specifically to help students organize thinking or develop a strategy for a specific question.
R4	with a student writing responses using separate paper, a word processer, computer, brailler, or similar device.	A student may use technological or other tools (e.g., large-spaced paper) to write responses to constructed-response items. A key distinction between this accommodation and R2 is that the student using this accommodation is responding in writing rather than dictating. When using a computer, word processing device, or other assistive technology, access to the Web must be turned off. This accommodation is intended for unique individual needs, not an entire class. Policies regarding recorded answers must be followed prior to returning test materials.
R5	with a student indicating responses to multiple-choice items to school personnel.	A student unable to write or otherwise unable to fill in answers to multiple-choice questions may indicate a response to trained school personnel. The school personnel records the student's response in the student answer booklet.
R6	with a student responding with the use of visual aids.	Visual aids include any optical or non-optical devices used to enhance visual capability. Examples include magnifiers, special lighting, markers, filters, large-spaced paper, color overlays, etc. An abacus may also be used for student with severe visual impairment or blindness on the Science tests. Note that the use of this accommodation still requires student responses to be recorded in a student answer booklet.
R7	with a student with limited English proficiency responding with use of a word-to-word dictionary.	A student with limited English proficiency may have a word-to-word dictionary available for individual use as needed when responding. A word-to-word dictionary is one that does not include any definitions. Information on acceptable dictionaries is provided on the Department's website.

O. Other This accommodation requires DOE approval or no credit will be given.		
Code	Tests were administered	Details on Delivery of Accommodations
O1	using other accommodation(s) not on this list, requested by the accommodations team.	An IEP team or other appropriate accommodation team may request that a student be provided an accommodation not included on this standard list of accommodations. Like all other accommodations, these should be consistent with the student's normal routine during instruction and/or assessment. Requests should be made to the DOE when accommodation plans are being made for a student prior to testing. DOE approval must be received for the requested accommodation to be coded as an O1 accommodation. Non-approved accommodations used during test administration will be coded as an M3 modification.

M. Modification This modification results in no credit being given.		
Code	Tests were administered	Details on Delivery of Accommodations
M3	using an accommodation on this list not approved for a particular test or an accommodation not included on this list without prior approval of the DOE.	Inappropriate use of an accommodation included on this list or use of another accommodation without prior approval of the DOE will result in impacted items being scored as incorrect.

Note: English Language Learners may qualify for any of the accommodations listed as appropriate and determined by a team. Refer to the *MEA-Science Accommodations Guide* for additional information.

Timing (T)

T1. With time to complete a session extended beyond the scheduled administration time within the same day.
T2. So that only a portion of the test session was administered on a particular day
T3. With short, supervised breaks
T4. At the time of day that takes into account the student's medical needs or learning style.

Considerations for Timing accommodations

Overall:

- **Students must be supervised at all times during an active testing session, including breaks.**
- **Timing accommodations may not be used specifically to change the required order of administration of test sessions.**

T1 With time to complete a session extended beyond the scheduled administration time within the same day.

- All students are given additional time on the MEA tests. The guidelines in the Principal/Test Coordinator and Test Administrator Manuals indicate the amount of time that must be scheduled for each test session. The scheduled time has been calculated to provide students sufficient additional time to complete the test session beyond the time that the vast majority of students will require to complete the test session. Completing the test within the scheduled time for the test session should not be marked as an accommodation.
- The extended time accommodation should be used for students who routinely take one third or more of the allotted time to complete class projects and tests. For other students, the decision to use this accommodation should include consideration of the amount of additional time built into the scheduled time for the test session as well as the student's normal timing requirements. The decision to use this accommodation may also require the use of an alternative setting accommodation.
- Refer to the Principal/Test Coordinator Manual for a complete discussion of the amount of additional time that has been built into the required time to be scheduled for a particular test and test session.

T2 So that only a portion of the test session was administered on a particular day

- The use of this accommodation should be **very rare and limited to severe cases** in which even with the use of other accommodations such as extended time (T1) and short, supervised breaks (T3), a student would be unable to complete a test session within a single day. Use of this accommodation requires the test administrator to ensure that the student only has access to the set of test items that will be completed on a particular day in order to maintain test security. It is likely that use of this accommodation will also require an individual administration (P1) in which a test administrator can closely monitor the student. Prior to use of this accommodation schools should contact the Department of Education to discuss appropriate strategies for its use on a particular test or test session.

- This accommodation is **NOT** intended for students who begin a test session and are unable to complete it because they become ill or must be removed from the testing environment for some other reason.

T3 With short, supervised breaks

- Be sure students who need frequent breaks (T3) are supervised during these breaks. They should not be allowed opportunities to interact with other students. Recess and lunch may not be used as breaks during a single testing session. Sessions must be completed within the length of the day.

Setting (S)

S1. In a separate location within the school by trained school personnel
S2. In an out-of-school setting by trained school personnel

Considerations for Setting accommodations

Overall:

- **Make sure the alternative setting is conducive to test-taking, and that the setting is quiet, has minimal distractions, and is reserved for a sufficient block of time.**
- **The decision to administer the test in an alternative setting is often linked to the need to properly administer other accommodations. Therefore, the setting accommodations are often bundled with other accommodations. Make sure the setting matches the intent of the other accommodation. For example,**
  - **an individual student who needs to have sections of the test read aloud (P3) should be in a location that will not disturb other students.**
  - **it may not be possible to test a student who needs short, supervised breaks (T3) as an accommodation during testing in a small group.**
- **All tests must be administered by trained school personnel.**

S2 In an out-of-school setting by trained school personnel

- The use of a non-school setting is intended for unique situations, such as a student who is incarcerated or a student with a long-term illness receiving instruction at home. This accommodation is not intended for students who are home-schooled.

REMINDER

Test security must be maintained in all alternative settings. The administrator must secure all test materials during transfer to and from the alternative setting. If the student is being tested in a separate location within the school, do not send the student to the alternative setting or back to the classroom on his/her own with testing materials.

Presentation (P)

P1. Individually
P2. In a small group
P3. With test and directions read aloud in English or signed to the student
P4. With only test directions read aloud or signed to the student
P5. With administrator verification of student understanding following the reading of test directions
P6. Using alternative or assistive technology that is part of the student's communication system
P7. By trained school personnel known to the student other than the student's classroom teacher
P8. Using a large-print version of assessment
P9. Using Braille version of assessment
P10. Using a word-to-word translation dictionary for ELL students (NOT allowed for the Reading test)
P11. Using visual or auditory supports

Considerations for Presentation accommodations

Overall:

- **Many presentation accommodations need to be bundled with an alternative setting (S1) accommodation. Make sure to take this into consideration when planning needed accommodations for a student, and when determining the number of individual (P1) and small group (P2) accommodations that may be needed.**

P2 In a small group

- The size of a small group of students will vary depending on the other accommodations that are bundled with it.
- Students who need the test and directions read aloud (P3) should probably be limited to a group size of no more than five. Students will need to work the problems out at their own individual pace. The test administrator will need to read a test item when the student is ready to work on that item.
- Students who have been identified in advance as needing an extended time accommodation (T1) may be able to have the test administered in a group larger than five students because students are working at their own pace.

P3 With test and directions read aloud in English or signed to the student.

- This accommodation should be administered in a manner that provides the amount of support required by the student and in a manner most similar to routine classroom instruction and test-taking for the student. For example,
  - The test administrator may read the entire test and sessions to the student,

- The student may ask for only portions of the test to be read aloud by the test administrator, as needed.

P4 With only test directions read aloud or signed to the student.

- Scripted test directions contained in the Test Administrator Manuals are read aloud by test administrators to all students being tested. These scripted directions may be repeated for any student as a standard procedure. Reading these scripted test directions should not be marked as an accommodation.
- On MEA tests, test directions inside test booklets are very limited and are generally printed in bold in a page header, page footer, or above a test item number. These directions may be repeated as often as needed. Examples of these directions include
  - the session header or footer "Science – Session 1"
  - "Answer questions 1 through 13 on page 12 in your Student Answer Booklet"
  - navigation directions such as "Go on." or "Do not go on. Stop."
  - general notes such as "No test materials on this page."
- Test directions **DO NOT** include test items, response options, or similar materials. Those materials may not be read aloud under accommodation P4.
- With the exception of sign language and the case of students enrolled in a program where the test administrator routinely presents information in a foreign language, test directions may not be translated.

P5 With administrator verification of student understanding following the reading of test directions

- Test administrators are expected to ask all students in a class whether they understand scripted test directions that are read by test administrator. That practice should not be marked as an accommodation and is not the intent of accommodation P5.
- This accommodation should be used with students who routinely have trouble interpreting and following directions during normal classroom instruction and test-taking situations.

P8 Using a large-print version of assessment

- Prior to each test administration the Department of Education provides schools with information needed to order large-print tests for the upcoming administration.
- Make sure to consider whether the student uses large-print materials routinely for classroom instruction and test-taking. The MEA tests should not be the first time the student is introduced to large-print materials.
- If the use of large-print materials results in the student responding outside of the Student Answer Booklet (for example R4), policies regarding recorded answers must be followed prior to returning test materials.

P9 Using Braille version of assessment

- Prior to each test administration the Department of Education provides schools with information needed to order Braille tests for the upcoming administration.
- Make sure to consider whether the student is using Braille routinely for classroom instruction and test-taking. The MEA tests are translated into both contracted and

uncontracted Braille. Uncontracted Braille is usually used by younger or beginning Braille readers. The school personnel ordering Braille test(s) will need to know what type of Braille the student is accustomed to using.
- Use of Braille tests is likely to be bundled with a response accommodation such as R4. Policies regarding recorded answers must be followed prior to returning test materials.

P10 Using a word-to-word translation dictionary for ELL students

- This accommodation is most appropriate for intermediate-stage English language learners. Research has shown that this accommodation is not helpful for beginning-stage learners.
- Make sure to consider whether the student uses a word-to-word translation dictionary routinely during classroom instruction and test-taking.
- A word-to-word translation dictionary does not include any definitions. Additional information on appropriate word-to-word translation dictionaries can be found in Section 6: Accommodation Resources.
- It is likely that this accommodation will be bundled with the corresponding response accommodation R7.

P11 Using visual or auditory supports

- This accommodation includes a variety of visual and auditory supports designed to minimize distractions and help students focus during the test administration.
- Materials that block other print on the page of the Test Booklet or Student Answer Booklet must not permanently alter the booklet. The use of Post-it Notes or plain paper, for example, is acceptable, but must be removed prior to returning materials.

Response (R)

R1. With a student dictating responses to school personnel
R2. With a student dictating responses using alternative or assistive technology/devices that are part of the student's communication system
R3. With a student using approved tools or devices to minimize distractions
R4. With a student writing responses using separate paper, a word processer, computer, Brailler, or similar device
R5. With a student indicating responses to multiple-choice items to school personnel
R6. With a student responding with the use of visual aids
R7. With a student with limited English proficiency responding with use of a word-to-word dictionary

Considerations for Response accommodations
Overall:

- **Note that response accommodations R1, R2, and R4 may result in the student responding outside of the Student Answer Booklet. Policies regarding recorded answers must be followed prior to returning test materials. Separate sheets of paper stapled, taped, or glued into the Student Answer Booklet are not acceptable.**
- **Students responding outside of the Student Answer Booklet should be shown the provided answer space in the Student Answer Booklet prior to responding, in order for them to gauge how much to write.**

R1 With a student dictating responses to school personnel.

- When using this accommodation, the student is dictating his/her responses to the constructed response questions directly to trained school personnel or recording oral responses. In most cases, school personnel will be able to scribe student responses directly into the Student Answer Booklet at the same time as the student is responding.
- Responses must be scribed exactly as dictated.
- In the event that responses are not scribed in the Student Answer Booklet at the same time as they are produced, policies regarding recorded answers must be followed prior to returning test materials.
- This response accommodation should require an individual administration (P1) and may also require the use of an alternative setting (S1) and possible the use of extended time (T1).

R4 With a student writing responses using separate paper, a word processor, computer, brailler, or similar device

- When any R4 accommodation is used, there are specific instructions you must follow after testing has been completed to prepare the materials to be returned for scoring.

These procedures permit the student work to be scored and, *if the appropriate procedure is not followed, student work will not be scored or credited*. These instructions are provided in both the Test Coordinator Manual and in the Test Administrator Manual in the section titled: **After Testing: Special Materials.**

R5 With a student indicating responses to multiple-choice items to school personnel

- When using this accommodation, the student is indicating by pointing to or naming/reading aloud his/her answers to multiple-choice questions and at the same time school personnel is marking those answers in the Student Answer Booklet.
- This response accommodation should require an individual administration (P1) and may also require the use of an alternative setting (S1) and possible the use of extended time (T1).

R6 With a student responding with the use of visual aids

R7 With a student with limited English proficiency responding with use of a word-to-word dictionary

- This accommodation is most appropriate for intermediate-stage English language learners. Research has shown that this accommodation is not helpful for beginning-stage learners.
- Make sure to consider whether the student uses a word-to-word translation dictionary routinely during classroom instruction and test-taking.
- A word-to-word translation dictionary does not include any definitions. Additional information on appropriate word-to-word translation dictionaries can be found in Section 6: Accommodation Resources.
- It is likely that this accommodation will be bundled with the corresponding presentation accommodation P10.

Other Accommodations (O) and Modifications (M)

The school must contact the Department of Education to request approval for any accommodation not listed as a standard accommodation under the categories for Timing (T), Setting (S), Presentation (P), and Response (R) (see contact information on page ii). The approval process and any consequences that result from the use of the proposed accommodation will be discussed. Based on the information provided by school personnel, the Department of Education will determine if the proposed accommodation will preserve the comparability of the test items to which it is being applied. The Department of Education will not override IEP Team decisions regarding the use of an accommodation during testing. However, if the school decides to use an accommodation that was determined to be non-comparable, then the proposed accommodation will be treated as a modification. All impacted items using unapproved O accommodations or modifications result in no credit being given.

Other Accommodations (O)

O1. Using other accommodation(s) not on this list, requested by the accommodations team.

Considerations for Other Accommodations

Overall:

- Use the "Other Accommodation" Discussion Worksheet in Appendix B to gather the information needed for a discussion with state personnel on accommodations.
- Contact state personnel about the accommodation being proposed. State personnel will determine whether the accommodation is comparable or not.
- Proposed accommodations that are verified as comparable by state personnel will need to be recorded as O1 on the Student Answer Booklet.
- If the proposed accommodation is deemed not comparable, then it is classified as a modification (see "Modifications" on page 20) and coded as M3.
- Verification of the comparability of a proposed accommodation is made on a test-by-test and student-by-student basis. Do not assume that an accommodation verified as comparable for one student will be deemed comparable for another. Do not assume that an accommodation verified as comparable one year will be approved for future assessments.
- The O1 code may only be used after written approval from the DOE. The O1 code used without approval will be treated as a modification for purposes of scoring and reporting (see "Modifications" on page 20).

Modifications (M)

M3. Using an accommodation on this list not approved for a particular test or an accommodation not included on this list without prior approval of the DOE

Considerations for Modifications

Overall:

- **Note that the use of any modification invalidates any and all items impacted within the session(s) in which it is used, and no credit will be given for student performance on those items. It is important for the school to take this into consideration when choosing whether to allow a modification.**

M3. Using an accommodation on this list not approved for a particular test or an accommodation not included on this list without prior approval of the DOE

- Proposed accommodations that are determined to be non-comparable are considered modifications and should be coded as M3. All items impacted by the accommodation will be scored as incorrect.

Section 5: Accommodation Decision Teams

Procedures

- A school team meets to discuss which, if any, accommodations are appropriate for the student during a state testing situation. The following questions should be discussed when considering appropriate accommodations for students:
    - What accommodation(s) does this student typically need while taking tests in the classroom?
    - Are there unique circumstances created by the MEA assessment that might require accommodations for this student?
    - Which accommodations, if any, from the *Table of Standard Test Accommodations* will meet this student's participation requirements?
    - If none of the standard accommodations meet the student's participation requirements, what other accommodations might be proposed?
    - Are there different accommodations needed for the student to participate in each content area tested?
- Decisions concerning appropriate accommodations for MEA testing are documented and communicated in order to ensure that the accommodation is implemented effectively and available for future instructional planning.
- The *Accommodation Decision Procedures Worksheet* found in Appendix A may be used and copied to facilitate and communicate accommodations planning. Note that this worksheet is keyed to the *Table of Standard Test Accommodations*. The lettered sections of the *Accommodation Decision Procedures Worksheet* refer to the same letters in the *Table of Standard Test Accommodations* beginning on page 8 of this document.

REMINDER:

Not all accommodations that are used for routine instruction are appropriate for MEA testing. If you are unsure about the appropriateness use of an accommodation for MEA testing, contact the Department of Education to discuss how best to use the accommodation or whether the proposed accommodation is actually a modification.

Administration: Planning for Needed Resources

Each school will need to prepare for large scale test administration because multiple grades are being tested and multiple accommodations are being implemented. Some important questions to consider are:

- How many spaces will you need to secure for small group and individual administrations of the assessment?
- How many students need a place where they can read aloud or be read to without disturbing other students who may be working?
- How many people will then be needed and who should they be?
- How and when will you train the people who will be administering accommodations?

Administration: Qualified Personnel

Accommodations must be administered by school personnel who are employed by the district and have been trained to administer the assessment. It is preferable that the person administering the accommodation(s) be familiar with and to the student(s). This is especially true for accommodation situations that call for individual settings.

The following are individuals who may **not** administer tests:

- Parents and other community volunteers
- Peer tutors
- Other students

Administration: Preparing Test Administrators

Equally as important as identifying appropriate school personnel to administer the test with accommodations is the training and knowledge provided to the administrator prior to administration of the state test.

The following are recommendations for ways to help school personnel prepare to administer the MEA assessments with one or more accommodations:

- Attend a training session, implemented by the school, which explains and reviews at minimum the Test Administrator Manual and this Accommodations Guide.
- Read both the Test Administrator Manual and this training guide prior to test administration.
- Experience implementing classroom accommodations that are similar to MEA accommodations.
- Provide the test booklet on the day of testing, prior to the testing session, so that the test administrator can become familiar with the test form in advance of administration. This is especially true of accommodation administrators who will need to read particular test sessions aloud to a student(s).

REMINDER:

Building administrators are required to sign off on the *Principal's Certification of Proper Test Administration* form, verifying that all test administrators are school personnel and have been properly trained.

Section 6: Accommodation Resources

This document is available in electronic format at:
www.maine.gov/education/mea/admininfo/htm. http:///

External Resources

National Center on Educational Outcomes
Special Topic Area: Accommodations for Students with Disabilities:
http://www.cehd.umn.edu/NCEO/TopicAreas/Accommodations/accomtopic.htm

The National Center for Educational Outcomes, affiliated with the University of Minnesota, is a central repository of research studies and general information focusing on the use of accommodations that support students with disabilities.

George Washington University
ELL Accommodations Online Toolkit and Database: http://ells.ceee.gwu.edu

This is an excellent and recently updated guide, developed by George Washington University, that will help educators better understand the nature of accommodations that are specifically responsive to the needs of ELL students. Typically these accommodations include both direct and indirect linguistic supports, which to be effective, need to be combined with other specific accommodations. Accommodations that are uniquely responsive to the needs of this population often differ from those most effective in supporting other student groups.

Center for Applied Special Technology (CAST)
Universal Design for Learning: http://www.CAST.org

CAST, a research and development organization affiliated with Harvard University, has pioneered development of the area known as "Universal Design for Learning (UDL)". This site provides valuable information for educators who wish to learn more about factors that must be considered to provide meaningful access for ALL students to curriculum materials and assessment.

Council of Chief State School Officers (CCSS0)
Accommodations Manual: How to Select, Administer, and Evaluate Use of Accommodations for Instruction and Assessment of Students with Disabilities, Second Edition (August 2005).
www.ccsso.org/projects/SCASS/projects/assessing_special_education_students/11302.cfm

Additional External Organizations

Council for Exceptional Children (CEC)
www.cec.sped.org

The CEC is the largest international professional organization dedicated to improving educational outcomes for individuals with exceptionalities, students with disabilities, and/or the gifted. The CEC advocates for appropriate governmental policies, sets professional standards, provides continual professional development, advocates for newly and historically underserved individuals with exceptionalities, and helps professionals obtain conditions and resources necessary for effective professional practice.

LD Online
www.ldonline.org

LD Online has many articles dealing with state assessments, large-scale assessments, and assessing achievement in skill areas.

National Information Center for Children and Youth with Disabilities (NICHCY)
www.nichcy.org

The NICHCY serves the nation as a central source of information on: disabilities in infants, toddlers, children, and youth; IDEA, which is the law authorizing special education; No Child Left Behind (as it relates to children with disabilities); and research-based information on effective educational practices.

Special Education Resources on the Internet (SERI)
www.seriweb.com

SERI houses a collection of Internet-accessible information resources of interest to those involved in the fields related to special education. This collection exists in order to make online special education resources more easily and readily available in one location. This site will continually modify, update, and add additional informative links.

Resources for Students with Blindness or Visual Impairment

American Printing House for the Blind, **Accessible Tests Department**
http://www.aph.org/tests/index.html

National Agenda for the Education of Children and Youths with Visual Impairments, Including Those with Multiple Disabilities
http://www.tsbvi.edu/agenda/

Resources for the Deaf and Hard of Hearing

Laurent Clerc National Deaf Education Center, Gallaudet University
http://clerccenter.gallaudet.edu/

Rochester Institute of Technology Libraries, Subject-Based Deaf and Hard of Hearing Internet Resources
http://wally.rit.edu/internet/subject/deafness.html

Appendix A: Accommodation Decision Procedures Worksheet

(Refer to the *Table of Standard Test Accommodations* on beginning on page 8 of this document)

<table>
<tr><td colspan="2">Team Members Present:</td><td>Student Name:

Date:</td></tr>
<tr><td colspan="3">A. Does the student use alternative Settings accommodations during routine classroom testing?</td></tr>
<tr><td>Yes
No</td><td colspan="2">If yes, describe:</td></tr>
<tr><td colspan="3">Will the student need alternative Settings accommodations during MEA testing?</td></tr>
<tr><td>Yes
No</td><td colspan="2">If yes, which accommodations will be needed?</td></tr>
<tr><td colspan="3">B. Does the student use scheduling and Timing accommodations during routine classroom testing?</td></tr>
<tr><td>Yes
No</td><td colspan="2">If yes, describe:</td></tr>
<tr><td colspan="3">Will the student need scheduling and Timing accommodations during MEA testing?</td></tr>
<tr><td>Yes
No</td><td colspan="2">If yes, which accommodations will be needed?</td></tr>
<tr><td colspan="3">C. Does the student use Presentation Formats accommodations during routine classroom testing?</td></tr>
<tr><td>Yes
No</td><td colspan="2">If yes, describe:</td></tr>
</table>

Will the student need **Presentation Formats** accommodations during MEA testing?	
Yes No	If yes, which accommodations will be needed?
D. Does the student use **Response Formats** accommodations during routine classroom testing?	
Yes No	If yes, describe:
Will the student need **Response Formats** accommodations during MEA testing?	
Yes No	If yes, which accommodations will be needed?
E. Does the student use **Other Accommodations** during routine classroom testing?	
Yes No	If yes, describe:
Will the student need **Other Accommodations** during MEA testing?	
Yes No	If yes, which accommodations will be needed?
Contact the State Department of Education staff listed in the contact information on page ii to discuss and receive approval for any accommodation not listed on the *Table of Standard Test Accommodations*.	
F. Does the student use **Modifications** during routine classroom testing?	
Yes No	If yes, describe:
Will the student need **Modifications** during MEA testing?	
Yes No	If yes, which modifications will be needed?

Appendix B: "Other Accommodation" Discussion Worksheet

Please complete this form before contacting the Department of Education with your proposal.

Student Information

Student Name:	Grade Level/Content Area(s)/Session(s):
Student Identification Number:	

Contact Information

Contact Name:	Contact Title:
Contact Phone:	Contact E-Mail:
Contact School Name and Address:	

Assurances:

- ☐ The school team has met and has considered all standard accommodations prior to proposing other accommodations.
- ☐ Parent(s)/guardian(s) were provided an opportunity to participate in the decision-making process.
- ☐ The proposed accommodation is used for routine class instruction and/or test-taking.

Description of the proposed accommodation and why it is deemed necessary:

Result of discussion with DOE personnel: Name of DOE personnel: Date:

Appendix D:

Appendix D: Supporting Students with Limited English Proficiency (LEP/ELL)

MEA Policy: MEA test accommodations are available to all students, regardless of whether or not a disability has been identified. Accommodations allowed in MEA testing are not group specific. For example many students with limited English proficiency benefit from certain language-based accommodations, but like any other student, they sometimes break their arms or develop visual difficulties, and may need accommodations during testing that are very different than the linguistic accommodations often recommended for students in this group. Before they are members of any subgroup, each student is first an individual with unique learning needs. MEA assessment accommodations policy treats students this way. The decision to allow all students to use the full range of accommodations, as needed, is consistent with prior research on best practice in the provision of accommodations (c.f. Elbaum, Aguelles, Campbell & Saleh, 2004, pp. 71-87). The MEA management team believes strongly that a fair and valid path of access to a universally designed test should not require that a student carry any specific group label or disability. Rather, much like differentiated instruction, accommodated conditions of test participation that preserve the essential construct of the standard being assessed should be supported for any student who has been shown to need these differentiated test conditions. This philosophy is consistent with the MEA team's commitment to building a universally accessible test that provides an accurate measure of what each student knows.

MEA accommodations policy supports the use of many accommodations that are being found to be effective with LEP/ELL students (as identified by the ongoing and growing national research effort). The purpose of this section is to help teams identify those allowable linguistic accommodations that may prove particularly useful to helping LEP/ELL students show us what they know and are able to do.

It is important to review the LEP *MEA* Accommodations Support tables that follow in the context provided by the six statements below, while remembering that research continues to inform and update our understanding of these issues. For more information on this issue, see also: Resource Note, below.

1. The stage of English language acquisition demonstrated by the individual student must be taken into consideration when choosing the most appropriate accommodations. For example, students who are beginning English language learners (ELL) do not usually benefit from the use of commonly used 'word-to-word translation' tools. Development of English language vocabulary skills must be further advanced for this accommodation to prove useful. In fact, use of word-to-word translation tools too soon can worsen confusion for students very new to the English language.

2. The structure of the first language of the student should be considered when choosing accommodations that may prove useful. For example, some languages (i.e. Japanese Kanji, Mandarin Chinese, and American Sign) are structured ideographically (contain picture-like symbols). Other languages (i.e. English, Spanish, French, German) are structured very differently and emphasize other representation formats (i.e. phonetically

based symbols). Depending upon the cultural communication experience of the student, the types of linguistic or other accommodations considered might be different.

3. Accommodation decisions for LEP/ELL students are best informed when educators who have specific training and expertise in second language acquisition are part of the decision making team. Knowledge of how students acquire a new language after the first language is a specialized area that differs significantly from knowledge of first language acquisition.

4. Translation 'on the fly' is not recommended for most instructional purposes and is not permitted during *MEA* assessment. Misunderstanding and miscommunication is the norm when this technique is used and can create many additional problems. Teams using this technique must do so with as much training and information as possible. *MEA* permits only limited exceptions to this no translation rule: see items P3, P4, and P5 in the chart below for clarification.

5. In addition, please note that American Sign and Braille languages are not treated as 'foreign language' translations for purposes of MEA test administration. These languages are the only means by which some students are able to access certain portions of academic tests. For this reason, these languages differ from other formal languages for some assessment purposes. Where permitted, American Sign Language ("sign") and Braille are specifically noted in the *Table of Standard Test Accommodations*.

Use of 14 Common Linguistic Accommodations & Supports on the MEA-Science Test

The following tables describe the use on the MEA tests of 14 common linguistic accommodations and supports. The primary sources used to construct these tables were

1. MEA *Table of Standard Accommodations, Revised 2010*; and
2. The George Washington University Center for Equity and Excellence in Education, *Guide for Refining State Assessment Policies for Accommodating English Language Learners*, C. Rivera, B.D. Acosta & L.S. Willner, 2008.

For more information regarding the appropriate use of accommodations will students who show limited English proficiency, see: *The Guide for Reining State Assessment Policies for Accommodating English Language Learners*, 2008, available at http://ceee.gwu.edu.

1. Plain English text used in items and passages
Type of Support Direct, English language
Recommended for Intermediate and Advanced English language learners
Use on MEA: A Bias/Sensitivity Committee that includes ELL specialists and Item Review Committees composed of review the appropriateness of language used in every test item considered for inclusion on the tests.
Coding as an accommodation on the MEA Student Answer Booklet Not applicable.

2. English language reference materials and supports
Type of Support Direct, English language
Recommended for Intermediate and Advanced English language learners
Use on MEA: A number of English language reference materials are built into the MEA tests or provided for use by all students during MEA testing**.**
Coding as an accommodation on the MEA Student Answer Booklet These are considered Generally Allowable Supports and do not require documentation as an accommodation.

3. Customized glossary or dictionary with word meanings or definitions (English only or dual language)
Type of Support Direct, English language or Direct, Native language
Recommended for Intermediate and Advanced English language learners
Use on MEA: Dictionaries or glossaries containing word meanings or definitions of any kind are **NOT ALLOWED** during MEA testing. Use of a glossary or dictionary with word meanings or definitions would be considered a modification resulting in impacted items being scored as incorrect.
Coding as an accommodation on the MEA Student Answer Booklet Use of dictionaries or glossaries containing word meanings or definitions must be coded as modification **M3**.

4. Commercial word-to-word dual language translation dictionary without definitions
Type of Support Direct, Native language
Recommended for Intermediate and Advanced English language learners
Use on MEA: This is a standard accommodation allowed on the MEA Science test.
Coding as an accommodation on the MEA Student Answer Booklet Code as standard accommodation **P10** and **R7** if used on the Science test.

5. Individually customized word-to-word translation lists without definitions
Type of Support Direct, Native language
Recommended for May be helpful for some Beginning level English language learners
Use on MEA: Use of customized or "homemade" lists is not included on the list of standard accommodations and requires prior approval of the Department of Education on a case by case basis as an "Other" allowable accommodation for use on the Science test
Coding as an accommodation on the MEA Student Answer Booklet If approved by the Department of Education for use on the Science test, code as an "Other" accommodation **O1**.

6. Test *directions* are read aloud in English or signed to a student
Type of Support Direct, English language
Recommended for Intermediate English language learners
Use on MEA: Reading test directions aloud in English or signing directions to a student is a standard accommodation on the MEA tests. Refer to the appropriate documentation in the *Accommodations Guide* for additional details on what are considered directions on the MEA tests. Note that native language translation of test directions is NOT ALLOWED except in the case of a student enrolled in a program where the test administrator routinely presents information in a foreign language.
Coding as an accommodation on the MEA Student Answer Booklet Code as standard accommodation **P4**.

7. Provide written version of test directions in native language
Type of Support Direct, Native language
Recommended for Beginning English language learners
Use on MEA: Written translations of test directions are not allowed on the MEA tests. Contact the Department of Education for additional information if you have questions.
Coding as an accommodation on the MEA Student Answer Booklet Not applicable

8. Administrator verifies student understanding of scripted test directions (Administrator may then clarify or paraphrase directions, if needed.)
Type of Support Direct, English language
Recommended for May be helpful to some Beginning and Intermediate English language learners
Use on MEA: Verification of an individual student's understanding of test directions is a standard accommodation on the MEA tests.
Coding as an accommodation on the MEA Student Answer Booklet Code as standard accommodation **P5**.

9. Entire test and directions are read aloud in English or signed to a student (Repetition is permitted as needed.)
Type of Support Direct, English language
Recommended for May be helpful to some Intermediate English language learners
Use on MEA: Reading the entire test and directions as needed is a standard accommodation on the MEA Science test
Coding as an accommodation on the MEA Student Answer Booklet Code as standard accommodation **P3** if used on the Science test.

10. Provide pre-recorded English or signed version of the entire test and directions to student.
Type of Support Direct, English language
Recommended for May be helpful to some Intermediate ELL students
Use on MEA: Use of technology by individual schools or districts to pre-record the test items or directions raises security concerns and is not considered a standard accommodation on the MEA test. Contact the Department of Education to discuss the specific technology being proposed and request approval as an "Other" accommodation.
Coding as an accommodation on the MEA Student Answer Booklet If approved for use by the Department of Education code as "Other" accommodation **O1**.

11. Allow student to respond in writing in native language
Type of Support Direct, Native language
Recommended for May be helpful to some Beginning and Intermediate English language learners
Use on MEA: Responses in languages other than English are not allowed on the MEA tests. Note that students are allowed to respond using Braille, but that applicable policies regarding recorded answers must be followed prior to returning test materials.
Coding as an accommodation on the MEA Student Answer Booklet In the case of Braille, code as standard accommodation **R4**. Not applicable for other languages.

12. Allow student to dictate responses to constructed response items in English
Type of Support Direct, English language
Recommended for May be helpful to some Intermediate English language learners
Use on MEA: Dictating responses to constructed response items is a standard accommodation on the MEA Science test.
Coding as an accommodation on the MEA Student Answer Booklet Code as standard accommodation R1 if used on the Science test.

13. Allow student to dictate responses to multiple choice items orally in English
Type of Support Direct, English language
Recommended for May be helpful to some Intermediate English language learners
Use on MEA: Dictating responses to multiple-choice items is a standard accommodation on all MEA tests.
Coding as an accommodation on the MEA Student Answer Booklet Code as standard accommodation **R5**

14. Allow extended time to complete a test session beyond the scheduled administration time within the same day
Type of Support Indirect
Recommended for Beginning, Intermediate, and Advanced English language learners
Use on MEA: Use of extended time is a standard accommodation on all MEA tests. Note that the scheduled administration time already includes a sufficient amount of time beyond the time expected for the majority of students to complete the test. Refer to applicable documentation in the Accommodations Guide and Manuals for additional information on scheduled administration time and the use of the extended time accommodation.
Coding as an accommodation on the MEA Student Answer Booklet Code as standard accommodation **T1**.

APPENDIX F—SAMPLE REPORTS

Suggestions for Helping Your Child in Science

• **Talk with your child** about what he or she finds really interesting about science. Use this interest as a starting point to learn more together about this subject, then encourage your child to question and explain the bigger ideas and concepts related to this area.

• **Does your child have an interest in space**, how the earth changes, magnets, gravity, pendulums, etc.? If so, check out Section D, “Earth’s Physical Setting,” of the state standards for more information about what knowledge is expected at each grade span. Make observations about the moon, experiment with magnets, discover why metal gets rusty, test rolling a ball on different surfaces and see what happens. Ask questions about what you see.

• **Is your child interested in living things**, how they change and survive, the differences and similarities among them, what they are made of and how they interact with each other and the environment? Refer to Section E, “The Living Environment," of the state standards for more detailed explanations of the knowledge your child should have.

• **Be observers of the world around you**: walk outdoors, visit the library and museums, and look for information on the internet and television. Question what you see and talk together about detailed facts and the bigger ideas connected to those facts.

• **Use your child’s textbook**, library books or other resources to identify and complete some simple and safe experiments together.

• **Talk with your child about areas in science** that are confusing and ask for help from your child’s teacher.

• **Always keep in close contact** with your child’s teacher and school about your child’s progress.

Maine Educational Assessment
Student Report

MEA
Maine Educational Assessment

May 2014 Science Results for
Colton A. Brooks

Grade 08
Demonstration School 1

Dear Parents and Guardians,
As you review this report, you will learn how your child scored on the May MEA science test and what the results mean. The report also offers ideas on how to encourage your child’s interest in science and help increase your child’s knowledge in specific areas.

These MEA science results should be used together with your child’s grades and daily schoolwork to gain a complete picture of how well your child is learning science concepts. If you have any questions about your child’s progress, I encourage you to meet with your child’s science teacher to discuss these results and ideas of how to support your child’s success.

Sincerely,

James E. Rier, Jr.
Commissioner of Education

General MEA Information

What is the MEA?

The MEA measures student progress in achieving the State science standards, known as the Maine Learning Results. These standards are the goals for what all students should know and be able to do at certain times in their school careers. You may review these standards at: http://www.maine.gov/education/lres/scitech/natlstandards.html.

Do all Students Take this Test?

All public school students in grades 5 and 8 must take this test to meet state assessment requirements and the federal requirements of the *No Child Left Behind Act*. Approximately 14,000 students took the test in each grade level.

What Does this Test Look Like?

This report provides one example of a test question, but you can see more questions online at: http://www.maine.gov/education/mea/mearelitems.htm. The test includes multiple-choice questions, where students can choose one out of 4 possible answers and earn 1 point for each correct answer. There are also questions that require students to write out their answer. Students can earn from 0 to 4 points on these constructed-response questions, depending on the completeness and accuracy of their answers.

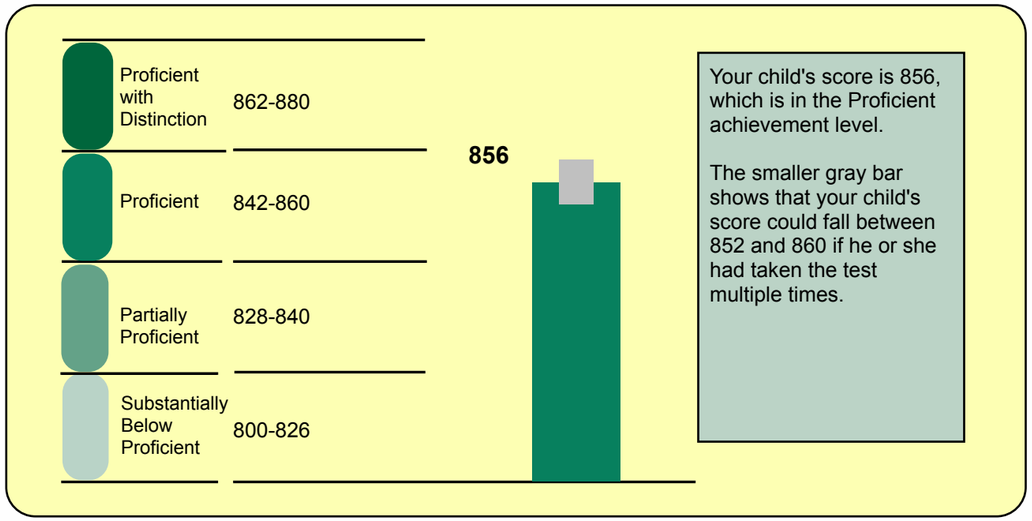
Who Decides what Questions are on this Test?

Test questions are developed by the test contractor and the Department of Education science specialist and are designed to measure a particular State standard. The questions are also reviewed by two groups of Maine teachers. One group makes sure that the questions measure the intended standard and that the language is appropriate for the grade level. The other group reviews all questions to make sure that they are fair and unbiased. All questions are field tested before a student is given a score on the question.

How are these Scores Used?

It is important to remember that this test score is just one measure of your child’s science knowledge. The results from this standardized test should only be used along with other school science grades to offer a more complete picture of your student’s learning. However, since it is the only test given to all students in this grade across the state of Maine, it gives you the opportunity to compare your student and school scores with other scores across the state. Schools and districts may use these scores to review science curriculum and instruction. The State uses the scores to determine how well schools are helping their students to meet State standards.

Your Child's Scaled Score and Achievement Level



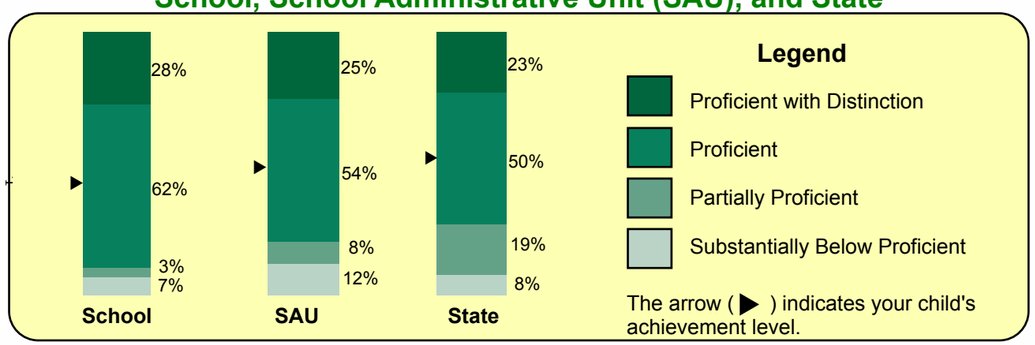
Student scores on the MEA fall into one of the four achievement levels listed above. These levels describe the quality of a student's answers on this test. Your child's score falls within the achievement level described below. The letters and numbers at the end of each bullet refer to particular science standards. If you would like more information about the other three levels, you may go to this website: http://www.maine.gov/doe/mea/resources/achievementleveldefinitions.pdf, or contact the Department of Education.

Proficient –

The student's work demonstrates a general understanding of essential concepts in science, including the ability to make connections among central ideas. The student's responses demonstrate the ability to analyze and solve routine problems, and explain central concepts with sufficient clarity and accuracy to demonstrate general understanding. Evidence of student's work at this achievement level may be provided by, but not limited to, examples from the Partially Proficient and Substantially Below Proficient levels, in addition to these examples that illustrate the following science ideas:

- describing the location of our solar system in its galaxy (D1c);
- describing ways in which two types of organisms (populations) may interact (E2b);
- describing Earth's atmosphere (D2b);
- explaining the relationship of the motion of molecules to the states of matter for gases, liquids, and solids (D3d); and/or
- comparing the structures that allow single-celled organisms and multicelled organisms to acquire and use energy (E3c).

Your Child's Performance Compared to Students in Your School, School Administrative Unit (SAU), and State



Your Child's Answers to Some Test Questions

Question Number	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
Science Standard	D2	E3	E1	D2	E4	E2	D1	D3	D3	E4	E2	D3	E5	D2	D4	D1	E5	D4	E2	E3	D2	E4
Your Child's Answer	✓	✓	D	✓	✓	D	D	✓	C	✓	✓	✓	✓	D	✓	✓	C	✓	✓	✓	1	3

Science Standards Included in the Test

The Physical Setting:
D1 = Universe and Solar System
D2 = Earth
D3 = Matter and Energy
D4 = Force and Motion

The Living Environment:
E1 = Biodiversity
E2 = Ecosystems
E3 = Cells
E4 = Heredity and Reproduction
E5 = Evolution

Key

✓ = Correct Answer
A Letter = Incorrect Answer Choice
A Number = Number of Points Earned Out of 4
* = Multiple Answer Choices
Blank = No Answer

Every year, one half of the questions from the test are released (made public). The chart above reports the results of how your child answered these released questions. One example of these questions is included here, along with the percentage of students in the state who chose each answer option.

If you would like to see other released questions to better understand your child's performance, they are located at: http://www.maine.gov/doe/mea/resources/released/index.html. *Please remember that you can see only half of the questions that were in this test. The other half of the questions may be used again in the future.*

MEA Released Question

Question Number 18

Lori owns a house next to the lake. She uses lots of fertilizer to keep her lawn green. Which impact could fertilizing her lawn have on the lake?

77% A. an increase in the algae population

4% B. an increase in the fish population

14% C. an increase in the mosquito population

5% D. an increase in the lake's depth

Your child chose A.
77% of Maine students chose the correct answer A.

A Closer Look at Your Child's Performance

Maine's goal is for all students to reach the proficient level. The chart below provides information about how your child performed on each science standard that was tested compared to how proficient students performed.

	Total Possible Points	Points Your Child Earned	Below Proficient	Similar To Proficient Students	Above Proficient Students
The Physical Setting	34	22			✓
Earth/Space	17	11		✓	
Matter and Energy/Force and Motion	17	11			✓
The Living Enviroment	22	15			✓

Name: Fowler, Shane J.
State ID: D05100023
School: Demonstration School 1
SAU: Demonstration District A

		---------- Achievement Levels	---- Scaled Scores
Date: 05/2014	**Science:**	Proficient	546

Name: Malik, Doren M.
State ID: D05100043
School: Demonstration School 1
SAU: Demonstration District A

		---------- Achievement Levels	---- Scaled Scores
Date: 05/2014	**Science:**	Partially Proficient	540

Name: French, April T.
State ID: D05100047
School: Demonstration School 1
SAU: Demonstration District A

		---------- Achievement Levels	---- Scaled Scores
Date: 05/2014	**Science:**	Proficient	548

Name: Morrison, Amanda
State ID: D05100034
School: Demonstration School 1
SAU: Demonstration District A

		---------- Achievement Levels	---- Scaled Scores
Date: 05/2014	**Science:**	Proficient	550

Name: Kacher, Georgia M.
State ID: D05100055
School: Demonstration School 1
SAU: Demonstration District A

		---------- Achievement Levels	---- Scaled Scores
Date: 05/2014	**Science:**	Partially Proficient	536

Name: Mortenson, Eliza M.
State ID: D05100063
School: Demonstration School 1
SAU: Demonstration District A

		---------- Achievement Levels	---- Scaled Scores
Date: 05/2014	**Science:**	Partially Proficient	530

Name: Keeley, Dawna U.
State ID: D05100068
School: Demonstration School 1
SAU: Demonstration District A

		---------- Achievement Levels	---- Scaled Scores
Date: 05/2014	**Science:**	Partially Proficient	538

Name: Murrell, Abigail
State ID: D05100016
School: Demonstration School 1
SAU: Demonstration District A

		---------- Achievement Levels	---- Scaled Scores
Date: 05/2014	**Science:**	Partially Proficient	534

Name: Madden, Alicia M.
State ID: D05100062
School: Demonstration School 1
SAU: Demonstration District A

		---------- Achievement Levels	---- Scaled Scores
Date: 05/2014	**Science:**	Substantially Below Proficient	526

Name: Obrien, Lindsey Q.
State ID: D05100036
School: Demonstration School 1
SAU: Demonstration District A

		---------- Achievement Levels	---- Scaled Scores
Date: 05/2014	**Science:**	Proficient	552

August 2014

Maine Department of Education

2013-2014 School Year Reports

Dear School Board Members and School Personnel:

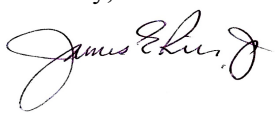
This 2013-2014 MEA Summary Report contains the results of student achievement in science at grades 5 and 8, as well as disaggregations by student and school characteristics. This report, together with MEA individual student and item analysis reports, provides support for use in program evaluation and planning.

The Maine Educational Assessment (MEA) is the State's measure of student progress in achieving the State science accountability standards of Maine's *Learning Results*. The *Learning Results* contain goals for what all students should know and be able to do at certain times in their school careers and include the accountability standards that are assessed for each grade. The MEA science test is administered to students in grades 5 and 8 to meet state assessment requirements and the federal requirements of the *No Child Left Behind* Act.

MEA results reflect scores based on the common science test questions that are taken by the approximately 14,000 students in each grade level. Students' scores are based on answers to a combination of multiple-choice questions and questions that require students to construct an answer. More information about the MEA is available at www.maine.gov/education/mea/index.htm.

Thank you for your continuing commitment to improve the quality and effectiveness of the instructional opportunities in your schools. These assuredly will help all students achieve the high standards of the *Learning Results* as demonstrated on classroom, district, and state assessments.

Sincerely,



James E. Rier, Jr.
Commissioner of Education

MEA
Maine Educational Assessment

Grade 8 Science State Report

Test Date: May 2014

Contents of the Report

The report is divided into four main sections including a section describing the students tested and a separate section for the content area results.

May 2014 - Grade 8 MEA Science Test

Grade Level Summary Report

State: Maine

Schools and SAUs administered the MEA Science tests to every enrolled student with the following exceptions: students who participated in the alternate assessment for the 2013-14 school year, students for whom a special consideration was granted through the state Department of Education, and other students for reasons not approved. On this page, and throughout this report, results are only reported for groups of students that are larger than nine (9).

PARTICIPATION SUMMARY	Number			Percentage		
	School	SAU	State	School	SAU	State
Students enrolled			**13,920**			**100**
	Science			Science		
Students tested			**13,473**			**97**
With an approved accommodation			1,780			13
Current LEP Students			399			3
With an approved accommodation			126			32
IEP Students			2,170			16
With an approved accommodation			1,443			66
Students not tested in MEA			**447**			**3**
State Approved			290			65
Alternate Assessment			240			83
Special Consideration			50			17
Other			157			35

PERFORMANCE SUMMARY

State																							
Enrolled	Not Tested Approved	Not Tested Other	Tested	Level 4		Level 3		Level 2		Level 1		Mean Scaled Score	Tested	Level 4	Level 3	Level 2	Level 1	Mean Scaled Score	Tested	Level 4	Level 3	Level 2	Level 1	Mean Scaled Score
N	N	N	N	N	%	N	%	N	%	N	%		N	%	%	%	%		N	%	%	%	%	
13,920	290	157	13,473	3,078	23	6,774	50	2,600	19	1,021	8	850												

Level 4 = Proficient with Distinction; Level 3 = Proficient; Level 2 = Partially Proficient; Level 1 = Substantially Below Proficient
Note: Throughout this report, percentages may not total 100 since each percentage is rounded to the nearest whole number.

May 2014 - Grade 8 MEA Science Test

Science Results

State: Maine

Proficient with Distinction (Level 4)
The student's work demonstrates in-depth understanding of essential concepts in science, including the ability to make multiple connections among central ideas. The student's responses demonstrate the ability to synthesize information, analyze and solve difficult problems, and explain complex concepts using evidence and proper terminology to support and communicate logical conclusions. (Scaled Score 862-880)

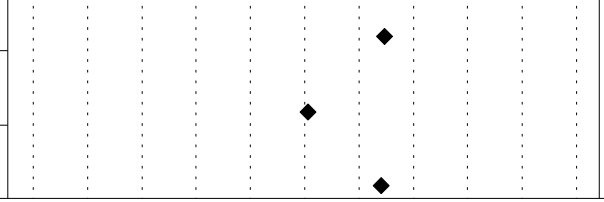
Proficient (Level 3)
The student's work demonstrates a general understanding of essential concepts in science, including the ability to make connections among central ideas. The student's responses demonstrate the ability to analyze and solve routine problems and explain central concepts with sufficient clarity and accuracy to demonstrate general understanding. (Scaled Score 842-860)

Partially Proficient (Level 2)
The student's work demonstrates incomplete understanding of essential concepts in science and inconsistent connections among central ideas. The student's responses demonstrate some ability to analyze and solve problems but the quality of responses is inconsistent. Explanation of concepts may be incomplete or unclear.
(Scaled Score 828-840)

Substantially Below Proficient (Level 1)
The student's work demonstrates limited understanding of essential concepts in science and infrequent or inaccurate connections among central ideas. The student's responses demonstrate minimal ability to solve problems. Explanations are illogical, incomplete, or missing. There are many inaccuracies. (Scaled Score 800-826)

	Enrolled	Not Tested Approved	Not Tested Other	Tested	Level 4		Level 3		Level 2		Level 1		Mean Scaled Score
	N	N	N	N	N	%	N	%	N	%	N	%	
School													
2011-12													
2012-13													
2013-14													
Cumulative Total													
SAU													
2011-12													
2012-13													
2013-14													
Cumulative Total													
State													
2011-12	14,388	238	195	13,955	3,120	22	6,942	50	2,956	21	937	7	850
2012-13	14,026	227	156	13,643	2,965	22	6,582	48	2,822	21	1,274	9	849
2013-14	**13,920**	**290**	**157**	**13,473**	**3,078**	**23**	**6,774**	**50**	**2,600**	**19**	**1,021**	**8**	**850**
Cumulative Total	42,334	755	508	41,071	9,163	22	20,298	49	8,378	20	3,232	8	850

Subtopic	Total Possible Points	Percent of Total Possible Points (0 10 20 30 40 50 60 70 80 90 100)
D. The Physical Setting	34	
D1/D2 Space/Earth	17	
D3/D4 Matter and Energy/Force and Motion	17	
E. The Living Environment	22	



The MEA assesses students' science knowledge based on questions that measure the science accountability content strands highlighted in Maine's 2007 Learning Results: Parameters for Essential Instruction, which can be found at: http://www.maine.gov/education/lres/pei/index.html.

§The standard error bar indicates how much the percent of points earned could vary if the students were examined multiple times with the same test.

May 2014 - Grade 8 MEA Science Test

Disaggregated Science Results

State: Maine

REPORTING CATEGORIES	State																											
	Enrolled	Not Tested Approved	Not Tested Other	Tested	Level 4		Level 3		Level 2		Level 1		Mean Scaled Score	Tested	Level 4	Level 3	Level 2	Level 1	Mean Scaled Score	Tested	Level 4	Level 3	Level 2	Level 1	Mean Scaled Score				
	N	N	N	N	N	%	N	%	N	%	N	%		N	%	%	%	%		N	%	%	%	%						
All Students	**13,920**	**290**	**157**	**13,473**	**3,078**	**23**	**6,774**	**50**	**2,600**	**19**	**1,021**	**8**	**850**																	
Gender																														
Male	7,171	184	94	6,893	1,713	25	3,318	48	1,281	19	581	8	850																	
Female	6,749	106	63	6,580	1,365	21	3,456	53	1,319	20	440	7	850																	
Not Reported	0	0	0	0																										
Race/Ethnicity																														
Hispanic or Latino	224	7	3	214	29	14	101	47	62	29	22	10	846																	
Not Hispanic or Latino																														
American Indian or Alaskan Native	109	2	9	98	16	16	52	53	22	22	8	8	849																	
Asian	207	1	2	204	74	36	84	41	35	17	11	5	853																	
Black or African American	419	9	13	397	24	6	142	36	134	34	97	24	838																	
Native Hawaiian or Pacific Islander	17	1	0	16	1	6	15	94	0	0	0	0	854																	
White	12,744	269	128	12,347	2,902	24	6,274	51	2,307	19	864	7	850																	
Two or more races	200	1	2	197	32	16	106	54	40	20	19	10	848																	
No Race/Ethnicity Reported	0	0	0	0																										
LEP Status																														
Current LEP student	426	8	19	399	10	3	140	35	146	37	103	26	836																	
Former LEP student - monitoring year 1	20	0	0	20	7	35	13	65	0	0	0	0	858																	
Former LEP student - monitoring year 2	9	0	0	9																										
All Other Students	13,465	282	138	13,045	3,056	23	6,617	51	2,454	19	918	7	850																	
IEP																														
Students with an IEP	2,499	262	67	2,170	110	5	628	29	826	38	606	28	836																	
All Other Students	11,421	28	90	11,303	2,968	26	6,146	54	1,774	16	415	4	853																	
SES																														
Economically Disadvantaged Students	6,181	175	106	5,900	748	13	2,831	48	1,598	27	723	12	845																	
All Other Students	7,739	115	51	7,573	2,330	31	3,943	52	1,002	13	298	4	854																	
Migrant																														
Migrant Students	2	0	0	2																										
All Other Students	13,918	290	157	13,471	3,078	23	6,772	50	2,600	19	1,021	8	850																	
Title I																														
Students Receiving Title I Services	965	14	14	937	80	9	444	47	305	33	108	12	844																	
All Other Students	12,955	276	143	12,536	2,998	24	6,330	50	2,295	18	913	7	851																	
504 Plan																														
Students with a 504 Plan	672	8	9	655	135	21	357	55	132	20	31	5	851																	
All Other Students	13,248	282	148	12,818	2,943	23	6,417	50	2,468	19	990	8	850																	

Level 4 = Proficient with Distinction; Level 3 = Proficient; Level 2 = Partially Proficient; Level 1 = Substantially Below Proficient
Note: Some numbers may have been left blank because fewer than ten (10) students were tested.

May 2014 - Grade 8 MEA Science Test

Questionnaire Results

State: Maine

QUESTIONNAIRE ITEMS	State																						
	Students in Each Category	Level 4		Level 3		Level 2		Level 1		Mean Scaled Score	Students in Each Category	Level 4	Level 3	Level 2	Level 1	Mean Scaled Score	Students in Each Category	Level 4	Level 3	Level 2	Level 1	Mean Scaled Score
	%	N	%	N	%	N	%	N	%		%	%	%	%	%		%	%	%	%	%	
Which statement best describes how often and how long your science class meets?																						
A. We meet every day for 45 minutes to an hour.	74	2,415	25	4,896	51	1,740	18	578	6	851												
B. We meet on alternate days for 80 to 90 minutes.	16	422	20	1,074	52	416	20	151	7	849												
C. We meet every day for 45 minutes, plus a longer lab period each week.	3	67	17	166	43	91	24	62	16	845												
D. We have a flexible schedule depending on the activities.	7	130	13	453	47	256	26	128	13	845												
Which statement best describes how you learn science?																						
A. I read a textbook and answer questions and/or take notes and do assignments.	18	400	17	1,070	46	575	25	274	12	846												
B. I work in groups to design and conduct experiments.	16	260	12	1,000	48	549	26	288	14	844												
C. I do a combination of A and B, mostly A.	39	1,423	28	2,588	51	819	16	203	4	853												
D. I do a combination of A and B, mostly B.	28	952	26	1,938	54	551	15	153	4	853												
How often do you make observations and collect data in science class?																						
A. a few times a week	47	1,217	20	3,178	52	1,250	20	463	8	849												
B. a few times a month	35	1,231	27	2,296	50	793	17	237	5	852												
C. once a month	10	377	29	584	45	242	19	95	7	852												
D. never or almost never	8	212	20	541	50	207	19	125	12	848												
How do you feel about the following statement?																						
My knowledge of science will be useful to me as an adult.																						
A. strongly agree	20	940	36	1,182	45	363	14	162	6	854												
B. agree	58	1,734	23	3,878	52	1,430	19	476	6	851												
C. disagree	17	303	14	1,216	55	521	23	191	9	847												
D. strongly disagree	5	60	9	320	49	185	28	94	14	843												
Which of the following best describes how you rate yourself as a student in science?																						
A. very good	18	992	42	1,016	43	248	11	103	4	857												
B. good	52	1,665	24	3,711	54	1,110	16	380	6	852												
C. fair	25	363	11	1,675	50	966	29	317	10	845												
D. poor	4	20	4	207	39	181	34	128	24	838												
How well do the questions that you have just been given on this MEA test match what you have learned in school about science?																						
A. The questions on the test match what I have learned in science class.	17	661	30	1,029	47	336	15	142	7	853												
B. They match some of what I have learned.	57	1,889	25	3,789	51	1,358	18	425	6	851												
C. They match just a little of what I have learned.	23	458	15	1,578	53	679	23	260	9	847												
D. There is no match.	4	31	6	214	45	133	28	100	21	840												
Which courses do you plan to take before you graduate from high school?																						
A. earth and space science and/or biology	26	539	16	1,817	53	787	23	296	9	848												
B. the course(s) described in A, plus chemistry	23	843	28	1,539	50	491	16	175	6	852												
C. the course(s) described in B, plus physics	22	1,168	41	1,216	43	302	11	148	5	856												
D. a life science and physical science class	28	479	13	1,979	54	909	25	292	8	847												

Level 4 = Proficient with Distinction; Level 3 = Proficient; Level 2 = Partially Proficient; Level 1 = Substantially Below Proficient

Note: Some numbers may have been left blank because fewer than ten (10) students were tested.

August 2014

Maine Department of Education

2013-2014 School Year Reports

Dear School Board Members and School Personnel:

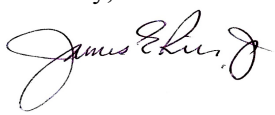
This 2013-2014 MEA Summary Report contains the results of student achievement in science at grades 5 and 8, as well as disaggregations by student and school characteristics. This report, together with MEA individual student and item analysis reports, provides support for use in program evaluation and planning.

The Maine Educational Assessment (MEA) is the State's measure of student progress in achieving the State science accountability standards of Maine's *Learning Results*. The *Learning Results* contain goals for what all students should know and be able to do at certain times in their school careers and include the accountability standards that are assessed for each grade. The MEA science test is administered to students in grades 5 and 8 to meet state assessment requirements and the federal requirements of the *No Child Left Behind* Act.

MEA results reflect scores based on the common science test questions that are taken by the approximately 14,000 students in each grade level. Students' scores are based on answers to a combination of multiple-choice questions and questions that require students to construct an answer. More information about the MEA is available at www.maine.gov/education/mea/index.htm.

Thank you for your continuing commitment to improve the quality and effectiveness of the instructional opportunities in your schools. These assuredly will help all students achieve the high standards of the *Learning Results* as demonstrated on classroom, district, and state assessments.

Sincerely,



James E. Rier, Jr.
Commissioner of Education

MEA
Maine Educational Assessment

Grade 8 Science SAU Report

Test Date: May 2014
Code: DEMA
SAU: Demonstration District A

Contents of the Report

The report is divided into four main sections including a section describing the students tested and a separate section for the content area results.

May 2014 - Grade 8 MEA Science Test

Grade Level Summary Report

SAU: Demonstration District A
State: Maine
Code: DEMA

Schools and SAUs administered the MEA Science tests to every enrolled student with the following exceptions: students who participated in the alternate assessment for the 2013-14 school year, students for whom a special consideration was granted through the state Department of Education, and other students for reasons not approved. On this page, and throughout this report, results are only reported for groups of students that are larger than nine (9).

PARTICIPATION SUMMARY	Number			Percentage		
	School	SAU	State	School	SAU	State
Students enrolled		**63**	**13,920**		**100**	**100**
	Science			Science		
Students tested		**59**	**13,473**		**94**	**97**
With an approved accommodation		6	1,780		10	13
Current LEP Students		2	399		3	3
With an approved accommodation		0	126		0	32
IEP Students		9	2,170		15	16
With an approved accommodation		5	1,443		56	66
Students not tested in MEA		**4**	**447**		**6**	**3**
State Approved		3	290		75	65
Alternate Assessment		2	240		67	83
Special Consideration		1	50		33	17
Other		1	157		25	35

PERFORMANCE SUMMARY

SAU													State											
Enrolled	Not Tested Approved	Not Tested Other	Tested	Level 4		Level 3		Level 2		Level 1		Mean Scaled Score	Tested	Level 4	Level 3	Level 2	Level 1	Mean Scaled Score	Tested	Level 4	Level 3	Level 2	Level 1	Mean Scaled Score
N	N	N	N	N	%	N	%	N	%	N	%		N	%	%	%	%		N	%	%	%	%	
63	3	1	59	15	25	32	54	5	8	7	12	851	13,473	23	50	19	8	850						

Level 4 = Proficient with Distinction; Level 3 = Proficient; Level 2 = Partially Proficient; Level 1 = Substantially Below Proficient
Note: Throughout this report, percentages may not total 100 since each percentage is rounded to the nearest whole number.

May 2014 - Grade 8 MEA Science Test

Science Results

SAU: Demonstration District A
State: Maine
Code: DEMA

Proficient with Distinction (Level 4)
The student's work demonstrates in-depth understanding of essential concepts in science, including the ability to make multiple connections among central ideas. The student's responses demonstrate the ability to synthesize information, analyze and solve difficult problems, and explain complex concepts using evidence and proper terminology to support and communicate logical conclusions. (Scaled Score 862-880)

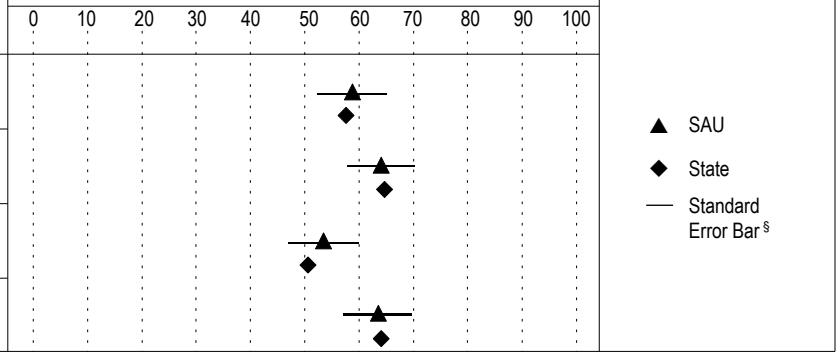
Proficient (Level 3)
The student's work demonstrates a general understanding of essential concepts in science, including the ability to make connections among central ideas. The student's responses demonstrate the ability to analyze and solve routine problems and explain central concepts with sufficient clarity and accuracy to demonstrate general understanding. (Scaled Score 842-860)

Partially Proficient (Level 2)
The student's work demonstrates incomplete understanding of essential concepts in science and inconsistent connections among central ideas. The student's responses demonstrate some ability to analyze and solve problems but the quality of responses is inconsistent. Explanation of concepts may be incomplete or unclear. (Scaled Score 828-840)

Substantially Below Proficient (Level 1)
The student's work demonstrates limited understanding of essential concepts in science and infrequent or inaccurate connections among central ideas. The student's responses demonstrate minimal ability to solve problems. Explanations are illogical, incomplete, or missing. There are many inaccuracies. (Scaled Score 800-826)

	Enrolled	Not Tested Approved	Not Tested Other	Tested	Level 4		Level 3		Level 2		Level 1		Mean Scaled Score
	N	N	N	N	N	%	N	%	N	%	N	%	
School													
2011-12													
2012-13													
2013-14													
Cumulative Total													
SAU													
2011-12	67	2	1	64	12	19	34	53	14	22	4	6	849
2012-13	66	2	2	62	15	24	29	47	12	19	6	10	850
2013-14	**63**	**3**	**1**	**59**	**15**	**25**	**32**	**54**	**5**	**8**	**7**	**12**	**851**
Cumulative Total	196	7	4	185	42	23	95	51	31	17	17	9	850
State													
2011-12	14,388	238	195	13,955	3,120	22	6,942	50	2,956	21	937	7	850
2012-13	14,026	227	156	13,643	2,965	22	6,582	48	2,822	21	1,274	9	849
2013-14	**13,920**	**290**	**157**	**13,473**	**3,078**	**23**	**6,774**	**50**	**2,600**	**19**	**1,021**	**8**	**850**
Cumulative Total	42,334	755	508	41,071	9,163	22	20,298	49	8,378	20	3,232	8	850

Subtopic	Total Possible Points
D. The Physical Setting	34
D1/D2 Space/Earth	17
D3/D4 Matter and Energy/Force and Motion	17
E. The Living Environment	22



The MEA assesses students' science knowledge based on questions that measure the science accountability content strands highlighted in Maine's 2007 Learning Results: Parameters for Essential Instruction, which can be found at: http://www.maine.gov/education/lres/pei/index.html.

§The standard error bar indicates how much the percent of points earned could vary if the students were examined multiple times with the same test.

May 2014 - Grade 8 MEA Science Test

Disaggregated Science Results

SAU: Demonstration District A
State: Maine
Code: DEMA

REPORTING CATEGORIES	SAU													State											
	Enrolled	Not Tested Approved	Not Tested Other	Tested	Level 4		Level 3		Level 2		Level 1		Mean Scaled Score	Tested	Level 4	Level 3	Level 2	Level 1	Mean Scaled Score	Tested	Level 4	Level 3	Level 2	Level 1	Mean Scaled Score
	N	N	N	N	N	%	N	%	N	%	N	%		N	%	%	%	%		N	%	%	%	%	
All Students	**63**	**3**	**1**	**59**	**15**	**25**	**32**	**54**	**5**	**8**	**7**	**12**	**851**	**13,473**	**23**	**50**	**19**	**8**	**850**						
Gender																									
Male	32	2	1	29	9	31	12	41	3	10	5	17	851	6,893	25	48	19	8	850						
Female	31	1	0	30	6	20	20	67	2	7	2	7	851	6,580	21	53	20	7	850						
Not Reported	0	0	0	0										0											
Race/Ethnicity																									
Hispanic or Latino	1	0	0	1										214	14	47	29	10	846						
Not Hispanic or Latino																									
American Indian or Alaskan Native	1	0	0	1										98	16	53	22	8	849						
Asian	1	0	0	1										204	36	41	17	5	853						
Black or African American	3	0	0	3										397	6	36	34	24	838						
Native Hawaiian or Pacific Islander	1	0	0	1										16	6	94	0	0	854						
White	55	3	1	51	14	27	27	53	4	8	6	12	851	12,347	24	51	19	7	850						
Two or more races	1	0	0	1										197	16	54	20	10	848						
No Race/Ethnicity Reported	0	0	0	0										0											
LEP Status																									
Current LEP student	2	0	0	2										399	3	35	37	26	836						
Former LEP student - monitoring year 1	1	0	0	1										20	35	65	0	0	858						
Former LEP student - monitoring year 2	1	0	0	1										9											
All Other Students	59	3	1	55	14	25	31	56	4	7	6	11	851	13,045	23	51	19	7	850						
IEP																									
Students with an IEP	11	2	0	9										2,170	5	29	38	28	836						
All Other Students	52	1	1	50	15	30	30	60	2	4	3	6	854	11,303	26	54	16	4	853						
SES																									
Economically Disadvantaged Students	28	1	0	27	5	19	14	52	4	15	4	15	847	5,900	13	48	27	12	845						
All Other Students	35	2	1	32	10	31	18	56	1	3	3	9	854	7,573	31	52	13	4	854						
Migrant																									
Migrant Students	1	0	0	1										2											
All Other Students	62	3	1	58	15	26	31	53	5	9	7	12	851	13,471	23	50	19	8	850						
Title I																									
Students Receiving Title I Services	2	0	0	2										937	9	47	33	12	844						
All Other Students	61	3	1	57	14	25	32	56	4	7	7	12	850	12,536	24	50	18	7	851						
504 Plan																									
Students with a 504 Plan	3	0	0	3										655	21	55	20	5	851						
All Other Students	60	3	1	56	14	25	30	54	5	9	7	13	850	12,818	23	50	19	8	850						

Level 4 = Proficient with Distinction; Level 3 = Proficient; Level 2 = Partially Proficient; Level 1 = Substantially Below Proficient
Note: Some numbers may have been left blank because fewer than ten (10) students were tested.

May 2014 - Grade 8 MEA Science Test

Questionnaire Results

SAU: Demonstration District A
State: Maine
Code: DEMA

QUESTIONNAIRE ITEMS	SAU										State											
	Students in Each Category	Level 4		Level 3		Level 2		Level 1		Mean Scaled Score	Students in Each Category	Level 4	Level 3	Level 2	Level 1	Mean Scaled Score	Students in Each Category	Level 4	Level 3	Level 2	Level 1	Mean Scaled Score
	%	N	%	N	%	N	%	N	%		%	%	%	%	%		%	%	%	%	%	
Which statement best describes how often and how long your science class meets?																						
A. We meet every day for 45 minutes to an hour.	79	13	30	24	55	2	5	5	11	853	74	25	51	18	6	851						
B. We meet on alternate days for 80 to 90 minutes.	9										16	20	52	20	7	849						
C. We meet every day for 45 minutes, plus a longer lab period each week.	5										3	17	43	24	16	845						
D. We have a flexible schedule depending on the activities.	7										7	13	47	26	13	845						
Which statement best describes how you learn science?																						
A. I read a textbook and answer questions and/or take notes and do assignments.	11										18	17	46	25	12	846						
B. I work in groups to design and conduct experiments.	18	3	30	4	40	1	10	2	20	847	16	12	48	26	14	844						
C. I do a combination of A and B, mostly A.	39	6	27	14	64	0	0	2	9	854	39	28	51	16	4	853						
D. I do a combination of A and B, mostly B.	32	6	33	12	67	0	0	0	0	860	28	26	54	15	4	853						
How often do you make observations and collect data in science class?																						
A. a few times a week	53	10	34	15	52	3	10	1	3	855	47	20	52	20	8	849						
B. a few times a month	33	3	17	11	61	1	6	3	17	849	35	27	50	17	5	852						
C. once a month	11										10	29	45	19	7	852						
D. never or almost never	4										8	20	50	19	12	848						
How do you feel about the following statement?																						
My knowledge of science will be useful to me as an adult.																						
A. strongly agree	23	3	23	6	46	1	8	3	23	848	20	36	45	14	6	854						
B. agree	55	9	29	18	58	2	6	2	6	854	58	23	52	19	6	851						
C. disagree	11										17	14	55	23	9	847						
D. strongly disagree	11										5	9	49	28	14	843						
Which of the following best describes how you rate yourself as a student in science?																						
A. very good	14										18	42	43	11	4	857						
B. good	55	7	23	20	65	1	3	3	10	853	52	24	54	16	6	852						
C. fair	27	5	33	7	47	3	20	0	0	854	25	11	50	29	10	845						
D. poor	4										4	4	39	34	24	838						
How well do the questions that you have just been given on this MEA test match what you have learned in school about science?																						
A. The questions on the test match what I have learned in science class.	14										17	30	47	15	7	853						
B. They match some of what I have learned.	57	9	28	18	56	2	6	3	9	852	57	25	51	18	6	851						
C. They match just a little of what I have learned.	23	5	38	6	46	1	8	1	8	857	23	15	53	23	9	847						
D. There is no match.	5										4	6	45	28	21	840						
Which courses do you plan to take before you graduate from high school?																						
A. earth and space science and/or biology	32	5	28	9	50	2	11	2	11	851	26	16	53	23	9	848						
B. the course(s) described in A, plus chemistry	14										23	28	50	16	6	852						
C. the course(s) described in B, plus physics	29	7	44	6	38	2	13	1	6	857	22	41	43	11	5	856						
D. a life science and physical science class	25	2	14	10	71	0	0	2	14	848	28	13	54	25	8	847						

Level 4 = Proficient with Distinction; Level 3 = Proficient; Level 2 = Partially Proficient; Level 1 = Substantially Below Proficient
Note: Some numbers may have been left blank because fewer than ten (10) students were tested.

August 2014

Maine Department of Education

2013-2014 School Year Reports

Dear School Board Members and School Personnel:

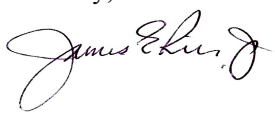
This 2013-2014 MEA Summary Report contains the results of student achievement in science at grades 5 and 8, as well as disaggregations by student and school characteristics. This report, together with MEA individual student and item analysis reports, provides support for use in program evaluation and planning.

The Maine Educational Assessment (MEA) is the State's measure of student progress in achieving the State science accountability standards of Maine's *Learning Results*. The *Learning Results* contain goals for what all students should know and be able to do at certain times in their school careers and include the accountability standards that are assessed for each grade. The MEA science test is administered to students in grades 5 and 8 to meet state assessment requirements and the federal requirements of the *No Child Left Behind* Act.

MEA results reflect scores based on the common science test questions that are taken by the approximately 14,000 students in each grade level. Students' scores are based on answers to a combination of multiple-choice questions and questions that require students to construct an answer. More information about the MEA is available at www.maine.gov/education/mea/index.htm.

Thank you for your continuing commitment to improve the quality and effectiveness of the instructional opportunities in your schools. These assuredly will help all students achieve the high standards of the *Learning Results* as demonstrated on classroom, district, and state assessments.

Sincerely,



James E. Rier, Jr.
Commissioner of Education

MEA
Maine Educational Assessment

Grade 8 Science School Report

Test Date: May 2014
Code: DEMA-DEM1
SAU: Demonstration District A
School: Demonstration School 1

Contents of the Report

The report is divided into four main sections including a section describing the students tested and a separate section for the content area results.

May 2014 - Grade 8 MEA Science Test

Grade Level Summary Report

School: Demonstration School 1
SAU: Demonstration District A
State: Maine
Code: DEMA-DEM1

Schools and SAUs administered the MEA Science tests to every enrolled student with the following exceptions: students who participated in the alternate assessment for the 2013-14 school year, students for whom a special consideration was granted through the state Department of Education, and other students for reasons not approved. On this page, and throughout this report, results are only reported for groups of students that are larger than nine (9).

PARTICIPATION SUMMARY	Number			Percentage		
	School	SAU	State	School	SAU	State
Students enrolled	**31**	**63**	**13,920**	**100**	**100**	**100**
	Science			Science		
Students tested	**29**	**59**	**13,473**	**94**	**94**	**97**
With an approved accommodation	1	6	1,780	3	10	13
Current LEP Students	0	2	399	0	3	3
With an approved accommodation	0	0	126		0	32
IEP Students	4	9	2,170	14	15	16
With an approved accommodation	1	5	1,443	25	56	66
Students not tested in MEA	**2**	**4**	**447**	**6**	**6**	**3**
State Approved	1	3	290	50	75	65
Alternate Assessment	1	2	240	100	67	83
Special Consideration	0	1	50	0	33	17
Other	1	1	157	50	25	35

PERFORMANCE SUMMARY

School													SAU						State					
Enrolled	Not Tested Approved	Not Tested Other	Tested	Level 4		Level 3		Level 2		Level 1		Mean Scaled Score	Tested	Level 4	Level 3	Level 2	Level 1	Mean Scaled Score	Tested	Level 4	Level 3	Level 2	Level 1	Mean Scaled Score
N	N	N	N	N	%	N	%	N	%	N	%		N	%	%	%	%		N	%	%	%	%	
31	1	1	29	8	28	18	62	1	3	2	7	854	59	25	54	8	12	851	13,473	23	50	19	8	850

Level 4 = Proficient with Distinction; Level 3 = Proficient; Level 2 = Partially Proficient; Level 1 = Substantially Below Proficient
Note: Throughout this report, percentages may not total 100 since each percentage is rounded to the nearest whole number.

May 2014 - Grade 8 MEA Science Test

Science Results

School: Demonstration School 1
SAU: Demonstration District A
State: Maine
Code: DEMA-DEM1

Proficient with Distinction (Level 4)
The student's work demonstrates in-depth understanding of essential concepts in science, including the ability to make multiple connections among central ideas. The student's responses demonstrate the ability to synthesize information, analyze and solve difficult problems, and explain complex concepts using evidence and proper terminology to support and communicate logical conclusions. (Scaled Score 862-880)

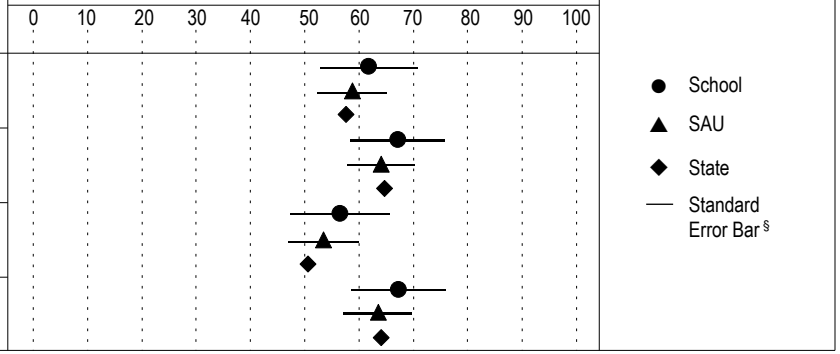
Proficient (Level 3)
The student's work demonstrates a general understanding of essential concepts in science, including the ability to make connections among central ideas. The student's responses demonstrate the ability to analyze and solve routine problems and explain central concepts with sufficient clarity and accuracy to demonstrate general understanding. (Scaled Score 842-860)

Partially Proficient (Level 2)
The student's work demonstrates incomplete understanding of essential concepts in science and inconsistent connections among central ideas. The student's responses demonstrate some ability to analyze and solve problems but the quality of responses is inconsistent. Explanation of concepts may be incomplete or unclear. (Scaled Score 828-840)

Substantially Below Proficient (Level 1)
The student's work demonstrates limited understanding of essential concepts in science and infrequent or inaccurate connections among central ideas. The student's responses demonstrate minimal ability to solve problems. Explanations are illogical, incomplete, or missing. There are many inaccuracies. (Scaled Score 800-826)

	Enrolled	Not Tested Approved	Not Tested Other	Tested	Level 4		Level 3		Level 2		Level 1		Mean Scaled Score
	N	N	N	N	N	%	N	%	N	%	N	%	
School													
2011-12	33	0	0	33	8	24	18	55	7	21	0	0	852
2012-13	33	1	1	31	4	13	15	48	9	29	3	10	847
2013-14	**31**	**1**	**1**	**29**	**8**	**28**	**18**	**62**	**1**	**3**	**2**	**7**	**854**
Cumulative Total	97	2	2	93	20	22	51	55	17	18	5	5	851
SAU													
2011-12	67	2	1	64	12	19	34	53	14	22	4	6	849
2012-13	66	2	2	62	15	24	29	47	12	19	6	10	850
2013-14	**63**	**3**	**1**	**59**	**15**	**25**	**32**	**54**	**5**	**8**	**7**	**12**	**851**
Cumulative Total	196	7	4	185	42	23	95	51	31	17	17	9	850
State													
2011-12	14,388	238	195	13,955	3,120	22	6,942	50	2,956	21	937	7	850
2012-13	14,026	227	156	13,643	2,965	22	6,582	48	2,822	21	1,274	9	849
2013-14	**13,920**	**290**	**157**	**13,473**	**3,078**	**23**	**6,774**	**50**	**2,600**	**19**	**1,021**	**8**	**850**
Cumulative Total	42,334	755	508	41,071	9,163	22	20,298	49	8,378	20	3,232	8	850

Subtopic	Total Possible Points
D. The Physical Setting	34
D1/D2 Space/Earth	17
D3/D4 Matter and Energy/Force and Motion	17
E. The Living Environment	22



The MEA assesses students' science knowledge based on questions that measure the science accountability content strands highlighted in Maine's 2007 Learning Results: Parameters for Essential Instruction, which can be found at: http://www.maine.gov/education/lres/pei/index.html.

§The standard error bar indicates how much the percent of points earned could vary if the students were examined multiple times with the same test.

May 2014 - Grade 8 MEA Science Test

Disaggregated Science Results

School: Demonstration School 1
SAU: Demonstration District A
State: Maine
Code: DEMA-DEM1

REPORTING CATEGORIES	School													SAU						State					
	Enrolled	Not Tested Approved	Not Tested Other	Tested	Level 4		Level 3		Level 2		Level 1		Mean Scaled Score	Tested	Level 4	Level 3	Level 2	Level 1	Mean Scaled Score	Tested	Level 4	Level 3	Level 2	Level 1	Mean Scaled Score
	N	N	N	N	N	%	N	%	N	%	N	%		N	%	%	%	%		N	%	%	%	%	
All Students	**31**	**1**	**1**	**29**	**8**	**28**	**18**	**62**	**1**	**3**	**2**	**7**	**854**	**59**	**25**	**54**	**8**	**12**	**851**	**13,473**	**23**	**50**	**19**	**8**	**850**
Gender																									
Male	17	1	1	15	4	27	9	60	1	7	1	7	856	29	31	41	10	17	851	6,893	25	48	19	8	850
Female	14	0	0	14	4	29	9	64	0	0	1	7	852	30	20	67	7	7	851	6,580	21	53	20	7	850
Not Reported	0	0	0	0										0						0					
Race/Ethnicity																									
Hispanic or Latino	1	0	0	1										1						214	14	47	29	10	846
Not Hispanic or Latino																									
American Indian or Alaskan Native	1	0	0	1										1						98	16	53	22	8	849
Asian	1	0	0	1										1						204	36	41	17	5	853
Black or African American	0	0	0	0										3						397	6	36	34	24	838
Native Hawaiian or Pacific Islander	1	0	0	1										1						16	6	94	0	0	854
White	26	1	1	24	7	29	14	58	1	4	2	8	854	51	27	53	8	12	851	12,347	24	51	19	7	850
Two or more races	1	0	0	1										1						197	16	54	20	10	848
No Race/Ethnicity Reported	0	0	0	0										0						0					
LEP Status																									
Current LEP student	0	0	0	0										2						399	3	35	37	26	836
Former LEP student - monitoring year 1	0	0	0	0										1						20	35	65	0	0	858
Former LEP student - monitoring year 2	0	0	0	0										1						9					
All Other Students	31	1	1	29	8	28	18	62	1	3	2	7	854	55	25	56	7	11	851	13,045	23	51	19	7	850
IEP																									
Students with an IEP	5	1	0	4										9						2,170	5	29	38	28	836
All Other Students	26	0	1	25	8	32	16	64	1	4	0	0	858	50	30	60	4	6	854	11,303	26	54	16	4	853
SES																									
Economically Disadvantaged Students	9	0	0	9										27	19	52	15	15	847	5,900	13	48	27	12	845
All Other Students	22	1	1	20	6	30	11	55	1	5	2	10	854	32	31	56	3	9	854	7,573	31	52	13	4	854
Migrant																									
Migrant Students	0	0	0	0										1						2					
All Other Students	31	1	1	29	8	28	18	62	1	3	2	7	854	58	26	53	9	12	851	13,471	23	50	19	8	850
Title I																									
Students Receiving Title I Services	1	0	0	1										2						937	9	47	33	12	844
All Other Students	30	1	1	28	7	25	18	64	1	4	2	7	853	57	25	56	7	12	850	12,536	24	50	18	7	851
504 Plan																									
Students with a 504 Plan	1	0	0	1										3						655	21	55	20	5	851
All Other Students	30	1	1	28	8	29	17	61	1	4	2	7	854	56	25	54	9	13	850	12,818	23	50	19	8	850

Level 4 = Proficient with Distinction; Level 3 = Proficient; Level 2 = Partially Proficient; Level 1 = Substantially Below Proficient
Note: Some numbers may have been left blank because fewer than ten (10) students were tested.

May 2014 - Grade 8 MEA Science Test

Questionnaire Results

School: Demonstration School 1
SAU: Demonstration District A
State: Maine
Code: DEMA-DEM1

QUESTIONNAIRE ITEMS	School										SAU						State					
	Students in Each Category	Level 4		Level 3		Level 2		Level 1		Mean Scaled Score	Students in Each Category	Level 4	Level 3	Level 2	Level 1	Mean Scaled Score	Students in Each Category	Level 4	Level 3	Level 2	Level 1	Mean Scaled Score
	%	N	%	N	%	N	%	N	%		%	%	%	%	%		%	%	%	%	%	
Which statement best describes how often and how long your science class meets?																						
A. We meet every day for 45 minutes to an hour.	86	7	29	15	63	1	4	1	4	856	79	30	55	5	11	853	74	25	51	18	6	851
B. We meet on alternate days for 80 to 90 minutes.	7										9						16	20	52	20	7	849
C. We meet every day for 45 minutes, plus a longer lab period each week.	4										5						3	17	43	24	16	845
D. We have a flexible schedule depending on the activities.	4										7						7	13	47	26	13	845
Which statement best describes how you learn science?																						
A. I read a textbook and answer questions and/or take notes and do assignments.	7										11						18	17	46	25	12	846
B. I work in groups to design and conduct experiments.	14										18	30	40	10	20	847	16	12	48	26	14	844
C. I do a combination of A and B, mostly A.	39	3	27	8	73	0	0	0	0	857	39	27	64	0	9	854	39	28	51	16	4	853
D. I do a combination of A and B, mostly B.	39	3	27	8	73	0	0	0	0	860	32	33	67	0	0	860	28	26	54	15	4	853
How often do you make observations and collect data in science class?																						
A. a few times a week	44	4	33	8	67	0	0	0	0	860	53	34	52	10	3	855	47	20	52	20	8	849
B. a few times a month	33										33	17	61	6	17	849	35	27	50	17	5	852
C. once a month	22										11						10	29	45	19	7	852
D. never or almost never	0										4						8	20	50	19	12	848
How do you feel about the following statement?																						
My knowledge of science will be useful to me as an adult.																						
A. strongly agree	18										23	23	46	8	23	848	20	36	45	14	6	854
B. agree	57	4	25	11	69	0	0	1	6	854	55	29	58	6	6	854	58	23	52	19	6	851
C. disagree	14										11						17	14	55	23	9	847
D. strongly disagree	11										11						5	9	49	28	14	843
Which of the following best describes how you rate yourself as a student in science?																						
A. very good	18										14						18	42	43	11	4	857
B. good	54	4	27	10	67	0	0	1	7	856	55	23	65	3	10	853	52	24	54	16	6	852
C. fair	25										27	33	47	20	0	854	25	11	50	29	10	845
D. poor	4										4						4	4	39	34	24	838
How well do the questions that you have just been given on this MEA test match what you have learned in school about science?																						
A. The questions on the test match what I have learned in science class.	14										14						17	30	47	15	7	853
B. They match some of what I have learned.	50	5	36	8	57	0	0	1	7	857	57	28	56	6	9	852	57	25	51	18	6	851
C. They match just a little of what I have learned.	32										23	38	46	8	8	857	23	15	53	23	9	847
D. There is no match.	4										5						4	6	45	28	21	840
Which courses do you plan to take before you graduate from high school?																						
A. earth and space science and/or biology	36	3	30	7	70	0	0	0	0	858	32	28	50	11	11	851	26	16	53	23	9	848
B. the course(s) described in A, plus chemistry	14										14						23	28	50	16	6	852
C. the course(s) described in B, plus physics	29										29	44	38	13	6	857	22	41	43	11	5	856
D. a life science and physical science class	21										25	14	71	0	14	848	28	13	54	25	8	847

Level 4 = Proficient with Distinction; Level 3 = Proficient; Level 2 = Partially Proficient; Level 1 = Substantially Below Proficient
Note: Some numbers may have been left blank because fewer than ten (10) students were tested.

State Summary

2013-2014 Science

State: Maine

Science	Enrolled	Not Tested Approved	Not Tested Other	Tested	Achievement Level								
	N	N	N	N	Level 4		Level 3		Level 2		Level 1		Mean Scaled Score
					N	%	N	%	N	%	N	%	
Maine	**40,790**	**706**	**870**	**39,214**	**4,873**	**12**	**18,723**	**48**	**9,867**	**25**	**5,751**	**15**	
Grade 5	13,296	215	101	12,980	1,301	10	6,859	53	3,783	29	1,037	8	546
Grade 8	13,920	290	157	13,473	3,078	23	6,774	50	2,600	19	1,021	8	850
High School	13,574	201	612	12,761	494	4	5,090	40	3,484	27	3,693	29	1141

Level 4 = Proficient with Distinction; Level 3 = Proficient; Level 2 = Partially Proficient; Level 1 = Substantially Below Proficient

SAU Summary

2013-2014 Science

SAU: Demonstration District A
State: Maine
Code: DEMA

Science	Enrolled	Not Tested Approved	Not Tested Other	Tested	Achievement Level									
					Level 4		Level 3		Level 2		Level 1		Mean Scaled Score	
	N	N	N	N	N	%	N	%	N	%	N	%	SAU	State
Demonstration District A	**257**	**7**	**7**	**243**	**22**	**9**	**125**	**51**	**48**	**20**	**48**	**20**		
Grade 5	67	2	1	64	1	2	39	61	18	28	6	9	544	546
Grade 8	63	3	1	59	15	25	32	54	5	8	7	12	851	850
High School	127	2	5	120	6	5	54	45	25	21	35	29	1142	1141

Level 4 = Proficient with Distinction; Level 3 = Proficient; Level 2 = Partially Proficient; Level 1 = Substantially Below Proficient

APPENDIX G—INTERACTIVE REPORTS

MEA
Maine Educational Assessment

C O N F I D E N T I A L

Science Item Analysis Report - May 2014

Grade 8

Date: 8/26/2014 9:30:49 AM
Code: DEMA-DEM1
Group Size: 31
SAU: Demonstration District A
School: Demonstration School 1

		Released Item	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	Content Strand Points Earned				Total Points Earned	Scaled Score	Achievement Level
		Content Strand	D2	E3	E1	D2	E4	E2	D1	D3	D3	E4	E2	D3	E5	D2	D4	D1	E5	D4	E2	E3	D2	E4							
		Depth of Knowledge Code	2	1	2	2	1	2	3	2	2	2	2	2	3	2	2	1	2	2	2	2	3	2	D. The Physical Setting			E. The Living Environment			
		Item Type	MC	MC	MC	MC	MC	MC	MC	MC	MC	MC	MC	MC	MC	MC	MC	MC	MC	MC	MC	MC	CR	CR							
		Correct MC Response	A	B	A	C	B	C	B	D	A	B	A	A	C	C	A	C	B	A	C	A			D. Total	D1/D2	D3/D4	E. Total			
Name/MEDMS ID		Possible Points	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	4	4	34	17	17	22	56		
Bholat, Abdullah	D08100024		+	+	D	+	+	D	+	C	+	D	+	C	+	+	+	D	+	+	+	B	2	2	24	12	12	14	38	858	3
Billham, Chelsea M	D08100058		B	+	B	+	+	+	+	+	+	+	+	C	A	+	+	+	+	+	+	+	3	4	23	12	11	18	41	862	4
Blair, Brandon	D08100040		+	+	C	+	+	D	A	+	+	A	+	B	D	+	D	+	+	+	+	+	2	3	23	14	9	15	38	858	3
Brooks, Colton A	D08100055		+	+	D	+	+	D	D	+	C	+	+	+	+	D	+	+	C	+	+	+	1	3	22	11	11	15	37	856	3
Brown, Shannon A	D08100005		+	+	+	+	+	+	+	+	C	A	C	D	A	A	C	+	+	+	D	+	2	3	20	11	9	13	33	850	3
Clark, Dilan I	D08100048		+	+	C	+	+	+	D	C	+	A	C	B	+	A	B	+	+	D	+	+	3	4	19	11	8	16	35	852	3
Clifford, Sara B	D08100046		+	+	+	+	+	+	+	+	+	+	+	D	+	+	D	+	+	+	+	+	4	4	31	17	14	20	51	880	4
Flanagan, Steven A	D08100057		C	A	D	B	A	D	+	+	C	A	+	+	A	+	C	D	C	+	A	+	2	2	19	12	7	9	28	840	2
Foster, Amy D	D08100004		C	+	C	+	+	+	A	A	+	+	+	+	D	D	C	+	+	B	A	+	1	1	17	10	7	14	31	846	3
Hatfield, Justina	D08100017		+	C	C	+	+	D	A	C	C	A	+	+	+	D	B	+	+	+	A	C	1	3	17	8	9	12	29	842	3
Healey, Williamg F	D08100035		+	+	B	+	+	+	+	C	+	D	+	C	+	D	+	+	+	+	+	+	2	3	22	14	8	17	39	858	3
Madden, Alicia R	D08100061		D	+	+	+	+	+	+	+	C	+	+	D	D	+	D	D	+	C	+	+	3	2	19	11	8	16	35	852	3
Malik, Doren D	D08100021		+	+	+	+	+	+	+	+	C	+	+	+	+	+	+	+	+	+	+	+	1	3	27	13	14	19	46	872	4
Merritt, Emma L	D08100023		-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	0	0	0	0	800	1
Miller, Steven T	D08100037		+	+	+	A	+	+	+	+	+	+	+	D	+	+	D	+	+	+	+	+	4	4	27	14	13	19	46	872	4
Nazarian, Jacob	D08100001		+	D	C	+	+	B	C	C	+	D	B	D	D	B	C	D	C	D	A	C	1	1	11	6	5	3	14	810	1
Phan, Jonathan V	D08100015		+	+	D	+	+	+	+	+	C	+	+	C	D	+	B	+	+	+	+	+	3	3	22	15	7	16	38	858	3
Pickett, Jasmine L	D08100032		C	+	D	+	+	+	D	C	B	+	+	D	D	+	+	D	+	+	+	+	2	3	16	8	8	17	33	850	3
Porter, Travis M	D08100025		+	C	+	+	+	A	C	C	D	+	+	C	+	+	B	+	+	B	+	C	2	3	18	11	7	13	31	846	3
Ramirez, Vanesa	D08100042		D	+	+	+	+	D	A	+	+	+	+	B	+	+	+	+	+	+	+	+	3	3	23	12	11	17	40	860	3
Rayborn, Jeremy I	D08100013		D	+	+	+	+	+	C	+	B	C	+	+	+	+	+	D	+	B	+	+	3	3	23	12	11	17	40	860	3
Roy, Briannah W	D08100053		D	+	B	+	+	D	+	+	+	+	+	+	+	+	+	D	+	C	+	+	3	4	26	13	13	17	43	866	4
Russell, Sam A	D08100002																								0	0	0	0	0		N
Soldano, Nicole	D08100038		D	+	B	+	+	+	+	+	+	C	+	+	+	+	+	+	+	B	A	+	3	3	28	13	15	16	44	868	4
Sparby, Brandt L	D08100063																								0	0	0	0	0		A

C O N F I D E N T I A L

Science Item Analysis Report - May 2014

Grade 8

Maine Educational Assessment

Date: 8/26/2014 9:30:49 AM
Code: DEMA-DEM1
Group Size: 31
SAU: Demonstration District A
School: Demonstration School 1

		Released Item	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	Content Strand Points Earned				Total Points Earned	Scaled Score	Achievement Level
		Content Strand	D2	E3	E1	D2	E4	E2	D1	D3	D3	E4	E2	D3	E5	D2	D4	D1	E5	D4	E2	E3	D2	E4							
		Depth of Knowledge Code	2	1	2	2	1	2	3	2	2	2	2	2	3	2	2	1	2	2	2	2	3	2	D. The Physical Setting			E. The Living Environment			
		Item Type	MC	MC	MC	MC	MC	MC	MC	MC	MC	MC	MC	MC	MC	MC	MC	MC	MC	MC	MC	MC	CR	CR							
		Correct MC Response	A	B	A	C	B	C	B	D	A	B	A	A	C	C	A	C	B	A	C	A			D. Total	D1/D2	D3/D4	E. Total			
Name/MEDMS ID		Possible Points	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	4	4	34	17	17	22	56		
Tuttle, Jacob X	D08100028		+	C	D	+	+	+	+	+	+	D	+	B	+	+	D	+	+	+	+	+	4	2	29	15	14	13	42	864	4
Upton, Heather J	D08100012		B	+	B	B	+	D	A	+	C	C	+	C	A	A	+	+	+	+	A	+	2	3	19	9	10	12	31	846	3
Waugh, Jaclyn	D08100059		B	+	+	+	+	+	D	B	C	+	D	D	+	D	B	+	+	+	+	+	3	3	21	12	9	17	38	858	3
Westlake, Michael D	D08100018		+	C	+	+	+	+	+	C	B	D	+	C	+	A	B	D	+	+	+	+	2	3	18	12	6	16	34	850	3
Woods, Elena S	D08100011		D	+	D	+	+	+	A	+	B	+	+	+	+	+	C	D	+	+	+	+	2	3	18	9	9	18	36	854	3
Worthington, Luc J	D08100006		+	C	+	+	+	+	+	+	+	+	+	+	+	+	C	+	+	+	+	+	2	4	28	14	14	21	49	880	4
Released Item			1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22							
Percent Correct/Avg. Score: Group			55	72	38	86	93	62	52	62	48	52	83	34	62	62	38	66	86	69	72	83	2.3	2.8	21.0	11.4	9.6	14.8			
Percent Correct/Avg. Score: School			55	72	38	86	93	62	52	62	48	52	83	34	62	62	38	66	86	69	72	83	2.3	2.8	21.0	11.4	9.6	14.8			
Percent Correct/Avg. Score: SAU			63	75	34	86	90	64	51	64	41	61	73	42	54	59	31	69	80	68	66	71	1.9	2.6	20.0	10.9	9.1	14.0			
Percent Correct/Avg. Score: State			60	73	38	86	89	68	61	67	30	59	77	33	56	59	30	66	70	68	68	78	1.9	2.6	19.6	11.0	8.6	14.1			

Legend for the Item Analysis Report – Science

May 2014

Released Item: This number corresponds to the item number in the released item documents. This report provides complete data on items that are being released, which are approximately 50% of the items used to calculate scores.

Content Strand: The letter indicates the standard with which the item is aligned as outlined in Maine's *2007 Learning Results: Parameters for Essential Instruction*. The performance indicator is also displayed.

Depth of Knowledge Code: This number indicates the Depth of Knowledge to which the item is coded.

Item Type: This indicates whether the question is multiple-choice (**MC**) or constructed-response (**CR**).

Correct MC Response: This is the correct letter response for multiple-choice questions.

Possible Points: The number indicates the maximum points awarded for the item: 1 point for a multiple-choice question and 0-4 points for a constructed-response question.

Student Item Results: Each student's name and state assigned student identification number are listed, followed by a score for each released item on the test included in this report.

- For multiple-choice (**MC**) questions only, a plus sign (+) indicates a correct response. If the student answered incorrectly, the letter of his or her response is indicated. An asterisk (*) indicates that the student selected more than one response.
- For constructed-response (**CR**) questions, a number indicates how many points a student earned for that item.
- For both MC and CR questions, a blank space indicates that the student left the question blank. A dash (–) means that the score was invalidated and that the student received no credit for parts of the test that were administered under non-standard conditions.

Content Strand Points Earned: These columns show the points the student earned in each content strand. The content strand points earned are based on all common items in the test and not just the released items.

Total Points Earned: This column shows the total number of points the student earned on all common items.

Scaled Score: This column shows the scaled score reported as a 3-digit number. The first digit is the grade and the next two digits are a score of 00-80. If the row is blank in this column, it means that the student was classified as Not Tested. (See Achievement Level below.)

Achievement Level: For Tested students, this column shows the achievement level into which the student's scores fall: **4** = Proficient with Distinction, **3** = Proficient, **2** = Partially Proficient, and **1** = Substantially Below Proficient. For Not Tested students, there are three reasons why a student did not participate: **A** = student participated in PAAP, **S** = state approved special consideration, and **N** = other reason.

Group/School/SAU/State Percent Correct/Average Score:

- **Released Items:** Percent correct refers to the percent of tested students who answered a multiple-choice item correctly. Average score refers to the average number of points awarded to all tested students for that constructed-response item.
- **Content Strand Points Earned:** Average score refers to the average number of points awarded to all tested students for that subcategory.

MEA
Maine Educational Assessment

Achievement Level Summary

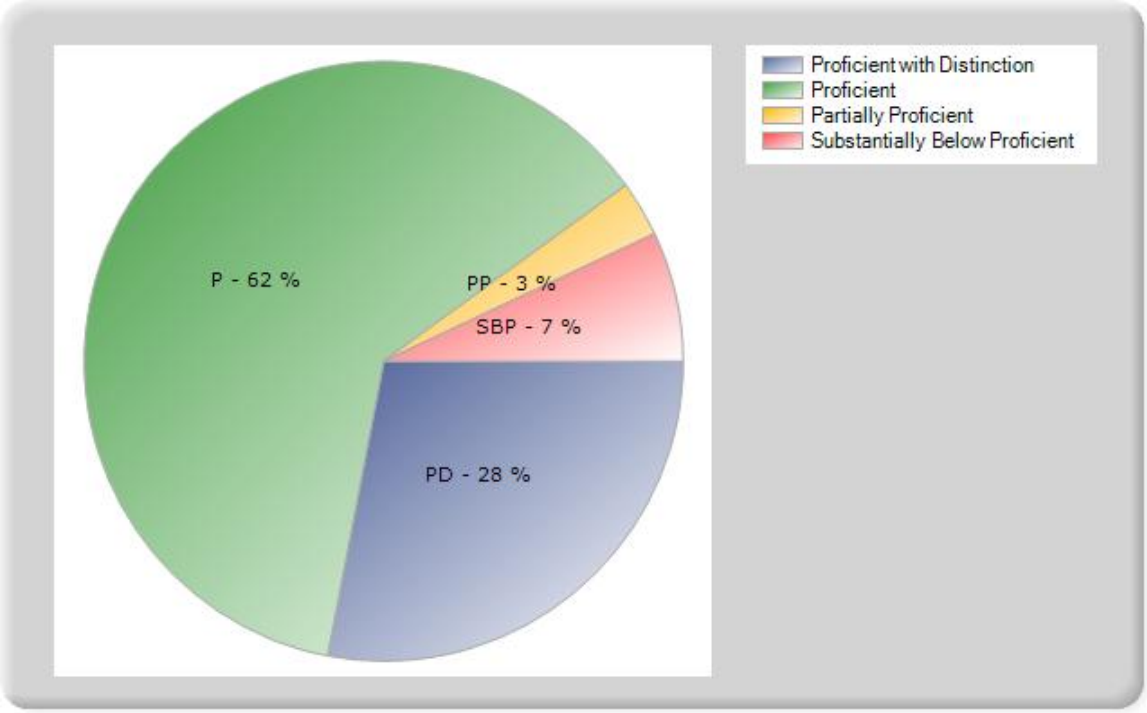
SAU: Demonstration District A

School: Demonstration School 1

Grade: 08

Date: 8/26/2014 9:34:46 AM

Science



Achievement Level	Count	Percentage %*
Proficient with Distinction	8	28
Proficient	18	62
Partially Proficient	1	3
Substantially Below Proficient	2	7

*Percentages may not total exactly 100% due to applied rounding.

Science Released Items Summary Data

SAU: Demonstration District A

School: Demonstration School 1

Grade: 08

Date: 8/26/2014 9:36:30 AM

Multiple Choice

Released Item	Content Strand	Correct (#)	A (#)	B (#)	C (#)	D (#)	IR (#)	Correct Response
1	D2	16	16	3	3	6	1	A
2	E3	21	1	21	5	1	1	B
3	E1	11	11	5	5	7	1	A
4	D2	25	1	2	25	0	1	C
5	E4	27	1	27	0	0	1	B
6	E2	18	1	1	18	8	1	C
7	D1	15	6	15	3	4	1	B
8	D3	18	1	1	8	18	1	D
9	D3	14	14	4	9	1	1	A
10	E4	15	5	15	3	5	1	B
11	E2	24	24	1	2	1	1	A
12	D3	10	10	4	7	7	1	A
13	E5	18	4	0	18	6	1	C
14	D2	18	4	1	18	5	1	C
15	D4	11	11	6	6	5	1	A
16	D1	19	0	0	19	9	1	C
17	E5	25	0	25	3	0	1	B
18	D4	20	20	4	2	2	1	A
19	E2	21	6	0	21	1	1	C
20	E3	24	24	1	3	0	1	A

Constructed Response

Released Item	Content Strand	Point Value	Average Score
21	D2	4	2.3
22	E4	4	2.8

CONFIDENTIAL

Longitudinal Data Report

Student Name
Kaelin Mitchell

Year	Enrolled Grade	School Name	Administration	Test Name	Content Area	Score	Achievement Level
1011	08	Demonstration School 1	NECAP Fall 2010	Grade 08 Mathematics	mat	864	Proficient with Distinction
1011	08	Demonstration School 1	NECAP Fall 2010	Grade 08 Reading	rea	878	Proficient with Distinction
1011	08	Demonstration School 1	MEA Science 2011	Grade 08 Science	sci	840	Partially Proficient
1011	08	Demonstration School 1	NECAP Fall 2010	Grade 08 Writing	wri	868	Proficient with Distinction
1112	08	Demonstration School 2	NECAP Fall 2011	Grade 08 Mathematics	mat	834	Partially Proficient
1112	08	Demonstration School 2	NECAP Fall 2011	Grade 08 Reading	rea	824	Substantially Below Proficient
1112	08	Demonstration School 1	MEA Science 2012	Grade 08 Science	sci	850	Proficient
1112	08	Demonstration School 2	NECAP Fall 2011	Grade 08 Writing	wri	807	Substantially Below Proficient
1213	08	Demonstration School 2	NECAP Fall 2012	Grade 08 Mathematics	mat	820	Substantially Below Proficient
1213	08	Demonstration School 2	NECAP Fall 2012	Grade 08 Reading	rea	838	Partially Proficient
1213	08	Demonstration School 2	MEA Science 2013	Grade 08 Science	sci	880	Proficient with Distinction
1213	08	Demonstration School 2	NECAP Fall 2012	Grade 08 Writing	wri	840	Proficient
1314	08	Demonstration School 1	NECAP Fall 2013	Grade 08 Mathematics	mat	845	Proficient
1314	08	Demonstration School 1	NECAP Fall 2013	Grade 08 Reading	rea	839	Partially Proficient
1314	08	Demonstration School 1	NECAP Fall 2013	Grade 08 Writing	wri	852	Proficient

Note: This report returns as many years of NECAP data as are available for this student beginning with 08-09.

APPENDIX H—DECISION RULES

Analysis and Reporting Decision Rules
Maine Educational Assessment
Spring 13-14 Administration

This document details rules for analysis and reporting. The final student level data set used for analysis and reporting is described in the "Data Processing Specifications." This document is considered a draft until the Maine State Department of Education (DOE) approves it. If there are rules that need to be added or modified after DOE approval, sign off will be obtained for each rule. Details of these additions and modifications will be in the Addendum section.

I. General Information

A. Test administered:

Grade	Subject	Items Included in Raw Score	DA Iref Table Reporting Categories		
			RepCat	Standard	Indicator
05	Science	Common	Cat3	Cat3	Cat4
08	Science	Common	Cat3, Cat4	Cat3	Cat4

B. Reports Produced:

1. Individual Student Report
2. Student Labels
3. Grade Level School/SAU/State Results
4. SAU All Grade Summary

C. Files Produced:

1. State Raw Data Files (by grade)
   a. (With names and without names)
2. State Scored Data Files (by grade)
   a. (With names and without names)
3. Invalidated Students Original Score
4. School Student Released Item Files (by grade)
5. SAU Student Released Item Files (by grade)
6. State Student Released Item Files (by grade)
7. Press Release (by grade)
   a. (School, SAU)
8. State Accommodation Frequency Report (by grade)
9. State Standard Deviations & Average Scaled Scores (by grade)
10. Teacher and Principal Questionnaire Raw Data
11. Teacher and Principal Questionnaire Frequency Distribution
12. Test Coordinator and Test Administrator Questionnaire Raw Data
13. Minimally Statistically Significant Differences for Scaled Scores (by grade and subject)
14. Standard Error of Measurement (by grade and subject)
15. Raw Score Ranges
16. Scaled Score Lookup
17. Released Item Percent Responses Data (For Program Management)
18. MEA School List for Mailing(For Program Management)

School Type:

SchType	Source: ICORE SubTypeID	Description
'PUB'	1	Public
'CHA'	11	Charter School
'PSP'	19	Public Special Purpose
'PSE'	15	Public Special Ed
'BIG'	6	Private with 60% or more Publicly Funded (Big 11
'PSN'	23	Private Special Purpose

School Type impact on Data Analysis and Reporting		
Level	**Impact on Analysis**	**Impact on Reporting**
Student	n/a	Report students based on testing discode and schcode. SAU data will be blank for students tested at BIG or PSN schools. Always print tested year state data.
School	Do not exclude any students based on school type using testing school code for aggregations	Generate a report for each school with at least one student enrolled using the tested school aggregate denominator. SAU data will be blank for BIG and PSN schools. Always print tested year state data.
SAU	For BIG and PSN schools, aggregate using the sending SAU. If BIG or PSN student does not have a sending SAU, do not include in aggregations.	Generate a report for each SAU with at least one student enrolled using the tested SAU aggregate denominator. Always report tested year state data.
State	Include all students regardless of schtype.	Always report testing year state data.

D. Student Status

StuStatus	Description
1	Homeschooled
2	Privately Funded
3	Exchange Student
4	Excluded State
0	Publicly Funded

StuStatus impact on Data Analysis and Reporting		
Level	**Impact on Analysis**	**Impact on Reporting**

Student	n/a	School and SAU data will be blank for students with a StuStatus value of 1,2 or 3. Always print tested year state data. For StuStatus values of 1, 2, and 3 print the description from the table above for the school and SAU names.
School	Exclude all students with a StuStatus value of 1, 2 or 3.	Students with a StuStatus value of 1, 2 or 3 are excluded from Interactive Reporting.
SAU	Exclude all students with a StuStatus value of 1, 2 or 3.	n/a
State	Exclude all students with a StuStatus value of 1, 2, 3 or 4.	n/a.

E. Requirements To Report Aggregate Data(Minimum N)

Calculation Description	**Rule**
Number and Percent at each achievement level, mean score by disaggregated category and aggregate level	If the number of tested students included in the denominator is less than 10, then do not report.
Content Area Subcategories Average Points Earned based on common items only by aggregate level	If the number of tested students included in the denominator is less than 10, then do not report.
Aggregate data on Item Analysis report	No required minimum number of students
Number and Percent of students in a participation category by aggregate level	No required minimum number of students
Content Area Cumulative Total Enrollment, Not tested, Tested, Number and Percent at each achievement level, mean score	Suppress all cumulative total data if at least one reported year has fewer than 10 tested students. The reported years are1112, 1213 and 1314.

F. Other Information

1. A non-public SAU code is a SAU associated with a school that is type BIG or PSN. Non-public testing sending SAU codes will be ignored.

2. Only students with a school type of BIG or PSN are allowed to have a sending SAU code. Sending SAU codes will be blanked for any other school type.

3. ICSEICSE(Infinte Campus State Edition) Linking

a. If a student is linked to ICSE, all demographic data of record are pulled from ICSE, with the exception of Homeschool. If the home school bubble was filled in but ICSE does not have the home school flag indicated for that student, the student will still be reported as homeschool.

b. If the student does not link to ICSE, then report the bubbled student number and demographics from the booklet. These students will be reported to the Login school and will be assigned to the 'not' group for all demographics that exist only in ICSE.

4. If a student tested off grade, the results for that student are suppressed and the student is added at the ICSE grade with no test data. The student is reported as defined by the rules described in this document.

5. If a student did not test and meets the following criteria, the student will be added to the enrollment at the school and grade indicated in ICSE, with no test data:

a. Is actively enrolled (Active = '1' in ICSE)

b. Is a Maine Resident (NonMaine = '0' in ICSE)

c. Enrolled at a 'PUB' , 'CHS', 'PSP' or 'PSE' school, or

d. Enrolled at a 'BIG' or 'PSN' school and has a sending SAU

i. The student is reported as defined by the rules described in this document.

II. **Student Participation / Exclusions**

A. Test Attempt Rules (by subject)

1. A valid multiple-choice response is A, B, C, D, or a multiple response (denoted by an asterisk).

2. A student attempted the test if either:

a. The student provided a valid response to at least one multiple choice item.

b. The student has a non-blank score to a common open-response item.

B. Not Tested Reasons (by subject)

1. If a student has more than one reason for not participating on the test, we will assign one participation code using the following hierarchy:

a. Special Consideration

i. If a student links to the demographic data file has content area "Not Tested State Approved Special Consideration" indicated, then the student is identified as "Not Tested State Approved Special Consideration".

b. Alternate Assessment

i. Students are identified as participating in the MEA Alternate Assessment based on the MEA Alternate Assessment Decision Rules for each subject assessed at the grade level in ICSE.

c. Not Tested Other

i. If content area test was not attempted, the student is identified as "Not Tested Other".

C. Special Circumstances (by Subject)

1. Item invalidation flags are provided to the DOE during data processing test clean up. The item invalidation flag variables are initially set using the rules below. The final values used for reporting are provided back to Measured Progress by the DOE and used in reporting.

a. If sciaccomM3 is marked, then mark sciInvSes1, sciInvSes2, and sciInvSes3.

2. A student is identified as content area tested if the student does not have any content area not tested reasons identified. Tested students are categorized in one of the four tested participation statuses: "Tested Damaged SRB", "Tested with Non-Standard Accommodations", "Tested Incomplete", and "Tested".

a. Students with a common item response of 'X' are identified as "Tested Damaged SRB".

b. Students identified as content area tested, are not identified as "Tested Damaged SRB", and have at least one of the content area invalidation session flags marked will be identified as "Tested with Non-Standard Accommodations".

c. Students identified as content area tested, are not identified as "Tested Damaged SRB", and not identified as "Tested with Non-Standard Accommodations" and did not attempt all sessions in the test are considered to be "Tested Incomplete."

d. All other tested students are identified as "Tested".

3. For students identified as "Tested Damaged SRB", the content area subcategories with at least one damaged item will not be reported. These students are excluded from all raw score aggregations (item, subcategory, and total raw score). They are included in participation, achievement level, and scaled score aggregations.

4. For students identified as "Tested with Non-Standard Accommodations" the content area sessions item responses which are marked for invalidation will be treated as a non-response.

5. Students identified as tested in a content area will receive released item scores, scaled score, scale score bounds, achievement level, raw total score, and subcategory scores.

6. Students identified as not tested in a content area will not receive a scaled score, scaled score bounds, or achievement level. They will receive released item scores, raw total score, and subcategory scores.

7. Item scores for students with an invalidation flag marked and have a not tested status will be blanked out based on the invalidation flag. For example, if the student is identified as "Not Tested: State Approved Alternate Assessment" for science and has SciInvSes1 marked, then all science session 1 item responses will be reported as a blank.

8. For students identified as any of the four tested participation categories they will be considered to be either tested with or without accommodations for participation purposes. A student is tested with accommodations if any of the accommodations were marked.

D. Student Participation Status Hierarchy (by subject)

1. Not Tested State Approved Special Consideration

2. Not Tested State Approved Alternate Assessment

3. Not Tested Other

4. Tested Damaged SRB

5. Tested with Non-Standard Accommodations

6. Tested Incomplete

7. Tested

Student Participation Summary (by subject)

Participation Status	**Description**	**Raw Score (*)**	**Scaled Score**	**Ach. Level**	**Student Report Ach. Level Text**	**Roster Ach. Level Text**
Z	Tested Damaged SRB(**)	✓	✓	✓	Substantially Below Proficient, Partially Proficient, Proficient, or Proficient with Distinction	1,2,3, or 4
A	Tested	✓	✓	✓	Substantially Below Proficient, Partially Proficient, Proficient, or Proficient with Distinction	1,2,3, or 4
C	Tested with Non-Standard Accommodations (%%)	✓	✓	✓	Substantially Below Proficient, Partially Proficient, Proficient, or Proficient with	1,2,3, or 4

					Distinction	
B	Tested Incomplete(%)	✓	✓	✓	Substantially Below Proficient, Partially Proficient, Proficient, or Proficient with Distinction	1,2,3, or 4
D	Not Tested State Approved Alternate Assessment	✓			Alternate Assessment	A
H	Not Tested State Approved Special Consideration	✓			Special Consideration	S
I	Not Tested Other	✓			Not Tested	N

* If a student has a participation status of Special Considerations and/or Alternate Assessment for all subjects assessed at the grade level, a Parent Letter is not produced.

(*) Raw scores are not printed on student report for students with a not tested status.

(**) Raw scores for Tested damaged SRB students will be reported based on the set of non-damaged items. Subcategory scores will not be reported if it includes a damaged item.

(%) Tested incomplete students will be identified on the student report with a footnote.

(%%) Tested with Non-standard accommodations students will be identified on student report with a footnote. The invalidated items will be stored as a '-'for item analysis.

III. **Calculations**

A. Rounding Table

Report	**Calculation**	**Rounded (to the nearest)**
Summary of Scores	Average Scaled Score	Whole value
Summary of Student Participation	All percents	Whole value
Results	Percent at each achievement level, Percent of points possible, Percent of students in each Category, Scaled Score	Whole value
	Cluster Average Points Attained (Number & Percent) Round Cluster Average Points Attained before calculating the percent.	Tenth
Item Analysis Report	Multiple Choice Percent Correct	Whole value
	Open Response Average Score	Tenth
	Content Standard Earned Averages	Tenth
All Grade Summary	Percent at each achievement level, Mean Scaled Score	Whole value

B. Raw scores

1. Raw scores are based on the scores for common items.

C. Released Item Data

1. The data for the released items are provided by Program Management or exist in IABS.

2. Details on how the standard/performance indicators are derived can be found later in the document under the Content Standards section.

D. Item Scores

1. For all analysis, non-response for an item by a tested student is treated as a score of 0.

2. For multiple choice released item data store a '+' for correct response, or A,B,C,D,* or blank

3. For open response released items, store the student score. If the score is not numeric ('B'), then store it as blank.

4. For students identified as content area tested with non-standard accommodations, then store the released item score as '-' for invalidated items.

E. Item Averages and Percents (on the Item Analysis Report)

1. Multiple-choice item averages are multiplied by 100 and rounded to the nearest whole number.

2. Open-response item averages are rounded to the nearest tenth.

F. Students included in calculations based on participation status

1. For number and percent of students enrolled, tested, and not tested categories include all students not excluded by other decision rules.

2. For number and percent at each achievement level, average scaled score, subcategories average points earned, percent/correct average score for each released item include all tested students not excluded by other decision rules.

3. Students identified as Tested Damaged SRB are excluded from all raw score aggregations (item, subcategory, and total raw score). They are included in participation, achievement level, and scaled score aggregations.

G. Cumulative Total

1. Include the yearly results where the number tested is greater than or equal to 10

2. Cumulative total N for Enrolled, Tested, for Not Tested Approved, and Not Tested Other number at each achievement level is the sum of the yearly results for each category where the number tested is greater than or equal to 10.

3. Cumulative percent for each achievement level is 100*(Number of students at the achievement level cumulative total / number of students tested cumulative total) rounded to the nearest whole number.

4. Cumulative mean scaled score is a weighted average. For years where the number tested is greater than or equal to 10, (sum of (yearly number tested * yearly mean scaled score)) / (sum of yearly number tested) rounded to the nearest whole number.

H. Average Points Earned Students at Proficient Level (Range)

1. Select all students across the states with Y40 scaled score, where Y=grade. Average the content area subcategories across the students and round to the nearest tenth. Add and subtract one standard error of measurement to get the range.

I. Content Standards

(Please note that all references to Cluster are for MP internal purposes only. Cluster is displayed as Content Standard/Performance Indicator for all content areas)

1. Science

a. The standard/performance indicator is calculated by concatenating the third and fourth portion of content framework. This is stored in tblIref in the Standard column.

b. Grade 5

i. RepCat1: D. The Physical Setting

- This is defined in IABS when standard ='D' .

ii. RepCat2: E. The Living Environment

- This is defined in IABS when standard = 'E' .

c. Grade 8

i. RepCat1: D. The Physical Setting

- This is defined in IABS when standard = 'D' .

ii. RepCat2: D1/D2 Earth/Space

- This is defined in IABS when standard = 'D' and performance indicator = '1' or '2'.

iii. RepCat3: D3/D4 Matter/Energy/Force/Motion

- This is defined in IABS when standard = 'D' and performance indicator = '3' or '4'.

iv. RepCat4: E. The Living Environment

- This is defined in IABS when standard = 'E' .

J. Participation

1. For participation calculate the number and percent of students in each of the following categories by school, district, and state according to schtype and stustatus decision rules.

2. Note that a student is tested with approved accommodations if one is tested, has a non-M accommodation marked, and does not have the M3 accommodation marked for that subject.

a. For Students Enrolled, Students Tested, and Students Not Tested the denominator will be the number of students enrolled

b. For Students Tested with approved Accommodations, Current LEP Students Tested (LEP=1), and IEP Students Tested the denominator will be the number of students tested.

c. For Current LEP Students Tested with approved accommodations (LEP=1 the denominator will be the number of current LEP students tested.

d. For IEP Students Tested with approved accommodations the denominator will be the number of IEP students tested.

e. For Students Not Tested State Approved and Not Tested Other the denominator will be the number of students not tested.

f. For Students Not Tested Alternate Assessment and Special Considerations the denominator will be the number of students not tested state approved.

K. Questionnaire

1. Only tested students will be included in the calculations.

2. Percent of students in this category is computed by the number of tested students that selected that response/number of tested students with a single response for the question *100. Students are considered to have a single response, if their response is not blank or '*'.

3. Achievement level data will be suppressed based on minimum N rules.

L. Scaling: Assignment of Scaled Score and Achievement Level

1. Scaling is accomplished by defining the unique set of test forms for the grade/subject. This is accomplished as follows:

a. Translate each form and position into the unique item number assigned to the form/position.

b. Order the items by

i. Type – multiple-choice, short-answer, constructed- response, extended-response, writing prompt.

ii. Form – common, then by ascending form number.

iii. Position

c. If an item number is on a form, then set the value for that item number to '1', otherwise set to '.'. Set the Exception field to '0' to indicate this is an original test form.

d. If an item number contains an 'X' (item is not included in scaling) then set the item number to '.'. Set the Exception field to '1' to indicate this is not an original test form.

e. Compress all of the item numbers together into one field in the order defined in step II to create the test for the student.

f. Select the distinct set of tests from the student data and order them by the exception field and the descending test field.

g. Check to see if the test has already been assigned a scale form by looking in the tblScaleForm table. If the test exists then assign the existing scale form. Otherwise assign the next available scale form number. All scale form numbering starts at 01 and increments by 1 up to 99.

2. Scaled Score assignment

a. Psychometrics provides data analysis with a lookup table for each scale form. The lookup table contains the raw score and the resulting scaled score.

b. Scaled Scores are rounded to even integers.

3. Achievement level coding:

a. 1 = Substantially Below Proficient

b. 2 = Partially Proficient

c. 3 = Proficient

d. 4 = Proficient with Distinction

M. Linking to Historical Aggregate Data

1. Some Maine schools were redistricted. Since the concatenation of discode and schcode is used to store school level aggregate data, to link historical data the SAU code for prior school data will be updated to the current SAU code.

II. **Report Specific Rules**

A. On all reports, grade is printed as 1 digit (5,8).

B. Always print data based on minimum N-size and school type decision rules.

C. For achievement level data if the number of students in an achievement level does not equal 0, and the percent of students is 0 then format the percent as <1.

D. Student Labels

1. Student name is printed last name, first name middle initial. If a student is missing a first and last name, then report as 'NAME NOT PROVIDED'.

2. If the student participated in the MEA, the scaled score is printed along with the achievement level text. Otherwise, the text is printed, based on the participation status.

3. If a student has a participation status of Special Consideration and/or Alternate Assessment a label is not produced (ParentLetter = '0').

4. If a student is Home schooled, a label is not produced.

5. SAU code concatenated with the school code is printed at the bottom of each page of student labels.

E. Individual Student Report

1. Cover Page

a. Print "FNAME MI. LNAME".

b. For school name do the following.

i. For students with a stustatus value of 0 or 4, print the abbreviated tested school and district ICORE name based on school type decision rules.

ii. Otherwise, for the school names print the "Description" in the StuStatus table presented earlier in this document.

2. Test Results

a. For the graphic display:

i. Display the scaled score ranges in the graphic display.

ii. Draw the vertical green bar up to the height indicated by the scaled score and print the scaled score. The shade of the vertical bar will match the shade associated with the performance level.

iii. Draw the vertical gray bar to indicate the scaled score bounds.

b. For the text box:

I. Print "Your child's score is [ScaledScore] which is in the [PerfLevel] achievement level."

II. Print "The smaller gray bar shows that you child's score could fall between [LowScaledScore] and [HighScaledScore] if he or she had taken the test multiple times."

3. Achievement Level Descriptor

a. Print the paragraph and bulleted list associated with the student's earned achievement level.

4. Your Child's Performance Compared to Other Students

a. For school, district, and state color portions of the bars the respective shade of green for each achievement level and print the percent in each achievement level to the right and in the vertical center of the colored portion.

I. If district data are suppressed because of schtype decision rules then, print '**' after the word District and "**SAU data are not available." at the bottom of the page.

II. If school or district data are suppressed because of minimum N decision rules then, print '*' after the word School or District and "*To protect student confidentiality, some numbers may have been left blank because fewer than ten (10) students were tested." at the bottom of the page.

b. Display the bars for school, district, and state based on student status, StuGrade, school type and minimum N rules.

c. Place an arrow to the left and in the vertical center of the portion of the bar associated with the students earned achievement level.

5. Your Child's Answers to Some Test Questions

a. Print the science standard for each released item.

b. For the row Your Child's Answer

i. For multiple choice released item data print a '+' for correct response, or A,B,C,D,* or blank.

ii. For open response released items, print the student score. If the score is not numeric ('B'), then print it as blank

iii. For students identified as content area tested with non-standard accommodations, then print the released item score as '-'for invalidated items.

6. MEA Released Question:

a. If the student chose a valid response, print "Your child chose [response]." Where [response] is the letter he chose.

b. If the student chose multiple responses, print "Your child chose multiple responses."

c. If the student left the item blank, print "Your child left this question blank."

7. A Closer Look at Your Child's Performance

a. Always print total possible points for each reporting category.

b. For students identified as "Tested Damaged SRB" do not report student points earned for reporting categories that have at least one damaged item. Also do not place a check in any of the boxes for this reporting category.

8. For all other tested students print the points earned for the reporting categories.

a. If the student's points earned are less than the lower bound of the range for proficient students, the place a check in the Below Proficient box.

b. If the student's points earned fall within the range for proficient students, the place a check in the Similar to Proficient Students box.

c. If the student's points earned are greater than the upper bound of the range for proficient students, the place a check in the Above Proficient Students box.

F. Grade Level School/SAU/State Results

1. Reports are run by state, SAU, school using the aggregate school and SAU codes described in the school type table.

2. Report Header Information

a. "Spring YYYY Grade XX MEA Science Tests" where XX is the grade level and YYYY is the year, will print as the title.

b. Use abbreviated school and SAU name from ICORE based on school type decision rules.

c. For ME print SAU and school codes separated by '-' for Code on first page for the school level. Print the SAU code for the SAU level. Print the full state name for the state level.

3. Report Section: Participation in MEA Science

a. Tested students will be aggregated as tested with or without accommodations and by IEP status.

4. Report Section: MEA Results

5. Report Section: Historical MEA Science Results

6. Report Section: Subtopic Results by content area

7. Report Section: Disaggregated Results

8. Report Section: Questionnaire Results

G. SAU All Grade Summary

1. Report Header Information

a. Use abbreviated school and SAU name from ICORE based on school type decision rules.

b. Print "Maine" to reference the state. The state graphic is printed on the first page.

c. Print SAU and school codes separated by '-' for Code on first page for the school level. Print the SAU code for the SAU level. Print the full state name for the state level.

2. Reports are run by SAU using the aggregate school and SAU codes described in the school type table.

3. Exclude students based student status, school type and participation status decision rules for aggregations.

4. Print entire aggregate group across grades tested and list grades tested results based on minimum N-size and school type decision rules. Mean scores across the grades is not calculated.

III. **Data Requirements Interactive Reporting**

A. Student Level

1. Refer to Sections III.E for decision rules on how student test data will be stored.

2. Students will be loaded into the Interactive System based on the Interactive flag in tblStuDemo. Students with Interactive flag set to 0 will not be loaded into the system. Students with Interactive set to 1 will be loaded.

a. Students with StuStatus value of 1, 2 or 3 will have the Interactive flag set to 0.

b. All others will have Interactive=1.

3. The Included flag will determine which students are included in school level aggregations. Students with Included=0 are excluded from all aggregations. Students with Included=2 will be included in Performance Level aggregations and excluded from raw score aggregations (item, subcategory, and total raw score). Students with Included=1 will be included in all school level aggregations.

a. Students with a Not Tested Participation Status, StuStatus=1, 2, or 3 will have their Included flag set to 0.

b. Students who do fall into the above group and have Participation Status of Tested Damaged SRB will have their Included flag set to 2.

c. All other students will have their Included flag set to 1.

4. Longitudinal Data

a. Only students with a valid StudentID and Interactive flag=1 will be loaded.

b. The complete achievement level name or not tested reason will be stored.

B. Aggregate Level

1. Data Analysis will compute Item Averages for the whole group only at the School and SAU Levels.

2. Data Analysis will compute Item Averages for all of the filter combinations that exist at the State Level.

3. Data Analysis will create a lookup table with all of the possible filter combinations. It will contain the variable Filter with length 5. Each position represents one of the filter variables. It will contain all the possible combinations of the values plus nulls for when variables are not selected. The first position will be Gender, second Ethnic, third IEP, fourth LEP, and fifth EconDis.

4. Data Analysis will compute Item Averages, Achievement Level Summary, and Item Summary data for the filter combinations for a sample of schools for quality assurance reveiw.

a. For this sample, percents will be rounded to the nearest whole number and open response average scores will be rounded to the nearest tenth.

b. For the Item Summary data, item responses other than A, B, C, and D will be counted in the IR column.

IV. **Data File Rules**

A. State Raw Data

1. Only students from 'PUB','CHA', 'PSP', or 'PSE' schools are included, or if they have a sending SAU. Students with all participation statuses are included.

2. Exclude students with StuStatus=1, 2, 3, or 4.

3. There are two files per grade; one with names and one without.

4. Field test item responses are not displayed.

5. Data is ordered by SAU code, School code, last name, first name.

B. State Scored Data

1. Only students from 'PUB','CHA','PSP', or 'PSE' schools are included, or if they have a sending SAU. Students with all participation statuses are included.

2. Exclude students with StuStatus=1, 2, 3, or 4.

3. There are two files per grade; one with names and one without.

4. Field test item responses are not displayed.

5. Data is ordered by subject, SAU code, School code, last name, first name.

C. SAU Student Released Item Files

1. CSV files will be created for each grade and SAU.
2. Only public school SAUs will receive SAU data files. (SAUs with at least one school with school subtypeid=1, 19, or 15)
3. Exclude students with StuStatus=1, 2, or 3.

4. Students with the Discode or SendDiscode will be in the SAU grade specific CSV file.

D. School Student Released Item Files

1. CSV files will be created for each grade and school.
2. Exclude students with StuStatus=1, 2, or 3.

3. Students with the SchCode will be in the SAU grade specific CSV file.

E. Press Release

1. The data reported in these files are the number of students tested, the number and percent of students performing at each achievement level, and the average scaled score.
2. The SAU files are only produced for public school SAUs. A student with a sending SAU is aggregated only to the sending SAU.

3. The school files are only produced for 'PUB','PSP', 'BIG' and 'PSE' schools.

4. Schools and SAUs that have less than 10 included students will only include data for the number of students tested.

F. State Accommodation Frequency Report

1. Exclude students with stustatus=1, 2, 3, or 4 and students with not tested participation status.

2. The data reported in these files are the counts of each accommodation for each subject.

G. State Standard Deviations & Average Scaled Scores

1. Exclude students with stustatus=1, 2, 3, or 4 and students with not tested participation status.

2. The data reported in these files are the number of students tested, the average scaled score, and the standard deviations for the following subgroups:

a. Identified Disability, No Identified Disability

b. LEP (1st year or 2nd year and beyond), Not LEP

c. Economically Disadvantaged, Not Economically Disadvantaged

d. Migrant, Not Migrant

e. Gender

f. Ethnicity

g. Title 1, Not Title 1

h. Total (All students)

H. Invalidated Students Original Score

1. A CSV file will be created for each grade.

2. Original raw scores for students whose responses were invalidated for reporting will be provided.

3. Exclude students with stustatus=1, 2, 3, or 4.

I. Teacher and Principal Questionnaire Raw Data

1. One CSV file will be created containing raw Teacher Questionnaire data.

2. One CSV file will be created containing raw Principal Questionnaire data.

J. Teacher and Principal Questionnaire Frequency Distribution

1. One CSV file will be created containing the distribution of responses of Teacher Questionnaire raw data.

2. One CSV file will be created containing the distribution of responses of Principal Questionnaire raw data.

K. Test Coordinator and Test Administrator Questionnaire Raw Data

1. One CSV file will be created containing raw Test Coordinator data.

2. One CSV file will be created containing raw Test Administrator data.

L. Minimally Statistically Significant Differences for Scaled Scores

The data reported in this file are the number of scaled score points denoting minimally statistically significant differences for average school/SAU results. This is calculated by the psychometricians.

M. Standard Error of Measurement

1. The data reported in this file are the number of students tested, the number of possible raw score points, the minimum, maximum, mean, standard deviation, reliability, and the SEM of the raw score.

N. Released Item Percent Responses Data

1. One excel file will be produced for each grade.

2. Will contain average score and percent at each response for the released items.

O. Data File Table
(YYYY indicates year, GG indicates grade, SSS indicates subject)

Data File	**Layout**	**File Name**
State Raw Data	MEA1314StateStudentRawDataLayout.xls	MEA1314StateStudentRawDataGrade[GR].csv MEA1314StateStudentRawDataNoNamesGrade[GR].csv
State Scored Data	MEA1314StateStudentScoredDataLayout.xls	MEA1314StateStudentScoredDataGrade[GR].csv MEA1314StateStudentScoredDataNoNamesGrade[GR].csv
School Student Released Item	MEA1314StudentReleasedItemLayout.xls	MEA1314SchoolSlice[GR]]_[SAU Code].[School Code].csv
SAU Student Released Item	MEA1314StudentReleasedItemLayout.xls	MEA1314DistrictSlice[GR]_[SAU Code].csv
State Student Released Item	MEA1314StudentReleasedItemLayout.xls	MEA1314StateStudentReleasedItem[GR].csv
Press Release	MEA1314PressReleaseLayout.xls	MEA1314SchoolPressRelease[GR].csv MEA1314DistrictPressRelease[GR].csv
State Accommodation Frequency Report	N/A	MEA1314Accommodation[GR].xls
State Standard Deviations & Average Scaled Scores	MEA1314StateStandardDeviationsLayout.xls	MEA1314StandardDeviation[GR].xls
Minimally Significant Differences for Scaled Scores	N/A	MEA1314SignificantDifferenceChart.xls
Standard Error of Measurement	N/A	MEA1314SEM.xls
Raw Score Ranges	N/A	MEA1314ScoreRanges.xls
Scaled Score Lookup	N/A	MEA1314ScaledScoreLookup.xls
MEA School Mailing List	N/A	MEA1314SchDisList.xls
Invalidated Students Original	MEA1314StateInvalidatedStudentOriginalScoredLayout.xls	MEA1314StateInvalidatedStudentOriginalScored.csv

Score		
Teacher and Principal Questionnaire Raw Data	MEA1314TeacherQuestionnaireRawLayout.xls MEA1314PrincipalQuestionnaireRawLayout.xls	MEA1314TeacherQuestionnaireRaw.csv MEA1314PrincipalQuestionnaireRaw.csv
Teacher and Principal Questionnaire Frequency Distribution	MEA1314TeacherPrincipalQuestionnaireFreqLayout.xls	MEA1314TeacherPrincipalQuestionnaireFreq.csv
Test Coordinator and Test Administrator Questionnaire Raw Data	TestCoordinatorQuestionnaireRawLayout.xlsx TestAdministratorQuestionnaireRawLayout.xlsx	MEA1314TestCoordinatorQuestionnaireRaw.xlsx MEA1314TestAdministratorQuestionnaireRaw.xlsx
Released Item Percent Responses Data	MEA1314ReleasedItemPercentResponsesLayout.xls	ReleasedItem[GR].xls

APPENDIX I—CLASSICAL ITEM ANALYSIS

Table I-1. 2013–14 MEA Science: Item-Level Classical Test Theory Statistics—Science Grade 5

Item		*Difficulty*	*Discrimination*	*Percent Omitted*
Number	*Type*			
64371	MC	0.93	0.23	0
64431	MC	0.53	0.24	0
64790	MC	0.81	0.32	0
64805	MC	0.87	0.36	0
64956	CR	0.50	0.39	0
66441	MC	0.67	0.37	0
67897	MC	0.84	0.22	0
67899	MC	0.68	0.20	0
67903	MC	0.90	0.27	0
96237	MC	0.74	0.35	0
96258	MC	0.76	0.17	0
96260	MC	0.72	0.37	0
96261	MC	0.57	0.35	0
96266	MC	0.47	0.28	0
96280	MC	0.69	0.35	0
96338	MC	0.87	0.32	0
96368	MC	0.77	0.25	0
96577	MC	0.87	0.33	1
97907	MC	0.32	0.19	0
158021	MC	0.68	0.26	0
158023	MC	0.85	0.24	1
189933	CR	0.44	0.35	2
190086	MC	0.54	0.25	0
190139	MC	0.91	0.14	1
190158	MC	0.85	0.23	0
190321	MC	0.61	0.30	1
228101	CR	0.37	0.45	1
228127	MC	0.76	0.33	0
228136	MC	0.59	0.22	0
228143	CR	0.33	0.44	1
228148	MC	0.70	0.35	0
228155	MC	0.74	0.19	0
228159	MC	0.44	0.18	1
248520	MC	0.47	0.19	0
248572	MC	0.87	0.23	0
248584	MC	0.69	0.25	0

Table I-2. 2013–14 MEA Science: Item-Level Classical Test Theory Statistics—Science Grade 8

Item		*Difficulty*	*Discrimination*	*Percent Omitted*
Number	*Type*			
2610	MC	0.77	0.30	0
63394	CR	0.65	0.33	1
65206	MC	0.61	0.35	1
96394	CR	0.45	0.45	1
96450	MC	0.47	0.36	0
96454	MC	0.78	0.44	0
96660	MC	0.72	0.32	0
97218	MC	0.93	0.33	0
97788	MC	0.58	0.32	0
97791	MC	0.84	0.30	0
99705	MC	0.55	0.36	0
99750	MC	0.46	0.25	0
99758	MC	0.69	0.26	1
100714	CR	0.32	0.51	1
153649	MC	0.72	0.26	0
153664	MC	0.87	0.40	0
153680	MC	0.68	0.41	0
153806	MC	0.85	0.32	0
158374	MC	0.49	0.20	0
158379	MC	0.73	0.31	0
158382	MC	0.86	0.25	0
187347	MC	0.60	0.19	0
187593	MC	0.67	0.39	0
187604	MC	0.68	0.27	0
228096	MC	0.77	0.30	0
228110	MC	0.64	0.27	0
228137	MC	0.86	0.29	0
228184	MC	0.70	0.40	1
228212	MC	0.30	0.19	0
228215	MC	0.53	0.38	0
250945	MC	0.53	0.23	0
250950	MC	0.30	0.25	0
250957	MC	0.66	0.18	0
250958	MC	0.59	0.41	0
250960	MC	0.61	0.32	0
250964	CR	0.48	0.53	1
250968	MC	0.84	0.32	0
250970	MC	0.72	0.35	0
250994	MC	0.59	0.22	0
251000	MC	0.89	0.33	0
251003	MC	0.56	0.40	0
251116	MC	0.68	0.33	0
251122	MC	0.33	0.16	0
251126	MC	0.38	0.24	0

Table I-3. 201–14 MEA Science: Item-Level Score Distributions for Constructed-Response Items—by Subject and Grade

Grade	*Item Number*	*Total Possible Points*	*Percent of Students at Score Point*				
			0	*1*	*2*	*3*	*4*
5	64956	4	5.21	14.76	56.74	20.58	2.42
	189933	4	13.95	11.53	56.43	13.01	3.06
	228101	4	14.03	38.98	31.74	11.81	2.75
	228143	4	30.66	29.50	19.28	12.99	6.63
8	63394	4	4.51	5.17	20.69	60.30	7.91
	96394	4	8.95	13.33	66.31	7.62	3.21
	100714	4	35.34	22.34	19.99	16.09	4.77
	250964	4	7.79	21.10	41.10	25.99	3.04

APPENDIX J—DIFFERENTIAL ITEM FUNCTIONING RESULTS

Table J-1. 2013–14 MEA Science: Number of Items Classified as "Low" or "High" DIF Overall and by Group Favored

Grade	Group		Item Type	Number of Items	Number "Low"			Number "High"		
	Reference	Focal			Total	Favoring		Total	Favoring	
						Reference	Focal		Reference	Focal
5	Male	Female	MC	32	5	5	0	1	1	0
			OR	4	1	0	1	0	0	0
	No Disability	Disability	MC	32	5	5	0	1	1	0
			OR	4	1	1	0	0	0	0
	Non-EconDis	EconDis	MC	32	1	1	0	0	0	0
			OR	4	0	0	0	0	0	0
	Non-LEP	LEP	MC	32	7	6	1	0	0	0
			OR	4	0	0	0	0	0	0
	White	Black	MC	32	8	7	1	1	1	0
			OR	4	0	0	0	0	0	0
		Hispanic	MC	32	5	3	2	0	0	0
			OR	4	0	0	0	0	0	0
8	Male	Female	MC	40	4	3	1	1	1	0
			OR	4	1	0	1	0	0	0
	No Disability	Disability	MC	40	3	3	0	0	0	0
			OR	4	0	0	0	1	1	0
	Non-EconDis	EconDis	MC	40	0	0	0	0	0	0
			OR	4	0	0	0	0	0	0
	Non-LEP	LEP	MC	40	6	6	0	0	0	0
			OR	4	1	1	0	0	0	0
	White	Asian	MC	40	4	2	2	0	0	0
		Asian	OR	4	1	1	0	0	0	0
		Black	MC	40	4	4	0	2	2	0
			OR	4	0	0	0	0	0	0
		Hispanic	MC	40	5	4	1	0	0	0
			OR	4	0	0	0	0	0	0

MC = Multiple-Choice; OR = Open Response.

APPENDIX K—ITEM RESPONSE THEORY PARAMETERS

Table K-1. 2013–14 MEA Science: IRT Parameters for Dichotomous Items—Science Grade 5

Item Number	a	S.E. (a)	b	S.E. (b)	c	S.E.(c)
96237	0.93438	0.04061	-0.12760	0.05793	0.35941	0.02213
64805	1.03983	0.04109	-0.97797	0.06954	0.30062	0.03602
158021	0.42799	0.02245	-0.56604	0.16343	0.13298	0.04747
228148	0.78652	0.03358	-0.17128	0.06697	0.25178	0.02630
96260	0.79725	0.03121	-0.35954	0.06486	0.19847	0.02795
96258	0.28837	0.01621	-1.86373	0.27998	0.13547	0.06002
190086	0.53062	0.03558	0.76804	0.09444	0.23866	0.02707
158023	0.49469	0.02183	-1.84511	0.17933	0.16257	0.06549
67897	0.44920	0.02177	-1.86120	0.21963	0.18298	0.07219
96338	0.76267	0.02979	-1.42614	0.10598	0.18123	0.05328
228136	0.36350	0.02432	0.08257	0.18992	0.13451	0.04626
248584	0.51485	0.03282	-0.13780	0.15418	0.27597	0.04291
228155	0.34779	0.02689	-0.92031	0.35773	0.25248	0.07783
190139	0.35782	0.02109	-3.43440	0.34342	0.22067	0.09140
64431	0.37577	0.01931	0.27879	0.11443	0.07376	0.02990
96261	0.80895	0.03439	0.40712	0.04632	0.21669	0.01778
67899	0.35844	0.02781	-0.41593	0.29046	0.22024	0.06509
228127	0.81871	0.03731	-0.31932	0.07650	0.33695	0.02845
97907	0.65571	0.04740	2.00757	0.05342	0.18115	0.01148
64371	0.63635	0.02536	-2.41691	0.12682	0.13083	0.05769
248572	0.53711	0.02624	-1.82591	0.21047	0.23168	0.07868
67903	0.66082	0.02680	-1.87752	0.13799	0.17533	0.06546
96368	0.44650	0.01786	-1.34826	0.13026	0.09695	0.04219
248520	0.34511	0.03162	1.16186	0.16933	0.14990	0.03988
190158	0.49108	0.02445	-1.73084	0.21860	0.21239	0.07553
96280	0.72328	0.03044	-0.25853	0.07345	0.19773	0.02922
66441	0.97161	0.03902	0.10587	0.04434	0.29538	0.01809
228159	0.41414	0.04064	1.63840	0.10928	0.21307	0.02876
96266	0.62474	0.03357	0.92323	0.05531	0.17111	0.01868
64790	0.66810	0.02519	-1.14371	0.09853	0.13463	0.04414
96577	0.77915	0.02702	-1.47482	0.08482	0.12446	0.04467
190321	0.54124	0.02675	-0.00779	0.09874	0.13962	0.03295

Table K-2. 2013–14 MEA Science: IRT Parameters for Polytomous Items—Science Grade 5

Item Number	a	S.E. (a)	b	S.E. (b)	D0	S.E. (D0)	D1	S.E. (D1)
189933	0.48938	0.00429	1.02266	0.02053	3.03885	0.02992	2.11853	0.02486
228101	0.64563	0.00561	1.21336	0.01494	2.91482	0.02407	0.79927	0.01757
228143	0.63630	0.00608	1.24001	0.01509	1.89698	0.01905	0.51947	0.01802
64956	0.56228	0.00478	0.41565	0.01803	3.54568	0.04105	1.85809	0.02416

Item Number	D2	S.E. (D2)	D3	S.E. (D3)	D4	S.E. (D4)
189933	-1.43631	0.02972	-3.72108	0.06149	0	0
228101	-0.96842	0.02403	-2.74567	0.04957	0	0
228143	-0.55456	0.02164	-1.86189	0.03372	0	0
64956	-1.28840	0.02312	-4.11537	0.06038	0	0

Table K-3. 2013–14 MEA Science: IRT Parameters for Dichotomous Items—Science Grade 8

Item Number	*a*	*S.E. (a)*	*b*	*S.E. (b)*	*c*	*S.E.(c)*
228110	0.66520	0.04309	0.41421	0.08706	0.36060	0.02495
250994	0.76090	0.05523	1.00420	0.06032	0.42072	0.01619
153806	0.66096	0.02689	-1.44642	0.11944	0.15484	0.05378
99750	0.66466	0.04372	1.18957	0.05207	0.22951	0.01687
250945	0.53351	0.04295	0.97029	0.09406	0.27659	0.02633
158382	0.53806	0.02724	-1.78922	0.21047	0.22727	0.07783
97218	0.95685	0.03634	-1.79243	0.08690	0.15689	0.05393
153649	0.42403	0.01894	-1.05949	0.13935	0.09774	0.04193
250970	0.99354	0.04722	0.09587	0.05192	0.39733	0.01871
187347	0.33354	0.03263	0.24652	0.28864	0.21224	0.05976
153680	0.87870	0.03390	-0.17885	0.05076	0.18781	0.02262
250968	0.87875	0.04482	-0.62081	0.09321	0.43614	0.03272
2610	0.56179	0.01992	-1.14440	0.08643	0.07744	0.03297
187593	0.67354	0.02117	-0.44536	0.04896	0.05167	0.01987
228137	0.59897	0.02208	-1.76937	0.10584	0.10311	0.04467
250958	1.08313	0.04101	0.35421	0.03053	0.21845	0.01361
158379	0.54521	0.02501	-0.76367	0.12148	0.13233	0.04362
250950	0.55971	0.03599	1.61329	0.04888	0.08295	0.01458
153664	0.99886	0.03197	-1.31765	0.05535	0.09951	0.03301
96450	0.84409	0.03674	0.70962	0.03497	0.15999	0.01392
250960	0.63488	0.03303	0.13808	0.08080	0.19539	0.02818
228184	0.86079	0.03280	-0.31523	0.05470	0.17551	0.02481
228096	0.57223	0.02719	-0.94820	0.13588	0.16868	0.05038
251000	0.81558	0.03403	-1.49763	0.11091	0.21364	0.05704
97791	0.60993	0.02230	-1.52371	0.10079	0.10277	0.04288
65206	0.98451	0.04470	0.45944	0.03932	0.30380	0.01509
250957	0.78894	0.06949	1.18087	0.06625	0.53590	0.01410
96454	1.16717	0.04387	-0.40882	0.04368	0.29212	0.02168
97788	0.62904	0.03358	0.28838	0.07648	0.19177	0.02638
158374	0.28455	0.01203	0.32369	0.03922	0.00000	0.00000
251126	0.78427	0.05016	1.48863	0.04068	0.21838	0.01178
228215	0.80218	0.03356	0.43059	0.04134	0.14869	0.01684
251122	0.61901	0.06095	2.17694	0.07773	0.22711	0.01333
96660	0.63673	0.03269	-0.44581	0.11096	0.23342	0.03933
251003	1.14841	0.04563	0.53921	0.02788	0.24759	0.01195
187604	0.48944	0.02767	-0.42102	0.14830	0.16135	0.04606
228212	1.22255	0.07454	1.72942	0.03197	0.21598	0.00660
99705	0.99920	0.04326	0.56295	0.03413	0.25518	0.01367
251116	0.60534	0.02984	-0.29381	0.10000	0.17009	0.03587
99758	0.46645	0.02905	-0.43928	0.17600	0.18853	0.05117

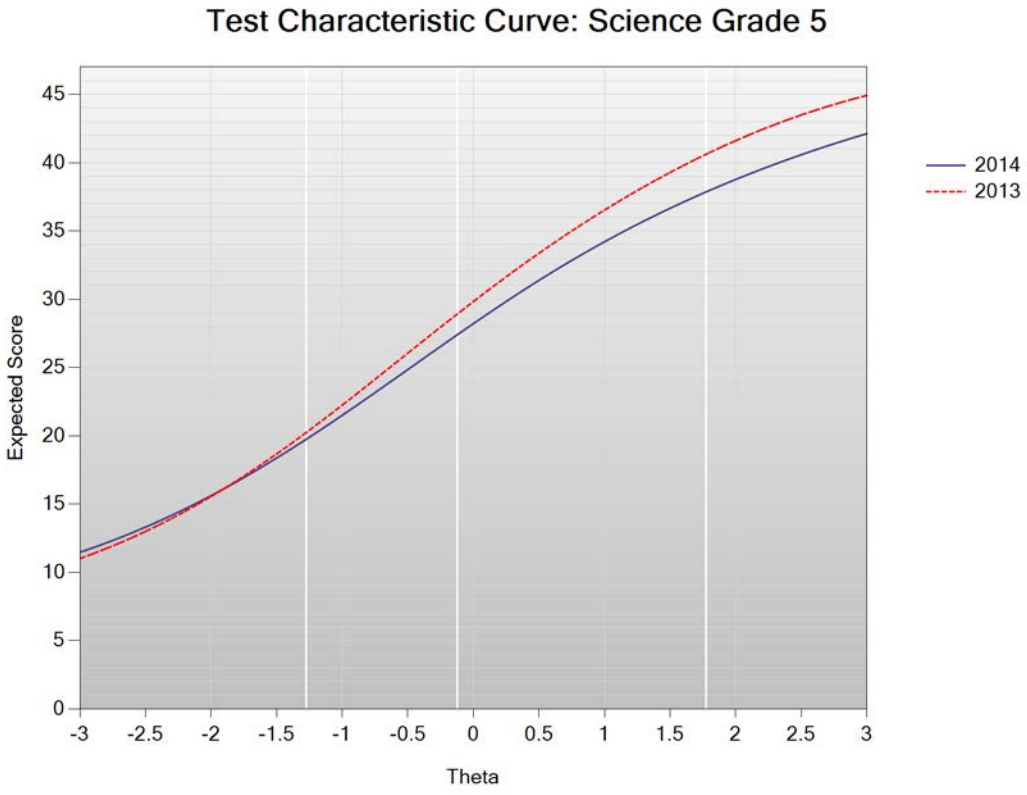
Table K-4. 2013–14 MEA Science: IRT Parameters for Polytomous Items—Science Grade 8

Item Number	*a*	*S.E. (a)*	*b*	*S.E. (b)*	*D0*	*S.E. (D0)*	*D1*	*S.E. (D1)*
100714	0.80050	0.00821	1.12386	0.01289	1.47470	0.01592	0.58604	0.01533
250964	0.79136	0.00718	0.38420	0.01288	2.46253	0.02673	1.05149	0.01684
63394	0.43251	0.00397	-0.97273	0.02514	2.92631	0.05415	1.90274	0.04045
96394	0.69636	0.00651	0.66923	0.01663	2.79057	0.02811	1.77691	0.02037

Item Number	*D2*	*S.E. (D2)*	*D3*	*S.E. (D3)*	*D4*	*S.E. (D4)*
100714	-0.32684	0.01802	-1.73391	0.03231	0	0
250964	-0.65903	0.01663	-2.85499	0.03976	0	0
63394	-0.05566	0.02787	-4.77339	0.04700	0	0
96394	-1.70198	0.02657	-2.86551	0.04491	0	0

APPENDIX L—TEST CHARACTERISTIC CURVES AND TEST INFORMATION FUNCTIONS

Figure L-1. 2013–14 MEA Science: Grade 5 Plots
Top: Test Characteristic Curve Bottom: Test Information Function



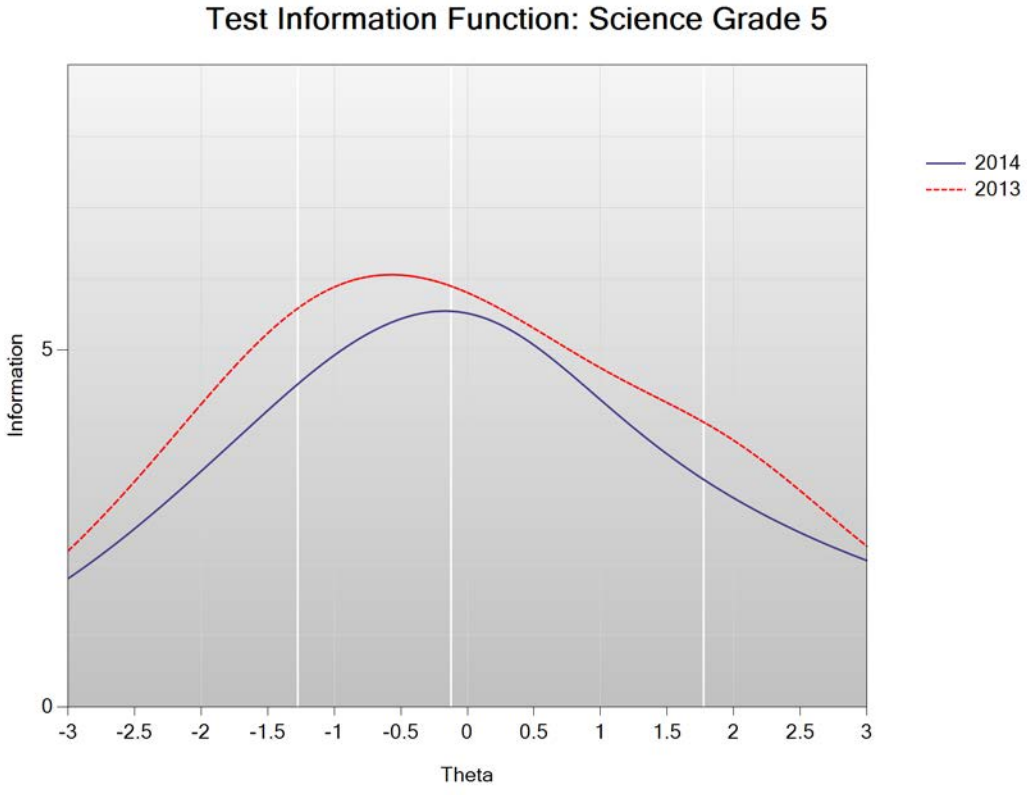
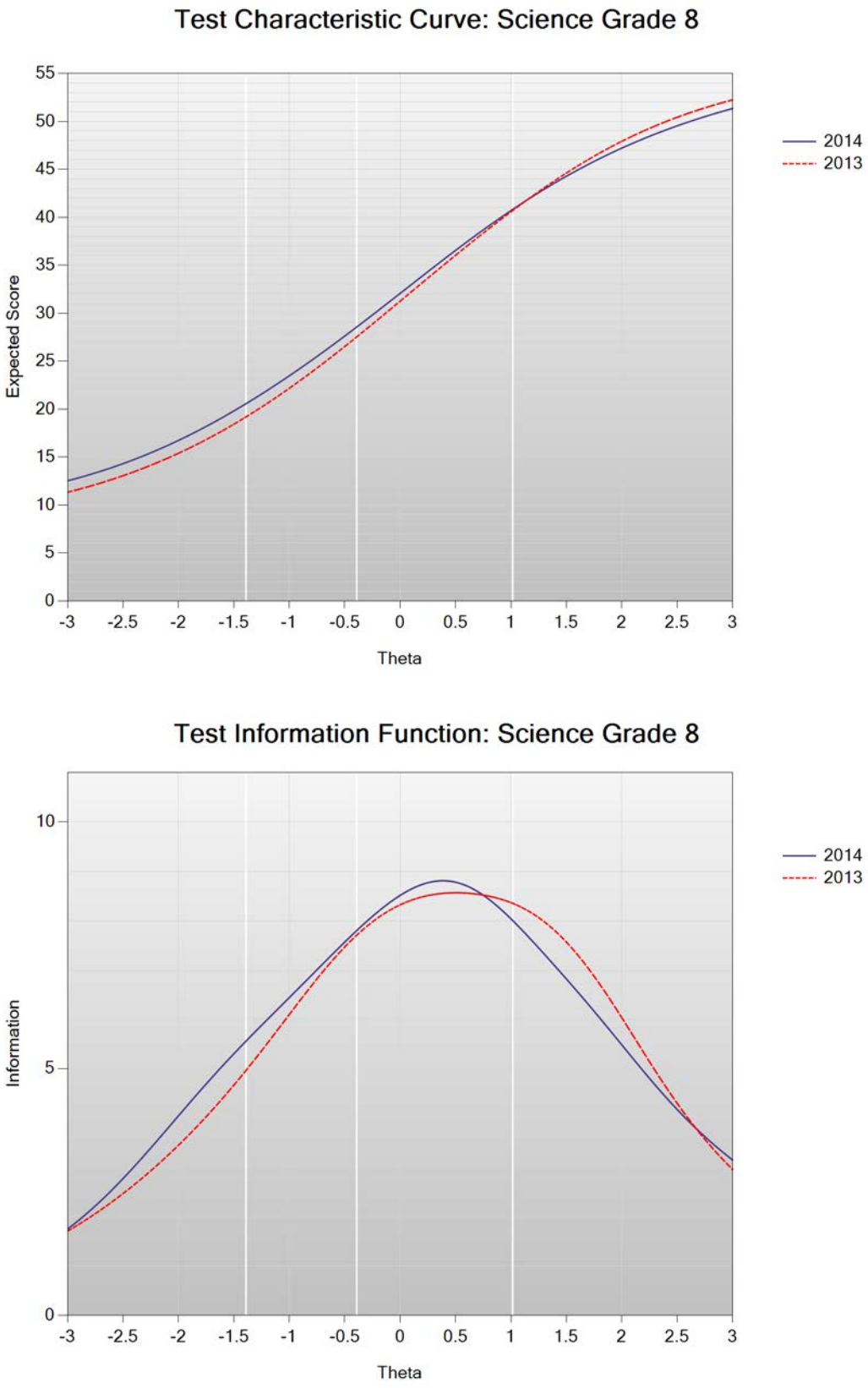


Figure L-2. 2013–14 MEA Science: Grade 8 Plots
Top: Test Characteristic Curve Bottom: Test Information Function



APPENDIX M—DELTA AND RESCORE ANALYSIS

Figure M-1. 2013–14 MEA Science: Delta Analysis Plots—Science
Top: Grade 5 Bottom: Grade 8

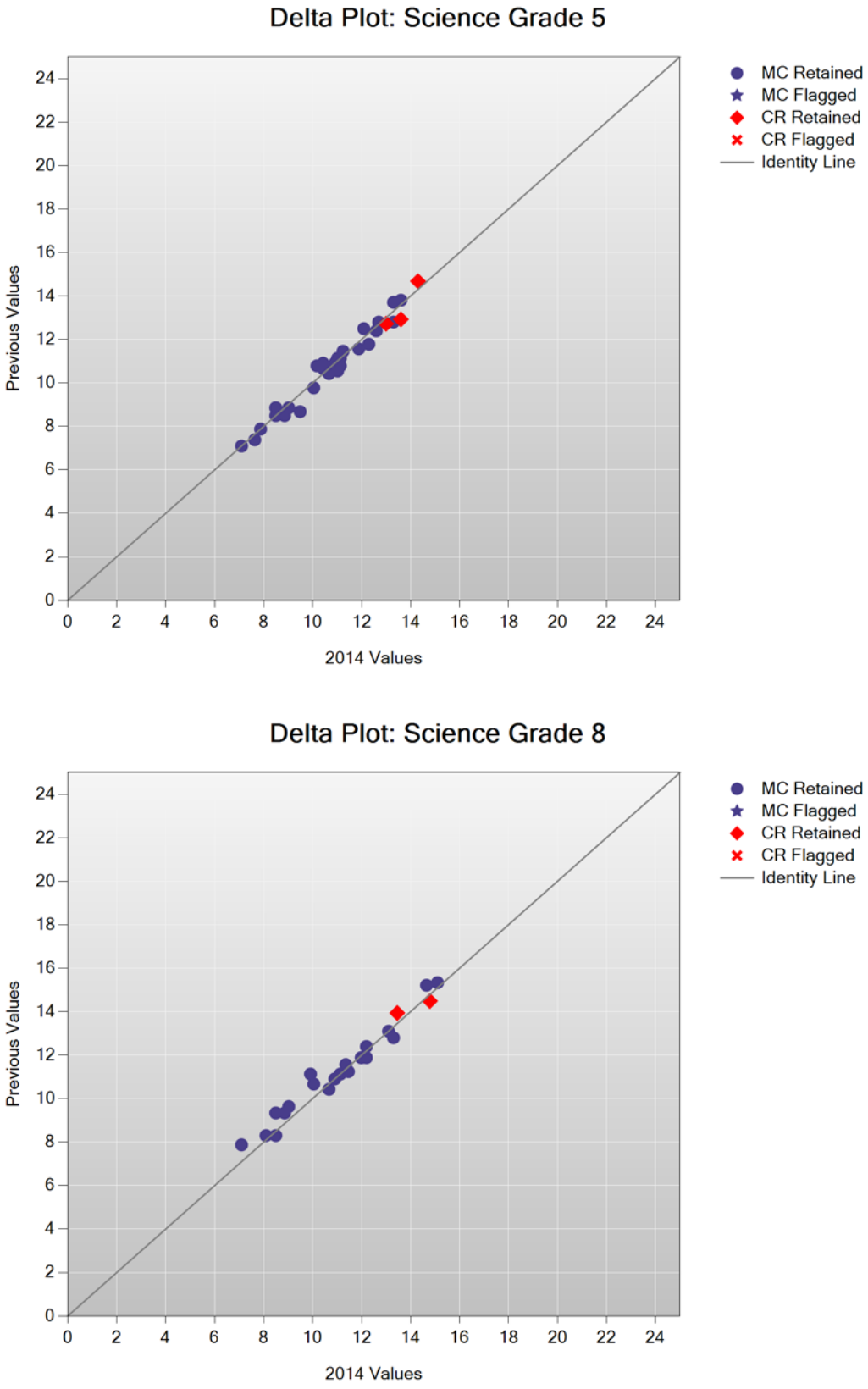


Table M-1. 2013–14 MEA Science: Delta Analysis Results—Science Grade 5

Item Number	Mean		Delta		Discard	Standardized Difference
	Old	New	Old	New		
158021	0.68000	0.68000	11.12920	11.12920	False	-1.35528
189933	0.50750	0.44000	12.92480	13.60388	False	1.83396
190086	0.56000	0.54000	12.39612	12.59827	False	-0.71263
190139	0.92000	0.91000	7.37971	7.63698	False	-0.67813
190158	0.87000	0.85000	8.49444	8.85427	False	-0.07926
190321	0.64000	0.61000	11.56616	11.88272	False	-0.15067
228101	0.33750	0.37250	14.67718	14.30096	False	0.45067
228136	0.55000	0.59000	12.49735	12.08982	False	0.72761
228148	0.70000	0.70000	10.90240	10.90240	False	-1.34367
228155	0.72000	0.74000	10.66863	10.42662	False	-0.05320
228159	0.42000	0.44000	13.80757	13.60388	False	-0.41625
248520	0.52000	0.47000	12.79939	13.30108	False	0.89046
248572	0.85000	0.87000	8.85427	8.49444	False	0.66203
248584	0.68000	0.69000	11.12920	11.01660	False	-0.76041
64371	0.93000	0.93000	7.09684	7.09684	False	-1.14893
64431	0.52000	0.53000	12.79939	12.69892	False	-0.91001
64790	0.86000	0.81000	8.67872	9.48841	False	2.30667
64805	0.85000	0.87000	8.85427	8.49444	False	0.66203
64956	0.52750	0.50000	12.72405	13.00000	False	-0.30596
66441	0.65000	0.67000	11.45872	11.24035	False	-0.21854
67897	0.85000	0.84000	8.85427	9.02217	False	-1.07476
67899	0.71000	0.68000	10.78646	11.12920	False	-0.05224
67903	0.90000	0.90000	7.87379	7.87379	False	-1.18869
96237	0.70000	0.74000	10.90240	10.42662	False	1.16975
96258	0.71000	0.76000	10.78646	10.17479	False	1.89357
96260	0.74000	0.72000	10.42662	10.66863	False	-0.60278
96261	0.62000	0.57000	11.77808	12.29450	False	0.91602
96266	0.43000	0.47000	13.70550	13.30108	False	0.64932
96280	0.73000	0.69000	10.54875	11.01660	False	0.59650
96338	0.87000	0.87000	8.49444	8.49444	False	-1.22045
96368	0.79000	0.77000	9.77432	10.04461	False	-0.48675

Table M-2. 2013–14 MEA Science: Delta Analysis Results—Science Grade 8

Item Number	Mean		Delta		Discard	Standardized Difference
	Old	New	Old	New		
100714	0.35500	0.32750	14.48742	14.78731	False	-0.31616
153664	0.88000	0.87000	8.30005	8.49444	False	1.06418
153680	0.68000	0.68000	11.12920	11.12920	False	-0.66569
153806	0.82000	0.85000	9.33854	8.85427	False	-0.95651
158374	0.49000	0.49000	13.10028	13.10028	False	-1.25762
158382	0.82000	0.87000	9.33854	8.49444	False	0.67299
187347	0.61000	0.60000	11.88272	11.98661	False	-0.42152
228110	0.67000	0.65000	11.24035	11.45872	False	0.28982
228184	0.70000	0.70000	10.90240	10.90240	False	-0.59758

continued

Item Number	Mean		Delta		Discard	Standardized Difference
	Old	New	Old	New		
250950	0.28000	0.30000	15.33137	15.09760	False	-0.29126
250957	0.64000	0.66000	11.56616	11.35015	False	-1.50234
250958	0.61000	0.58000	11.88272	12.19243	False	0.51051
250970	0.74000	0.72000	10.42662	10.66863	False	0.64127
251000	0.88000	0.89000	8.30005	8.09389	False	-0.74970
251122	0.29000	0.34000	15.21354	14.64985	False	1.16741
2610	0.72000	0.77000	10.66863	10.04461	False	0.07577
65206	0.61000	0.60000	11.88272	11.98661	False	-0.42152
96394	0.40750	0.45500	13.93592	13.45215	False	0.42182
96450	0.52000	0.47000	12.79939	13.30108	False	1.10466
96454	0.68000	0.78000	11.12920	9.91123	False	2.90382
97218	0.90000	0.93000	7.87379	7.09684	False	-0.07096
97788	0.56000	0.58000	12.39612	12.19243	False	-1.30889
97791	0.80000	0.84000	9.63352	9.02217	False	-0.29248

Table M-3. 2013–14 MEA Science: Rescore Analysis Results by Subject and Grade

Grade	Item Number	Maximum	Mean		Standard Deviation		Effect Size	Discard
			Old	New	Old	New		
5	228143	4	1.39702	1.44665	1.26420	1.24349	0.03926	False
	228101	4	1.36318	1.61194	1.04866	1.07507	0.23721	False
	189933	4	2.09756	1.78049	0.90651	0.90400	-0.34977	False
	64956	4	2.16749	2.04926	0.77078	0.73538	-0.15339	False
8	250964	4	2.10462	1.86618	1.05519	0.96226	-0.22597	False
	63394	4	3.05854	2.65366	0.90448	0.82861	-0.44764	False
	100714	4	1.31373	1.12255	1.34826	1.23012	-0.14180	False
	96394	4	1.66667	1.53922	0.92291	0.86620	-0.13810	False

APPENDIX N—α-PLOTS AND b-PLOTS

Figure N-1. 2013–14 MEA Science: Grade 5 Plots
Top: *α*-Plot Bottom: *b*-Plot

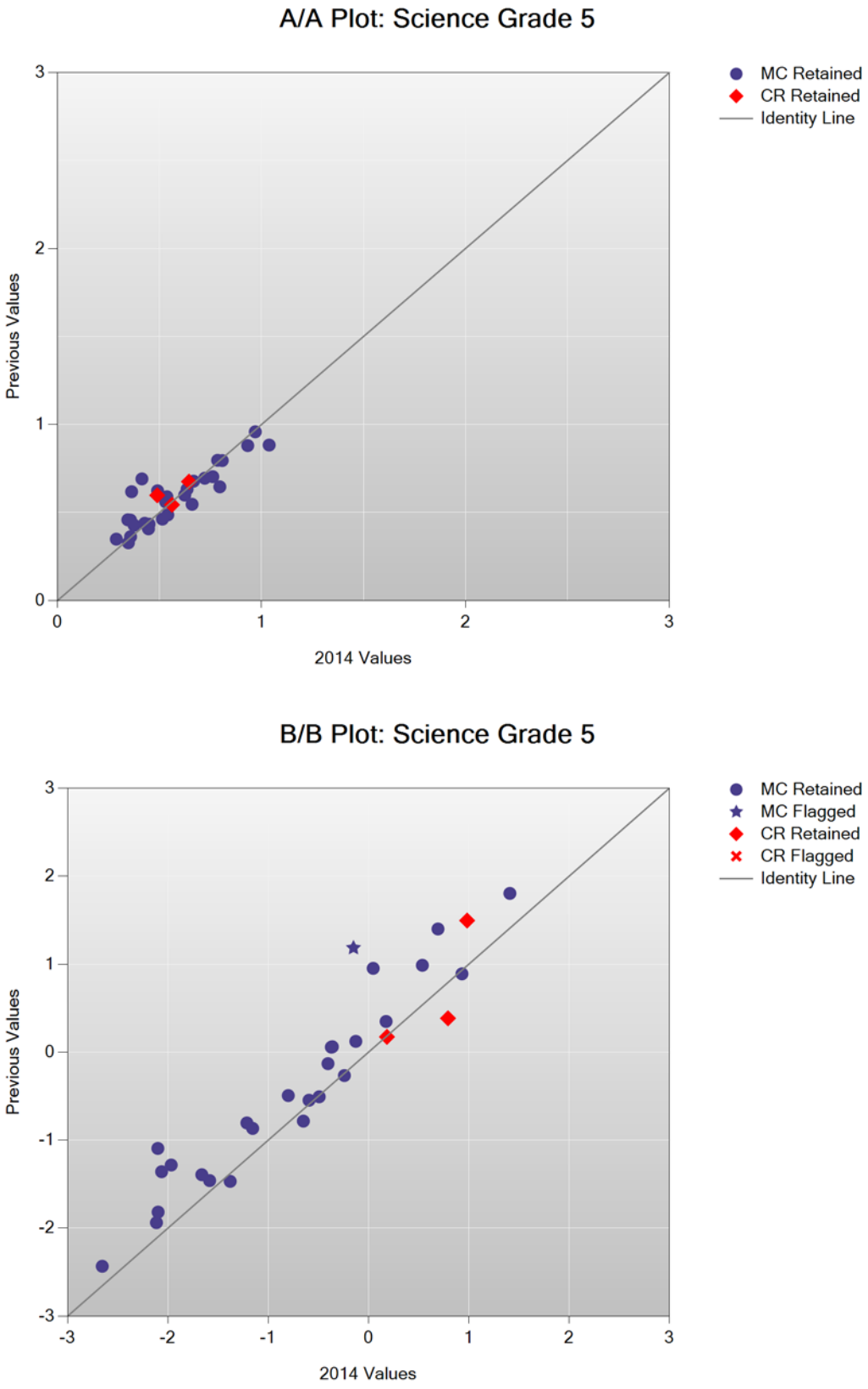
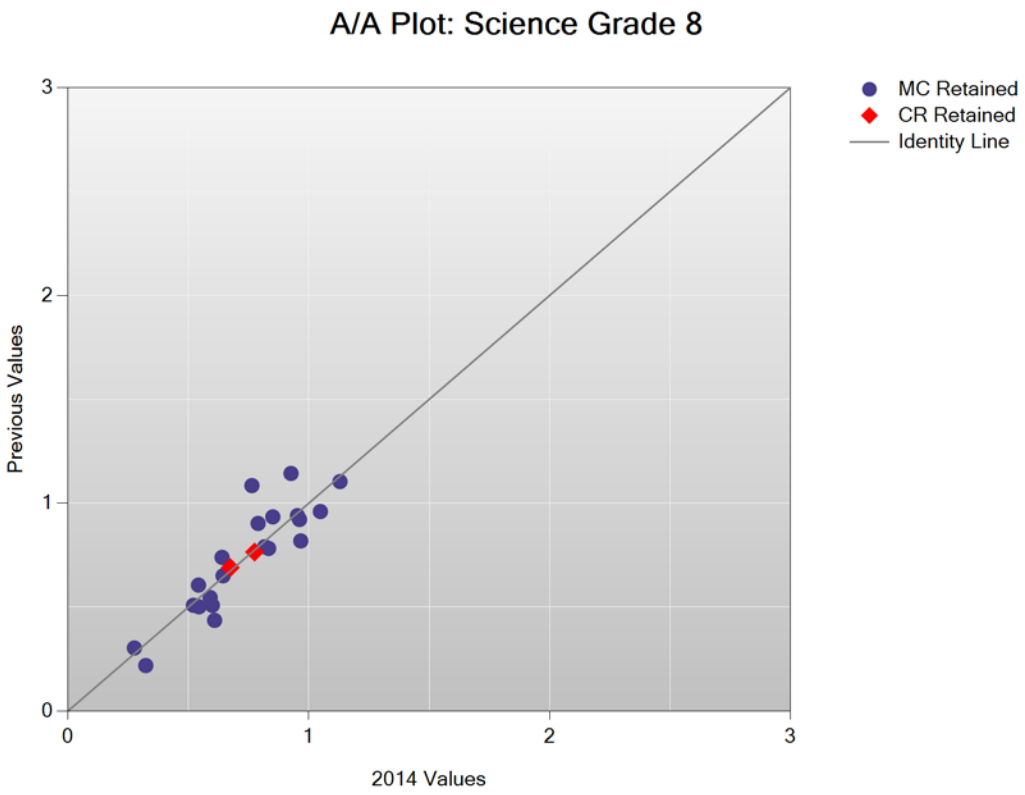
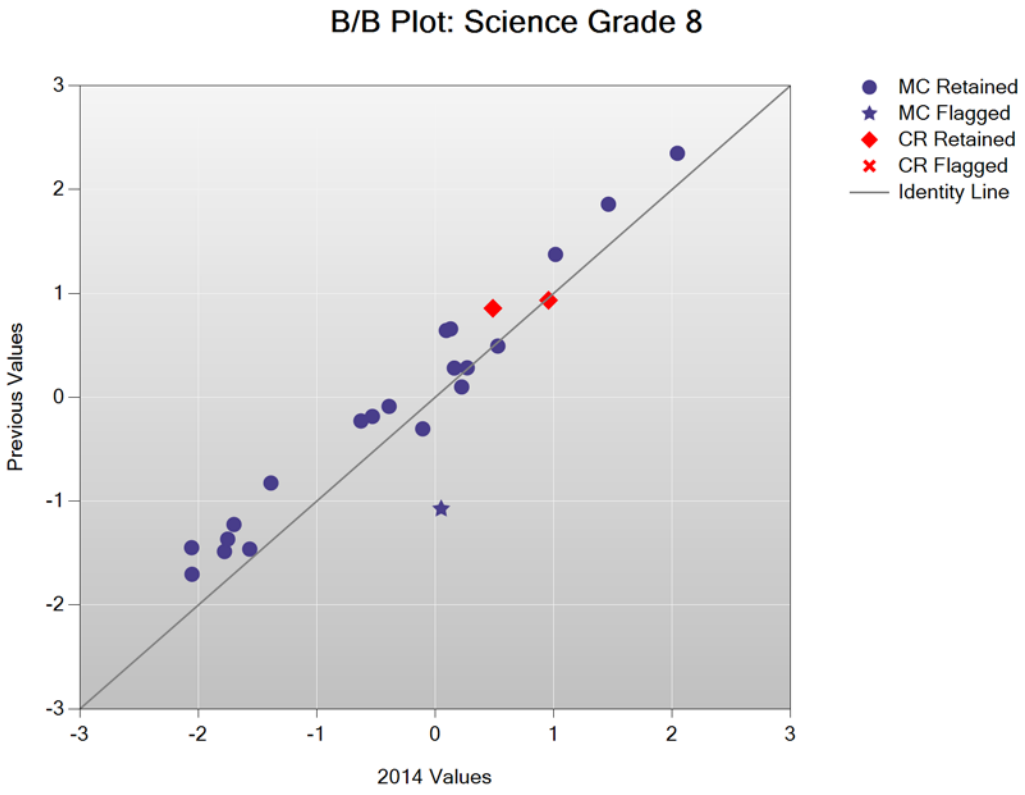


Figure N-2. 2013–14 MEA Science: Grade 8 Plots
Top: *α*-Plot Bottom: *b*-Plot





APPENDIX O—RAW TO SCALED SCORE LOOKUP TABLES

Table O-1. 2013–14 MEA Science: Raw to Scaled Score Look-up Table—Science Grade 5

Raw Score	This Year			Last Year		
	Scaled Score	Standard Error	Performance Level	Scaled Score	Standard Error	Performance Level
0	500	10.0	1	500	10.0	1
1	500	10.0	1	500	10.0	1
2	500	10.0	1	500	10.0	1
3	500	10.0	1	500	10.0	1
4	500	10.0	1	500	10.0	1
5	500	10.0	1	500	10.0	1
6	500	10.0	1	500	10.0	1
7	500	10.0	1	500	10.0	1
8	500	10.0	1	500	10.0	1
9	502	10.0	1	504	10.0	1
10	506	10.0	1	508	8.0	1
11	510	8.0	1	512	8.0	1
12	514	8.0	1	514	6.0	1
13	516	6.0	1	516	6.0	1
14	518	6.0	1	520	6.0	1
15	520	6.0	1	522	6.0	1
16	524	6.0	1	524	6.0	1
17	524	6.0	1	524	4.0	1
18	526	6.0	1	526	4.0	1
19	528	6.0	1	528	4.0	1
20	530	4.0	2	528	4.0	1
21	532	4.0	2	530	4.0	2
22	534	4.0	2	532	4.0	2
23	536	4.0	2	534	4.0	2
24	536	4.0	2	536	4.0	2
25	538	4.0	2	536	4.0	2
26	540	4.0	2	538	4.0	2
27	540	4.0	2	540	4.0	2
28	542	4.0	3	540	4.0	2
29	544	4.0	3	542	4.0	3
30	546	4.0	3	544	4.0	3
31	548	4.0	3	544	4.0	3
32	550	4.0	3	546	4.0	3
33	552	4.0	3	548	4.0	3
34	554	6.0	3	550	4.0	3
35	556	6.0	3	552	4.0	3
36	558	6.0	3	552	4.0	3
37	560	6.0	3	554	4.0	3
38	562	6.0	4	556	4.0	3
39	566	6.0	4	558	6.0	3
40	568	6.0	4	560	6.0	3
41	570	6.0	4	562	6.0	4
42	574	8.0	4	566	6.0	4
43	578	8.0	4	568	6.0	4
44	580	8.0	4	572	6.0	4

continued

Raw Score	This Year			Last Year		
	Scaled Score	Standard Error	Performance Level	Scaled Score	Standard Error	Performance Level
45	580	8.0	4	576	8.0	4
46	580	8.0	4	580	8.0	4
47	580	8.0	4	580	10.0	4
48	580	8.0	4	580	10.0	4

Table O-2. 2013–14 MEA Science: Raw to Scaled Score Look-up Table—Science Grade 8

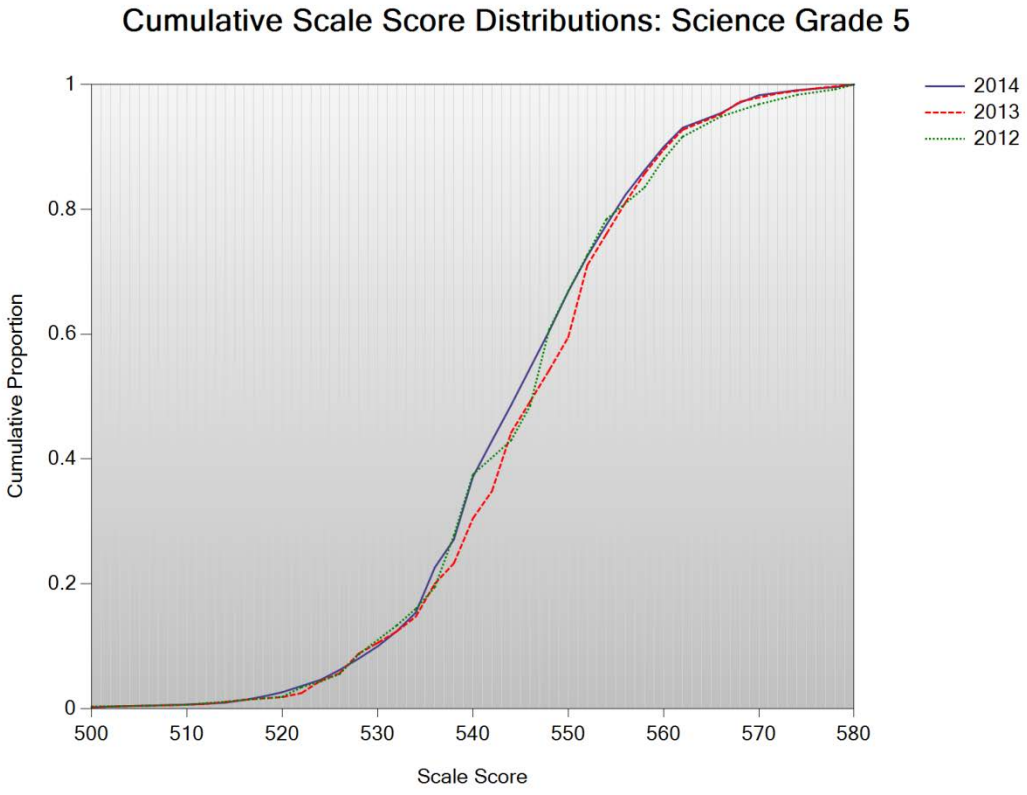
Raw Score	This Year			Last Year		
	Scaled Score	Standard Error	Performance Level	Scaled Score	Standard Error	Performance Level
0	800	10.0	1	800	10.0	1
1	800	10.0	1	800	10.0	1
2	800	10.0	1	800	10.0	1
3	800	10.0	1	800	10.0	1
4	800	10.0	1	800	10.0	1
5	800	10.0	1	800	10.0	1
6	800	10.0	1	800	10.0	1
7	800	10.0	1	800	10.0	1
8	800	10.0	1	800	10.0	1
9	800	10.0	1	800	10.0	1
10	800	10.0	1	800	10.0	1
11	800	10.0	1	804	10.0	1
12	802	10.0	1	808	10.0	1
13	808	10.0	1	812	10.0	1
14	810	8.0	1	816	8.0	1
15	814	8.0	1	818	8.0	1
16	818	8.0	1	820	8.0	1
17	820	6.0	1	824	8.0	1
18	822	6.0	1	826	6.0	1
19	824	6.0	1	826	6.0	1
20	826	6.0	1	830	6.0	2
21	828	6.0	2	832	6.0	2
22	830	6.0	2	834	6.0	2
23	832	6.0	2	834	6.0	2
24	834	6.0	2	836	6.0	2
25	836	6.0	2	838	6.0	2
26	838	6.0	2	840	6.0	2
27	840	6.0	2	840	6.0	2
28	840	6.0	2	842	6.0	3
29	842	6.0	3	844	6.0	3
30	844	4.0	3	846	4.0	3
31	846	4.0	3	848	4.0	3
32	848	4.0	3	848	4.0	3
33	850	4.0	3	850	4.0	3
34	850	4.0	3	852	4.0	3
35	852	4.0	3	854	4.0	3
36	854	4.0	3	854	4.0	3

continued

Raw Score	This Year			Last Year		
	Scaled Score	Standard Error	Performance Level	Scaled Score	Standard Error	Performance Level
37	856	4.0	3	856	4.0	3
38	858	4.0	3	858	4.0	3
39	858	4.0	3	860	4.0	3
40	860	4.0	3	860	4.0	3
41	862	6.0	4	862	4.0	4
42	864	6.0	4	864	4.0	4
43	866	6.0	4	866	6.0	4
44	868	6.0	4	868	6.0	4
45	870	6.0	4	870	6.0	4
46	872	6.0	4	872	6.0	4
47	876	6.0	4	874	6.0	4
48	878	6.0	4	876	6.0	4
49	880	6.0	4	878	6.0	4
50	880	8.0	4	880	6.0	4
51	880	8.0	4	880	8.0	4
52	880	8.0	4	880	8.0	4
53	880	10.0	4	880	10.0	4
54	880	10.0	4	880	10.0	4
55	880	10.0	4	880	10.0	4
56	880	10.0	4	880	10.0	4

APPENDIX P—SCORE DISTRIBUTIONS

Figure P-1. 2013–14 MEA Science: Score Distribution Plots
Top: Grade 5 Bottom: Grade 8



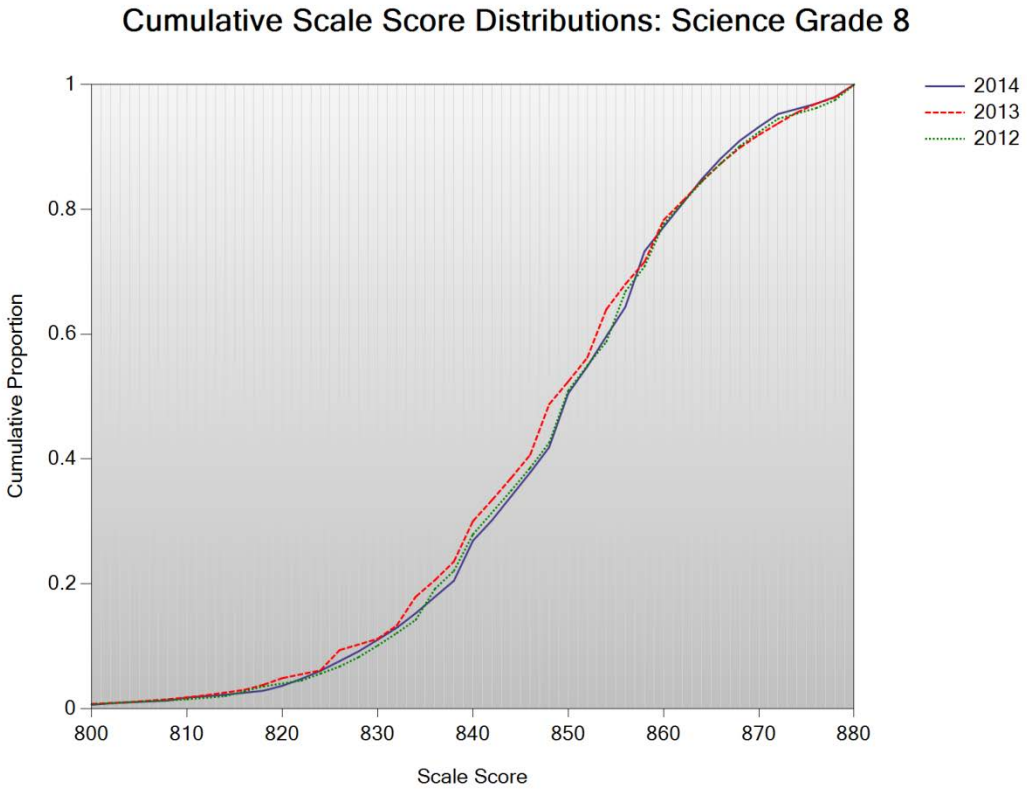


Table P-1. MEA Science: Achievement Level Distributions by Subject and Grade

Grade	*Performance Level*	*2013-14*	*2012-13*	*2011-12*
5	4	10.02	10.47	11.94
	3	52.84	59.07	50.56
	2	29.14	21.68	28.85
	1	7.99	8.78	8.65
8	4	22.85	21.73	22.36
	3	50.28	48.24	49.75
	2	19.30	20.68	21.18
	1	7.58	9.34	6.71

APPENDIX Q—RELIABILITY

Table Q-1. 2013–14 MEA Science: Subgroup Reliabilities

Grade	Group	Number of Students	Raw Score			Alpha	SEM
			Maximum	Mean	Standard Deviation		
5	All Students	12979	48	29.30	6.56	0.79	2.99
	Male	6653	48	29.23	6.59	0.80	2.97
	Female	6325	48	29.36	6.52	0.79	3.01
	Gender not reported	1	48				
	Hispanic or Latino	244	48	27.64	6.33	0.77	3.01
	American Indian or Alaskan Native	102	48	26.69	5.74	0.73	2.99
	Asian	189	48	30.72	6.67	0.8	2.99
	Black or African American	391	48	23.74	7.17	0.81	3.12
	Native Hawaiian or Pacific Islander	13	48	31.00	5.66	0.71	3.04
	White (non-Hispanic)	11845	48	29.52	6.44	0.79	2.98
	Two or More Races (non-Hispanic)	194	48	28.83	7.15	0.82	3.04
	Race not reported	1	48				
	Currently receiving LEP services	334	48	22.09	7.05	0.81	3.09
	Former LEP student – monitoring year 1	53	48	28.60	6.17	0.73	3.20
	Former LEP student – monitoring year 2	43	48	30.33	6.07	0.77	2.90
	LEP: All Other Students	12549	48	29.49	6.44	0.78	2.99
	Students with an IEP	2210	48	24.17	6.49	0.78	3.04
	IEP: All Other Students	10769	48	30.35	6.06	0.76	2.97
	Economically Disadvantaged Students	5943	48	27.33	6.50	0.78	3.02
	SES: All Other Students	7036	48	30.96	6.13	0.77	2.95
	Migrant Students	6	48				
	Migrant: All Other Students	12973	48	29.30	6.56	0.79	2.99
	Students receiving Title 1 services	2047	48	26.51	5.96	0.74	3.05
	Title 1: All Other Students	10932	48	29.82	6.53	0.79	2.98
	Plan 504:	449	48	29.87	5.90	0.75	2.94
	Plan 504: All Other Students	12530	48	29.28	6.58	0.79	3.00
8	All Students	13473	56	33.65	8.45	0.85	3.27
	Male	6893	56	33.79	8.79	0.86	3.28
	Female	6580	56	33.51	8.08	0.84	3.26
	Gender not reported	0	56				
	Hispanic or Latino	214	56	31.13	8.21	0.83	3.36
	American Indian or Alaskan Native	98	56	32.67	8.11	0.84	3.23
	Asian	204	56	35.60	8.81	0.87	3.21
	Black or African American	397	56	27.07	8.61	0.84	3.43
	Native Hawaiian or Pacific Islander	16	56	36.13	5.11	0.62	3.17
	White (non-Hispanic)	12347	56	33.90	8.35	0.85	3.27
	Two or More Races (non-Hispanic)	197	56	32.56	8.50	0.85	3.29
	Race not reported	0	56				
	Currently receiving LEP services	399	56	25.89	7.93	0.81	3.45
	Former LEP student – monitoring year 1	20	56	38.00	4.68	0.59	3.01
	Former LEP student – monitoring year 2	9	56				
	LEP: All Other Students	13045	56	33.88	8.36	0.85	3.27
	Students with an IEP	2170	56	25.76	8.13	0.82	3.44
	IEP: All Other Students	11303	56	35.17	7.63	0.82	3.21
	Economically Disadvantaged Students	5900	56	30.76	8.35	0.84	3.35
	SES: All Other Students	7573	56	35.91	7.81	0.83	3.19

continued

Grade	Group	Number of Students	Raw Score			Alpha	SEM
			Maximum	Mean	Standard Deviation		
8	Migrant Students	2	56				
	Migrant: All Other Students	13471	56	33.65	8.45	0.85	3.27
	Students receiving Title 1 services	937	56	30.01	7.73	0.81	3.38
	Title 1: All Other Students	12536	56	33.93	8.44	0.85	3.26
	Plan 504:	655	56	33.98	7.71	0.82	3.27
	Plan 504: All Other Students	12818	56	33.64	8.48	0.85	3.28

Table Q-2. 2013–14 MEA Science: Reliabilities by Reporting Category

Grade	Item Reporting Category	Number of Items	Raw Score			Alpha	SEM
			Maximum	Mean	Standard Deviation		
5	Physical Setting	18	24	15.48	3.33	0.64	1.99
	Living Environment	18	24	13.82	3.88	0.67	2.23
8	Physical Setting	28	34	19.58	5.52	0.77	2.65
	Earth/Space	14	17	10.96	2.93	0.64	1.77
	Matter/Energy/Force/Motion	14	17	8.62	3.18	0.61	1.98
	Living Environment	16	22	14.08	3.55	0.71	1.92

APPENDIX R—DECISION ACCURACY AND CONSISTENCY

Table R-1. 2013–14 MEA Science: DAC

Grade	Overall	Kappa	Conditional on Level			
			Substantially Below Proficient	Partially Proficient	Proficient	Proficient with Distinction
5	0.74 (0.64)	0.42	0.74 (0.53)	0.68 (0.58)	0.77 (0.71)	0.75 (0.54)
8	0.76 (0.67)	0.5	0.77 (0.59)	0.65 (0.53)	0.78 (0.73)	0.82 (0.70)

Table R-2. 2013–14 MEA Science: DAC

Grade	Substantially Below Proficient / Partially Proficient			Partially Proficient / Proficient			Proficient / Proficient with Distinction		
	Accuracy (consistency)	False		Accuracy (consistency)	False		Accuracy (consistency)	False	
		Positive	Negative		Positive	Negative		Positive	Negative
5	0.95 (0.93)	0.01	0.04	0.86 (0.80)	0.06	0.08	0.93 (0.90)	0.05	0.02
8	0.96 (0.94)	0.01	0.03	0.90 (0.86)	0.04	0.06	0.90 (0.87)	0.06	0.04

APPENDIX S—INTERRATER AGREEMENT

Table S-1. 2013–14 MEA Science: Item-Level Interrater Consistency Statistics by Grade

Grade	Item Number	Number of		Percent		Correlation	Percent of Third Scores
		Score Categories	Responses Scored Twice	Exact	Adjacent		
5	189933	5	1362	72.83	24.01	0.79	2.94
	228101	5	1411	55.71	39.97	0.68	3.19
	228143	5	1410	70.00	27.94	0.88	1.99
	64956	5	1327	73.25	24.79	0.74	1.66
8	100714	5	1349	73.91	21.94	0.87	3.78
	250964	5	1328	59.04	38.48	0.72	1.96
	63394	5	1316	79.33	18.92	0.83	1.67
	96394	5	1323	81.56	16.70	0.82	1.59