Colorado Alternate Assessment:
Science and Social Studies

Technical Report

Spring
2015

Table of Contents

PART I: HISTORICAL OVERVIEW AND SUMMARY OF PROCESSES

CHAPTER 1: INTRODUCTION AND BACKGROUND

All public school students enrolled in Colorado are required by state law to take a standards-based assessment each year in specified content areas and grade levels. Every student, regardless of language background or academic ability, must be provided with the opportunity to demonstrate their content knowledge of the Colorado Academic Standards (CAS). The CAS were adopted by the State in science and social studies in December of 2009 and outline the concepts and skills that students need in order to be successful in the current grade as well as to make academic progress from year to year.

In partnership with Colorado educators and Pearson, Inc., the Colorado Department of Education (CDE) developed a new assessment, the Colorado Alternate Assessment (CoAlt): Science and Social Studies assessments, to evaluate student mastery of the CAS in science and social studies for students with the most significant cognitive disabilities. For students who qualify, these assessments provide an indicator of student progress toward the Extended Evidence Outcomes (EEOs) of the CAS in the content areas of science and social studies.

Purpose of the Document

The purpose of the CoAlt: Science and Social Studies Technical Report is to inform users and other interested parties about the technical characteristics of this assessment program. This technical report provides information about the Fall 2014 and the Spring 2015 CoAlt: Science and Social Studies assessments, including content, assessment development, administration, scoring, and technical attributes.

The Spring 2015 CoAlt: Science and Social Studies Technical Report is divided into two parts. Part I presents an overview and summary of the components of the program. Information regarding the planning and administration of the assessments as well as details regarding item development, item banking, test construction, administration procedures, scoring, reporting, reliability, and validity are included in Part I of the document. Part II provides a statistical summary of the Fall 2014 and the Spring 2015 administrations, including results for both the operational items and the embedded field test items.

Overview of CoAlt

Purposes of the CoAlt Assessment Program

The Individuals with Disabilities Education Improvement Act of 2004 (IDEA) mandates that all students have access to the general curriculum and be included in each state's accountability system. The Reauthorization of the Elementary and Secondary Education Act (ESEA) of 2001 (also known as No Child Left Behind) specifies that states must provide an alternate assessment when implementing statewide accountability systems to help ensure the inclusion of all students

in a state's accountability system. To ensure the participation of all students with the most significant cognitive disabilities in the Colorado accountability system in the content areas of science and social studies, Colorado developed the CoAlt: Science and Social Studies assessments.

The goals of the Colorado Assessment System, including the CoAlt: Science and Social Studies assessments, are to measure and support student progress toward the content standards; provide students, parents, and other stakeholders with information regarding student achievement; and gauge the quality and efficiency of educational programs in public schools.

In addition to the goals noted above, CoAlt promotes improved instruction toward grade-level expectations, growth over time toward independent performance, and high expectations toward achievement in the content areas.

The Student Population

The CoAlt: Science and Social Studies assessments are designed for students with the most significant cognitive disabilities. These students are defined by having significant limitations in cognitive functioning and deficits in adaptive behavior. They also may exhibit limitations in communication, methods of response, sustaining attention, and short-term memory. A very small number of students with the most significant cognitive disabilities who cannot participate in the state summative assessment, the Colorado Measures of Academic Success (CMAS), even with accommodations may take CoAlt. These students are often identified as having an Intellectual Disability; however, students with other disability categories may also meet the participation criteria for CoAlt.

Participation in the CoAlt: Science and Social Studies assessments is determined by a student's Individualized Education Program (IEP) team. The IEP team will determine whether a student should participate in CoAlt or CMAS by determining if the student meets the criteria in the Alternate Academic Achievement Standards and Alternate Assessment Participation Guidelines Worksheet. The IEP team can decide that CoAlt is the most appropriate assessment for the student if the student meets all of the following participation criteria:

1. The student has been evaluated and determined to be eligible to receive special education services and has an IEP.
2. The student has documented evidence of a cognitive disability.
3. The student has a significant cognitive disability.
4. The student is receiving daily instruction based on the EEOs (alternate achievement standards).

The CoAlt eligibility guidelines can be found in Appendix A and are also available on the Exceptional Student Services Unit website at the following location:
http://www.cde.state.co.us/sites/default/files/accommodationsmanual_eligibility.pdf

Description of CoAlt: Science and Social Studies

CoAlt is a standards-based assessment designed specifically for students with the most significant cognitive disabilities. The primary purpose of the assessment program is to determine the level at which Colorado students with significant cognitive disabilities meet the EEOs of the CAS in the content areas of science and social studies. The EEOs are alternate academic standards that describe what students taking CoAlt are expected to know and be able to demonstrate at each grade level and in each content area.

The test design of the CoAlt: Science and Social Studies was developed to provide this unique population of students with an opportunity to demonstrate their knowledge of the EEOs. The CoAlt: Science and Social Studies assessments include paper-based test books used by the Test Examiner to administer test items to the students. The test books are oriented so that the Test Examiner administers the test while facing the student. The test book includes scripted text for the Test Examiner to read test questions and answer choices to the student. There is flexibility for presentation and response based on the student's mode of communication; however, the script and order in which the answer options are presented to the student must remain the same. During the course of the administration, the Test Examiner scores each item and records student performance within the test book or on the score recording form included with the test materials. At the conclusion of the administration, the Test Examiner enters the student's scores into PearsonAccess, an online score entry system.

Two item types are included as part of the CoAlt: Science and Social Studies assessments: selected response (SR) items and supported performance task (SPT) items. SR items have three answer options from which the student selects an answer to the question presented. The student works with the item until he or she provides the correct answer or the maximum number of attempts is reached. Teachers score the student's performance using a four-point scoring rubric that is built into the item.

SPT items consist of three related questions. Teachers are provided with specific prompts and the students respond to each prompt using a set of option cards. Students manipulate the option cards by placing them on a designated response page (e.g., placing option cards in designated boxes within a chart or diagram). Teachers score the student's performance on each of the three prompts using a two-point scoring rubric that is built into the item. The points for the three prompts are then added together to provide one score for the SPT item.

Field test items are embedded in the operational forms. Including field test items on the operational test forms reduces the need for future stand-alone field tests and allows newly-developed test items to be field tested with a relatively large participation count.

The CoAlt: Science and Social Studies High School (HS) assessments were administered for the first time in Fall 2014. The following CoAlt: Science and Social Studies Elementary and Middle School (ES/MS) assessments were administered during Spring 2015:

- Social studies: grades 4 and 7
- Science: grades 5 and 8

The Standards

A key element in ESEA is that alternate assessments must be aligned with the content standards for the grade level in which the student is enrolled. On August 3, 2011, the State Board of Education adopted the EEOs for students with the most significant cognitive disabilities who qualify for an alternate assessment. The EEOs are alternate academic standards aligned to the grade-level content standards (i.e., the CAS), but reduced in depth, breadth, and complexity. The EEOs can be found online at the following location:
http://www.cde.state.co.us/CoExtendedEO/StateStandards

CoAlt Assessment Frameworks were developed to better identify the content standards that may be assessed on the CoAlt: Science and Social Studies assessments. The frameworks were designed to assist educators, test developers, policy makers, and the public by clearly defining those elements of the EEOs that are suitable for state testing. However, the assessment frameworks are not designed to replace local curricula and should not be considered state curricula. The CoAlt: Science and Social Studies Assessment Frameworks can be found online at the following location:
http://www.cde.state.co.us/assessment/newassess-coaltsss

Descriptions of the content standards measured by the CoAlt: Science and Social Studies assessments are provided below.

- Science
    - Physical Science: Students know and understand common properties, forms, and changes in matter and energy.
    - Life Science: Students know and understand the characteristics and structure of living things, the processes of life, and how living things interact with each other and their environment.
    - Earth Systems Science: Students know and understand the processes and interactions of Earth's systems and the structure and dynamics of Earth and other objects in space.
- Social Studies
    - History: History develops moral understanding, defines identity, and creates an appreciation of how things change while building skills in judgment and decision-making. History enhances the ability to read varied sources and develop the skills to analyze, interpret, and communicate.
    - Geography: Geography provides students with an understanding of spatial perspectives and technologies for spatial analysis, awareness of interdependence of world regions and resources, and how places are connected at local, national, and global scales.

- Economics: Economics teaches how society manages its scarce resources, how people make decisions, how people interact in the domestic and international markets, and how forces and trends affect the economy as a whole. Personal financial literacy applies the economic way of thinking to help individuals understand how to manage their own scarce resources.

- Civics: Civics teaches the complexity of the origins, structure, and functions of governments; the rights, roles, and responsibilities of ethical citizenship; the importance of law; and the skills necessary to participate in all levels of government.

Item development for the CoAlt: Science and Social Studies assessments began in Summer 2012. The newly-developed items were then administered in a stand-alone field test in Spring 2013 for ES/MS and in Fall 2013 for HS. The goal of the stand-alone field tests was to collect student response data on the new items that would then be used to evaluate item quality.

After the newly-developed items were field tested and the item performance data were obtained, the items went through data review where CDE assessment specialists evaluated item performance to recommend if an item should be accepted or rejected based on the student performance data. The items that were accepted were re-classified in the item bank as available for use in future operational assessments.

Assessment Development Partners

The CoAlt: Science and Social Studies assessments are collaboratively developed by CDE, the Colorado educator community, and the assessment contractor, Pearson. Additional input and advice are provided by a Technical Advisory Committee (TAC).

Colorado Department of Education

CDE staff work closely with Pearson on each facet of the assessment with CDE serving as the ultimate approver.

Colorado Educator Community

Throughout the assessment development process, educators provide input into item and assessment development through participation in item writing, content and bias review, and standard setting meetings. For each meeting, an effort is made to involve educators who are representative of the entire state of Colorado, familiar with this population of students, and experts in the content areas assessed.

Pearson

Pearson is responsible for the content development, administration, and psychometrics of the CoAlt: Science and Social Studies assessments. This includes item and test development, enrollment, packaging and distribution, scoring, customer service, standard setting, score reporting, and psychometric services.

Technical Advisory Committee

The Technical Advisory Committee (TAC) is comprised of psychometric and assessment experts tasked with providing high-level consulting and expert advice regarding the creation of a reliable and valid assessment. Input is received on topics such as blueprint design, score reports, scaling and equating, and standard setting. The TAC members are as follows:

- Dr. Jamal Abedi, Professor, University of California, Davis
- Dr. Elliot Asp, Special Assistant to the Commissioner, Colorado Department of Education
- Dr. Jonathan Dings, Executive Director of Student Assessment and Program Evaluation, Boulder Valley School District
- Dr. Michael Kolen, Professor, University of Iowa
- Dr. Robert Linn, Distinguished Professor Emeritus, University of Colorado at Boulder
- Dr. Martha Thurlow, Director, National Center on Educational Outcomes

CHAPTER 2: ITEM DEVELOPMENT AND ITEM BANKING

The test development process involves various steps. To the extent possible, CoAlt: Science and Social Studies follows the same test development process as CMAS: Science and Social Studies. However, the CoAlt test development process reflects the unique characteristics of the assessment program, specifically the items types included in the assessments and the needs of the population of students who take alternate assessments. CDE relies greatly on input from Colorado educators (both general and special educators) and alternate assessment specialists throughout the development process to ensure that CoAlt: Science and Social Studies assessments are equitable for students and that they accurately measure the content.

The validity of a state assessment relies on the methodology that frames the development and design of the assessment. In support of that claim, Pearson upheld these considerations as the cornerstones of the CoAlt: Science and Social Studies item and test development:

- The test specifications ensure that the CoAlt: Science and Social Studies items align to the EEOs they are intended to measure.
- The CoAlt: Science and Social Studies item development plan (IDP) is designed to produce and maintain a robust item bank.
- The CoAlt: Science and Social Studies item and test development processes are compliant with industry standards.

Pearson's proprietary software Item Tracker Test Builder (ITTB) is used to support the item and test development process. As described in the following sections, items can be classified in different groups, each representing a step in the item development process.

Item-Writing Process

The item-writing process for the CoAlt: Science and Social Studies assessments began in Summer 2012. The items were written by Colorado educators, content specialists, and professional item writers with guidance and input from CDE. The SR and SPT items for each assessment were written to measure concepts and skills found in the EEOs. The initial item writing development effort was bolstered with an overage of items per standard in order to ensure depth of the operational item bank moving forward in the event that an item performed poorly during field testing.

The item-writing process included the following steps:

Specifications Development

Pearson created the test blueprint with input and approval from CDE. The CoAlt: Science and Social Studies test blueprints contain the number of test items by content standard and item type. The blueprints can be found in Appendix B. During this stage, Pearson also created an IDP which delineates the target number of items per standard, grade level expectation (GLE), and

EEO. The IDP helped to forecast the number of items that were needed to create a robust operational item bank that would be refreshed over time.

Item Development

After the test blueprints and IDPs were developed, item writers were trained using various guides and resources developed during specifications development. These documents included the content standards, item specifications, and item writing guidelines. Pearson's assessment specialists reviewed each batch of items and provided feedback as often as necessary, focusing on both the technical quality of the items and their match to the standards.

Item Reviews

After the items were written and uploaded into ITTB, they were subjected to content and editorial reviews, including inspection for adherence to universal design (UD) principles. Following field testing, each field tested item was further analyzed during a data review before inclusion in the operational assessment.

Content and Editorial Review

Pearson's Assessment Development Services Department conducted a content review to evaluate standard and knowledge-and-skill match, quality of the items, adherence to the principles of UD, cognitive demand, item relevance to the purpose of the test, readability, and appropriateness of graphics. Members of the development team performed additional fact-checking to ensure accuracy of item content.

The Editorial Department checked items for clarity, correctness of language, appropriateness of language for the grade level, adherence to style guidelines, and conformity with acceptable item-writing practices. In addition, editors with content expertise in the areas of science and social studies reviewed the items. The content editors added a valuable layer of content validation and fact-checking. Alternate assessment specialists, who have expertise in the areas of special education and students with disabilities, reviewed all items to ensure that the items were appropriate for students with significant cognitive disabilities.

Pearson performed a UD review to assess item accessibility irrespective of diversity of background, cultural tradition, and viewpoints; to evaluate changing roles and attitudes toward various groups; to review the role of language in setting and changing attitudes toward various groups; to appraise contributions of diverse groups (including ethnic and minority groups, individuals with disabilities, and women) to the history and culture of the United States and the achievements of individuals within these groups; and to edit for inappropriate language usage or stereotyping with regard to sex, race, culture, ethnicity, class, or geographic region. These reviews were conducted to ensure that all students would have an equal opportunity to demonstrate achievement regardless of their gender, ethnic background, religion, socio-economic status, or geographic region. Items that were accepted based on the Pearson reviews were re-classified in ITTB as ready for CDE review.

Once the Pearson reviews within each department were completed, the items were submitted to CDE for their review. CDE reviewed the items checking to make sure the content is accurate, the EEO alignment is appropriate, the language is appropriate for the grade level and student population, and the graphics are clear and relevant to the item. Items that were accepted based on the CDE review were re-classified in ITTB as ready for bias and sensitivity review.

Accepted items were then reviewed by Colorado educators to evaluate whether the items are properly aligned to the content standards and to identify if any potential bias exists in the items. The unique needs of students with significant cognitive disabilities were also considered in the content and bias reviews of assessment items. These reviews included content-specific general educators, special educators, and teachers of students who are culturally and linguistically diverse. Items that were accepted based on the educator committee recommendation were re-classified in ITTB as ready for field testing.

Data Review

After the development of the items, selected items were administered in a stand-alone field test. Following the field test administration, CDE and Pearson assessment specialists and psychometricians reviewed student performance on the items. Pearson provided the results of all statistical analyses. These analyses included classical statistics and item response theory statistics so that CDE and Pearson could make informed judgments. The statistical information provided included:

- Classical statistics, such as the item sample size, item mean score, item-total correlation, and response distribution
- Item response theory statistics, such as item difficulty and fit values

Due to small sample sizes, statistical bias analyses were not conducted. Statistical bias analyses by subgroup were conducted once operational data became available for assessment items.

Field test items that were accepted based on the evaluation of student performance were re-classified in the item bank as available for use on future operational assessments. Items that were rejected were re-classified to eliminate them from use on a test. These items may be modified and field tested again on future test forms.

Item Banking Process

Item banking is handled by the Pearson Item Tracker software, which houses the items from creation through retirement in a secure environment. The web-based secure item bank serves as the repository from which items for current and future forms of the assessment are drawn.

Following the stand-alone field test and data review process, content specialists met to conduct a final examination of items prior to their inclusion in the operational item bank. This review process provided content specialists with an opportunity to discuss their concerns about item content, format, bias, and fit. These discussions were used to make inclusion decisions about the items on the operational test forms. Items that passed all stages of the development process (e.g.,

item review, field test, data review, and bias review) were placed in the operational item bank to become eligible for use in future assessments.

Item Bank Statistics

The metadata for each item are included in the item bank, which includes: the item image, test date, cognitive level, the assessed content standard, the form on which the item appeared, the item position on the form, the item type, the correct key, and the maximum number of points possible for a correct answer.

The item summary statistics include the item sample size, item mean score, item-total correlation, and a response distribution that presents the percentage of students achieving each score point both overall and by ability level. When available, statistical bias analyses are also included. A more complete description of these variables is included in the Data Review section of this report.

CHAPTER 3: TEST CONSTRUCTION

Pearson is responsible for the implementation and monitoring of all phases of the test construction process. Test forms are constructed through an iterative process between Pearson content and Pearson psychometric staff. CDE then reviews the forms, provides feedback, and gives final approval as described below.

When building operational test forms, the assessment specialists select a set of operational items in accordance with the test blueprint and test construction specifications. Items selected for use operationally must meet the blueprint and should include a variety of topics and contexts with specified psychometric targets.

The following guidelines are used during form construction:

- adherence to the test blueprints
- review of the item statistics and adherence to the statistical criteria found in the test construction specifications
- balance of gender, ethnicity, geographic regions, and relevant demographic factors
- selection of items with various stimuli types throughout the test form to enhance the test-taker experience by providing variation in the items presented
- efficient and deliberate use of varied content representative of the knowledge and skills in the content standards
- review of full form, including field test items, for instances of clueing and/or content overlap

After the initial operational items are selected, the test form is reviewed by two Pearson assessment specialists. Each assessment specialist verifies that the form meets test blueprint (i.e., the required number of items, EEO coverage, and item types). The form is then presented to psychometrics for analysis; the psychometrician verifies that the form falls within the established psychometric and blueprint parameters.

Once the form is vetted internally, the form is presented to CDE for review. If needed, CDE and Pearson assessment specialists and psychometricians collaborate to finalize the form. This can be an iterative process with the end result being CDE's form approval.

After the operational form is approved, field test items are selected from the items in ITTB that are coded as ready for field testing. The assessment specialists assemble field test item sets so that they comprise the appropriate distribution of standards, item types, topic coverage, and key distributions. They also review item replacement for future years to ensure appropriate item rotation. Items chosen are embedded on the operational form in a designated location.

The specific responsibilities for Pearson and CDE during test construction are outlined below:

- Pearson responsibilities:
  - generate a test construction schedule
  - select and sequence a proposed set of operational items
  - select and sequence of a proposed set of field test items
  - conduct content and psychometric reviews of each proposed set of items
  - construct a customer test map that provides content and psychometric information for each proposed item
  - manage the customer review process
  - provide the customer with copies of proposed items and the associated customer test map
  - revise the proposed item set based on customer comments
  - document edits/comments provided by the customer

- CDE responsibilities:
  - review and approve item selection based on content and psychometric properties
  - review and approve test for layout, item sequencing, and avoidance of cueing

A high-level description of the number of operational test forms and the number of operational and embedded field test items is shown in Table 1.

Table 1. CoAlt: Science and Social Studies Operational Assessments

Assessment	Number of Operational Test Forms	Test Blueprint Length		Embedded FT Items Per Form		Total Test Length Per Form	Total Points Per Form
		4-Point SRs	6-Point SPTs	4-Point SRs	6-Point SPTs		
Grade 4 social studies	1	15	2	4	2	23	72
Grade 5 science	1	15	2	4	2	23	72
Grade 7 social studies	1	15	2	4	2	23	72
Grade 8 science	2	24	2	3	1	30	108
HS social studies	2	23	3	3	1	30	110
HS science	2	23	3	3	1	30	110

CHAPTER 4: TEST ADMINISTRATION PROCEDURES

This chapter provides information related to the CoAlt: Science and Social Studies administration procedures. Training of Colorado districts, schools, and teachers was a high priority because the assessments involve specifically-developed materials, administration requirements, and score entry steps. CoAlt: Science and Social Studies administration and training procedures were standardized to ensure that students would receive comparable assessment results. Test administration procedures and online score entry information were communicated via manuals and trainings as described below.

Manuals

Several manuals were created to support the CoAlt: Science and Social Studies administration. These manuals include the following:

- *Colorado Measures of Academic Success (CMAS) and Colorado Alternate Assessment (CoAlt): Science and Social Studies Procedures Manual*
- *CoAlt: Science and Social Studies Examiner's Manual*
- *CoAlt: Science and Social Studies Data Supplement*
- *Colorado Accommodations Manual and Accommodations Guide for English Learners*
- *PearsonAccess User Guide*

Training

CDE and Pearson conducted several in-person administration trainings for District Assessment Coordinators in Colorado. CoAlt training materials were posted to the Support tab of PearsonAccess to provide District Assessment Coordinators with access to materials well in advance of the administration of the assessment. In addition, Pearson customer service center staff were trained to answer questions thoroughly and knowledgably and to escalate inquiries as necessary. CDE hosted WebEx training sessions covering CoAlt eligibility requirements, the test design, accommodations, distribution of materials, test security, and PearsonAccess tasks necessary to set up and administer the assessment and access test results.

Accessibility and Accommodations

The CoAlt: Science and Social Studies assessments were developed to be accessible for students with significant cognitive disabilities. Accessibility was considered from the beginning of the test development process and is inherent within the CoAlt assessment and administration. For example, CoAlt assessments are read aloud to students and all students who take CoAlt are assessed individually. In addition, the assessment can be administered over several days for those students who need more time due to limitations in behavioral control, stamina, or

communication. Even though the assessments are designed to be accessible, students with disabilities taking the assessment may still require changes to the assessment procedures, or accommodations, in order to accurately demonstrate their knowledge and skills of the content. This also includes English learners (ELs) who need language supports to demonstrate their knowledge of the content.

Accommodations provide a student with an opportunity to engage with the assessment while not affecting the reliability or validity of the assessment. Accommodations can be adjustments to the test presentation, materials, environment, or response mode of the student and are based on student need. Accommodations should not provide an unfair advantage to any student. Providing an accommodation for the sole purpose of increasing test scores is not ethical. Accommodations must be documented in the student's IEP and used regularly during classroom instruction and assessments prior to the assessment window to ensure the student can successfully use the accommodation.

Although accommodations are used for classroom instruction and assessments, some may not be appropriate for use on statewide assessments. As a result, it is important that educators become familiar with the state assessment policies about the appropriate use of accommodations and that districts have a plan in place to ensure and monitor the appropriate use of accommodations. Accommodations recorded in the online scoring system for the CoAlt: Science and Social Studies could include the following:

- Assistive technology
- Braille
- Eye gaze
- Modified picture symbols (enlarged pictures and/or pictures of real objects)
- Objects (three-dimensional or representational objects)
- Translation into student's native language
- Other
- None

Test Security

Districts were trained on assessment security to ensure that security procedures were maintained during the test administration. Materials used during the administration of the assessment were to be kept in locked storage locations when not under the direct supervision of approved assessment coordinators or Test Examiners. All state, district, and/or school personnel signed the Security Agreement prior to handling test materials. By signing the Security Agreement, personnel agreed to a set of security guidelines that required them to follow all procedures set forth in manuals. Personnel could not divulge the contents of the assessment, copy any part of the assessment except for students with allowable CoAlt accommodations, or review test questions with students. They also could not allow students to remove test materials from the room where testing takes place or interfere with the independent work of any student taking the assessment.

CHAPTER 5: SCORING THE ASSESSMENTS

Test Examiners use two rubrics to evaluate student performance on the CoAlt: Science and Social Studies assessments. A unique rubric is built into each item type. The rubrics were developed taking into account the characteristics of the students taking CoAlt. Students with the most significant cognitive disabilities often require direct, structured learning experiences and various levels of support, in addition to their usual accommodations, in order to demonstrate their knowledge of the content. As a result, each rubric incorporates the level of independence (i.e., the level of teacher support needed to demonstrate performance on the item) and the student's response into the rubric's score points. This scoring method was developed to closely mirror the type of instruction and levels of support the students typically receive in the classroom.

Selected Response Scoring Rubric

SR items contain a primary prompt with a question and three answer options from which the student selects an answer. Test Examiners score the student's performance on the SR item using a four-point rubric found in Table 2. To administer the item, the Test Examiner presents scripted text containing the primary prompt and answer choices to the student. If the student responds correctly with no supports from the teacher, or after a single repetition of the primary prompt, the student receives a score point of 4. If the student responds incorrectly or does not respond to the primary prompt after the Test Examiner repeats it once, an additional prompt is presented to the student. The additional prompt provides the student with an example that is similar to the primary prompt and answer options. The Test Examiner then repeats the primary prompt after the additional prompt is presented. If the student responds correctly after the additional prompt is presented, the student receives a score point of 3. If the student responds incorrectly or does not respond, the student is presented with the correct response and is presented with the primary prompt again to have another opportunity to respond. If the student responds correctly after being presented with the correct answer, the student receives a score point of 2. If the student responds incorrectly after being presented with the correct response, the student receives a score point of 1. There are sometimes instances in which a student does not engage with the item even with the scaffolded supports provided within the item. If a student does not provide a response when provided with all of the supports for the item, the student receives an NR, or no response, which represents 0 points.

Table 2. Selected Response Scoring Rubric

Score Point	Selected Response Scoring Rubric
4	Student responds correctly, independently
3	Student responds correctly after being presented with an additional prompt
2	Student responds correctly after being presented with the correct response
1	Student responds incorrectly
NR	Student does not respond

Supported Performance Task Scoring Rubric

SPT items consist of three related questions called prompts. For this item type, students are required to manipulate option cards by placing them in designated areas on a diagram or chart in order to respond to each of the three prompts. Student performance on each prompt is scored using a two-point rubric found below in Table 3. To administer the item, the Test Examiner has the student response page and option cards ready for the student to engage with the item. The Test Examiner then presents the scripted text for the first prompt. If the student responds correctly, the student receives 2 points. If the student responds incorrectly, the student receives 1 point. If the student does not provide a response to the prompt, the student receives an NR, or no response, which represents 0 points. When an incorrect response is given or the student does not respond, the Test Examiner places the correct option card in the response box and tells the student the correct answer. After the first prompt is completed, the Test Examiner then completes the same steps for the remaining two prompts.

Table 3. Supported Performance Task Scoring Rubric

Score Point	Supported Performance Task Scoring Rubric (utilized for each of three prompts within each task)
2	Student responds correctly
1	Student responds incorrectly
NR	Student does not respond

Additional Scoring Information

Test Examiners record all student scores within the test book or on the score recording form that is included with the task manipulatives set provided for each test. Recorded responses are then entered into PearsonAccess, the online score entry system. The SPT items involve an additional step that occurs after the student's individual prompt scores are entered into PearsonAccess. The points for the three prompts are added together to provide one score for the SPT item, with the maximum of 6 points possible. On the CoAlt: Science and Social Studies assessments, SR and SPT items never have more than three answer options, but there can be as few as two answer

options for the prompts in the SPT items. The number of answer options available for the SPT items can vary by item and content area.

CHAPTER 6: STANDARD SETTING

To support the interpretation of student results, student performance on the CoAlt: Science and Social Studies assessments is described in terms of four performance levels: Novice, Developing, Emerging, and Exploring. After the first operational administration of the ES/MS assessments in Spring 2014, a standard setting meeting was held to determine the performance standards (see the *Spring 2014 CoAlt: Science and Social Studies Technical Report* for more details regarding the ES/MS standard setting). Likewise, after the first operational administration of the HS assessments, a standard setting meeting was also held to determine performance standards. Performance standards specify what level of performance on a test is required for a test taker to be classified in a given performance level. The HS standard setting meeting was held in February 2015 with Colorado educators.

The Modified Extended Angoff approach (Cizek, 2012; Cizek, Bunch, & Koons, 2004; Hambleton & Plake, 1995) was used to set performance standards on the assessments. With this methodology, panelists review performance level descriptors (PLDs) to conceptualize "threshold" students (students just barely in a particular performance level) and then make a judgment about what score a threshold student should receive on each item to be considered "just-barely" in a performance level. The individual item-level cut scores for each performance level are then summed to obtain the recommended cut score for each performance level. The Reasoned Judgment approach (Roeber, 2002) was also used in this methodology to help panelists think about the level of content knowledge that may be needed for a student to earn a specific rubric score, the patterns of performance (i.e., combinations of item scores) that lead to overall test scores, and whether various scoring patterns make sense for a given performance level. Different patterns of student performance, called score profiles, were presented to panelists with this approach. The score profile is a graphical representation of how a student could achieve a specific test score.

The standard setting meeting included approximately ten panelists for each subject-area committee. Panelists were grouped into tables of three within each meeting room. Panelists were selected for participation by CDE to represent the state in terms of gender and ethnicity as well as relevant demographic characteristics (e.g., school size, geographic location). The CoAlt panelists included educators who taught at the high school level, including special educators with experience working with students with significant cognitive disabilities, special educators with experience working with students with other types of disabilities, and content experts with knowledge of the subject-area curriculum. In addition to classroom teachers, special education administrators and higher education representatives also participated in the meeting. Panelists from the CMAS Science and Social Studies HS standard setting meeting were also recruited to participate in the CoAlt HS standard setting meeting. Including panelists from the prior CMAS standard setting meeting helped provide context to the CoAlt panelists regarding how the earlier recommended performance standards were determined.

The standard setting for the CoAlt: Science and Social Studies HS assessments was held on February 18–19, 2015. During the two-day meeting, panelists from each of the two subject-area standard setting committees received training on the assessment and the standard setting process, reviewed the grade-level PLDs, reviewed the Fall 2014 operational items, reviewed the threshold

student descriptors, and applied the Modified Extended Angoff method to establish cut score recommendations across three rounds of rating. During the process of establishing cut score recommendations, panelists also reviewed the content assessed by the CoAlt items and matched the items to performance levels based on the concepts and skills found in the PLDs, engaged in table and committee-level discussions, and considered the impact of their cut scores on student performance when making their cut score recommendations.

The proposed recommended cut scores from standard setting were presented to the State Board of Education for review. The Colorado State Board of Education approved the HS science recommended cut scores for one year and the approval of the HS social studies cut scores are currently pending.

For the HS science and social studies assessments, an estimated 37% and 48% of students, respectively, were in the top two performance levels (Novice Level and Developing Level). More details about the CoAlt: Science and Social Studies HS standard setting meeting and the final cut scores can be found in the full standard setting report in Appendix C.

CHAPTER 7: REPORTING

Several score reports are generated to communicate student performance on the CoAlt: Science and Social Studies assessments. The information below describes the types of scores given on reports and the types of reports available. For additional details on score reports, see the Spring 2015 Score Interpretive Guide at http://www.cde.state.co.us/assessment/newassess-sum.

Description of Scores

CoAlt: Science and Social Studies reports provide information about student performance in terms of scale scores, performance levels, and percent of points earned.

Scale Scores

A scale score is a conversion of a student's total test score (i.e., the total number of points earned on a test) onto a scale that is common to all test forms for that assessment. Scale scores are particularly useful for comparing assessment scores across years from different test administrations. For the CoAlt: Science and Social Studies assessments, students receive an overall test scale score. An indicator of the range of scale scores a student would likely receive if the assessment was taken multiple times is also provided. Each assessment's scales range from 0 to 250. Chapter 8 provides technical details related to scale development for the CoAlt: Science and Social Studies assessments.

Performance Levels

Performance levels are reported at the overall test level. Examinees are classified into performance levels based on their scale score as compared with the cut scores, which were obtained from standard setting. CoAlt: Science and Social Studies have four performance levels:

- Novice
- Developing
- Emerging
- Exploring

For those students who did not respond to any of the CoAlt assessment items, an "Inconclusive" designation is reported on their individual student reports. These students are given a scale score of zero and included in the Exploring Level for aggregation purposes.

Percent of Points Earned

The percent of points earned is provided for each assessment. Unlike scale scores, the percent of points earned cannot be compared across years because individual items change from year to year and the difficulty of the items may not be the same.

Score Reports

Two types of score reports are provided: student level and aggregate. Sample score reports can be found in Appendix D.

Student Performance Reports

The Student Performance Report provides information about the performance of a particular student on the CoAlt: Science and Social Studies assessment. The student's scale score, associated performance level, and percent of points earned are displayed on a one-page report along with comparative information related to state performance. In addition, performance level descriptors are provided. Student Performance Reports are printed and shipped to districts for distribution to students and parents.

Aggregate Reports

Two types of aggregate reports are produced for CoAlt:

- Content Standards Roster
- Performance Level Summary

These reports are produced at the school, district, and state levels and provide summary information for a given school or district. State, district, and school reports are provided electronically through PearsonAccess Test Results, and access to the reports is limited to users approved by CDE.

CHAPTER 8: CALIBRATION, EQUATING, AND SCALING

Item Response Theory (IRT) was used to develop, calibrate, equate, and scale the CoAlt: Science and Social Studies assessments. The Rasch Partial Credit Model was the measurement model used for test construction, calibration, scaling, and equating and to maintain and build the item bank. All calibration, scaling, and item-model fit analyses were accomplished within the IRT framework. The initial administration of the CoAlt: Science and Social Studies ES/MS assessments in Spring 2014 and the HS assessments in Fall 2014 determined the base scale for the assessments.

Calibration

The Rasch Partial Credit Model

Calibration is the process used to obtain item parameter estimates and then place all items and students on a common scale. For each grade-level assessment, the Rasch Partial-Credit Model (RPCM) was used to place the CoAlt items and student proficiency on the same Rasch scale. The model is an extension of the Rasch one-parameter IRT model attributed to Georg Rasch (1966), as extended by Wright and Stone (1979), Masters (1982), and Wright and Masters (1982). The RPCM was selected because of its flexibility in accommodating various item types (i.e., multiple-choice items and items with multiple response categories). The RPCM maintains a one-to-one relationship between scale scores and raw scores, meaning each raw score is associated with a unique scale score. It is the underlying Rasch scale that allows for comparisons of student performance across years and facilitates the maintenance of equivalent performance standards across years.

The RPCM is defined by the following mathematical measurement model where, for a given item involving *m*+1 score categories, the probability of person *n* scoring *x* on question *i* is given by:

$$P_{xni} = \frac{exp\sum_{j=0}^{x}(\theta_n - \delta_{ij})}{\sum_{k=0}^{m_i} exp\sum_{j=0}^{k}(\theta_n - \delta_{ij})} \quad x = 0, 1, \ldots m_i$$

The RPCM provides the probability of a student scoring *x* on *m* steps of question *i* as a function of the student's proficiency level, θ_n (sometimes referred to as "ability"), and the step difficulties, δ_{ij}, of the *m* steps in question *i*.

Equating and Scaling

Equating involves adjusting for differences in the difficulty of test forms, both within and across assessment administrations. Equating makes certain that students taking one form of a test are neither advantaged nor disadvantaged when compared to students taking a different form. Each time a new test form is constructed, equating is used to allow scores on the new form to be comparable to scores on the previous form by placing the scores on both forms on the same scale. It is the underlying Rasch scale obtained from calibration that facilitates equating of test forms. The Rasch scale can then be transformed to create scale scores to allow for the interpretation of test scores.

Equating and Scaling

The Fall 2014 administration of the CoAlt: Science and Social Studies HS assessments represents the first operational tests on the newly developed Rasch scale. In the following years, equating will be used to place the new HS test forms on this newly-developed operational scale. To obtain Rasch item parameter estimates for the Fall 2014 HS assessments, the RPCM was applied to the operational and embedded field test items. Winsteps (Linacre, 2011) was used for all calibrations. The calibrations of the operational and embedded field test items for each assessment occurred in several steps. First, the operational items were calibrated. Next, the embedded field test items were calibrated with the operational items using a fixed common items calibration approach. With this calibration method, the embedded field test items are calibrated with the operational item parameters fixed at their previously-estimated values in order to place the embedded field test items on the same scale as the operational items.

The fixed common items approach was also used to equate the Spring 2015 ES/MS assessments to their Spring 2014 operational scales. The operational items used to equate the 2015 assessments to the 2014 scales are called anchor items. The anchor items are a set of common items that are placed on forms from adjacent administrations. This set of items represents the CoAlt blueprint in terms of content and item types and represents approximately 60% of a full form. To obtain equated Rasch parameter estimates for the Spring 2015 ES/MS assessments, anchor item parameter estimates were fixed to their 2014 parameter estimates before calibrating the remaining non-anchor operational items. This method placed the non-anchor operational items on the same scale as the anchor items. The RPCM and Winsteps were used for all ES/MS calibrations.

The stability check for the anchor items was conducted using classical item analysis, scatter plots of item difficulties, and displacement estimates from Winsteps. Displacement estimates greater than or equal to ±0.30 was used as the flagging criteria. Items flagged from the stability check are examined and consideration is given to the impact of flagged item(s) on the content representativeness of the resulting anchor set. A flag alone is not the sole criteria for removing an item from the anchor item set. It is important to also make sure that the remaining anchor set continues to be representative of the overall content and structure of the test.

Ability Estimates

After the item parameter estimates were obtained for the ES/MS and the HS operational items, student proficiencies were estimated for each assessment by conducting an anchored calibration of the operational items' item parameter estimates. Estimates were obtained via the joint maximum likelihood method (JMLE) applied within the Winsteps software program.

Scale Scores

Student proficiencies were then transformed to scale scores ranging from 0 to 250 with a mean of 150 and standard deviation of 40. The CoAlt: Science and Social Studies scale scores represent linear transformations of the student proficiencies (θ). The transformation is made by first multiplying any given θ by a slope (a) and then adding an intercept (b). The following linear transformation was used to convert student proficiency estimates into scaled scores (SS):

$$SS = (a * \theta) + b$$

The a and b values are referred to as scaling constants. These scaling constants will be applied each year to the Rasch proficiency estimates for that year's set of operational items. After the scale scores were obtained, the lowest observable scale score (LOSS) and the highest observable scale score (HOSS) for the performance levels were applied. The LOSS and HOSS for the performance levels were set to 1 and 250, respectively.

Steps in the Calibration and Scaling Process

The entire process previously described was repeated for each CoAlt: Science and Social Studies assessment. All steps were independently replicated by at least two members of the Pearson psychometric team to ensure the accuracy of the processes.

Data Preparation

Prior to any analyses, several steps were completed in preparation.

- The data file containing student responses was verified and exclusion rules were applied.
- Traditional item analyses of all items were conducted prior to calibration.
- Incomplete data matrices (IDMs) were created.

A traditional item analysis of all operational and embedded field test items was conducted prior to calibration. The purpose of this analysis was to obtain classical statistics used to evaluate item performance. The following statistics were calculated:

- Item sample size
- Response distribution

- Item mean score
- Item-total correlation

Calibration

Several different calibrations were done to obtain item parameter estimates for the operational and embedded field test items.

- Operational Items
    - Used Winsteps control files and IDM to obtain operational item parameter estimates
        - Obtained operational Rasch item difficulty values, step deviation values, and item fit values
- Embedded Field Test Items
    - Used Winsteps control files and IDM to scale the embedded field test item parameter estimates to the operational scale by fixing the item parameter estimates of the operational items
        - Obtained embedded field test Rasch item difficulty values, step deviation values, and item fit values

CHAPTER 9: RELIABILITY

A variety of statistics can be calculated that pertain to the reliability of the CoAlt: Science and Social Studies assessments. In this report, Cronbach's alpha, standard error of measurement (SEM), conditional standard error of measurement (CSEM), decision consistency and accuracy, and inter-rater agreement will be described. For these statistical estimates for the Fall 2014 and Spring 2015 administrations, see Part II of this document.

Cronbach's Alpha

Within the framework of Classical Test Theory, an observed test score is defined as the sum of a student's true score and error ($X = T + E$, where X = the observed score, T = the true score, and E = error). A true score is considered the student's true standing on the measure, while the error score reflects a random error component. Thus, error is the discrepancy between a student's observed and true score.

The reliability coefficient of a measure is the proportion of variance in observed scores accounted for by the variance in true scores. The coefficient can be interpreted as the degree to which scores remain consistent over parallel forms of an assessment (Ferguson & Takane, 1989; Crocker & Algina, 1986). There are several methods for estimating reliability; however, in this report, an internal consistency method is used. In this method, a single form is administered to the same group of subjects to determine whether examinees respond consistently across the items within a test. A basic estimate of internal consistency reliability is *Cronbach's Coefficient Alpha* statistic (Cronbach, 1951). Coefficient alpha is equivalent to the average split-half correlation based on all possible divisions of a test into two halves. Coefficient alpha can be used on any combination of dichotomous (two score values) and polytomous (two or more score values) test items and is computed using the following formula:

$$\alpha = \frac{n}{n-1}\left(1 - \frac{\sum_{j=1}^{n} S_j^2}{S_X^2}\right)$$

where n is the number of items,

S_j^2 is the variance of students' scores on item j, and

S_X^2 is the variance of the total-test scores.

Cronbach's alpha ranges in value from 0.0 to 1.0, where higher values indicate a greater proportion of observed score variance is true score variance. Two factors affect estimates of internal consistency: test length and homogeneity of items. The longer the test, the more observed score variance is likely to be true score variance. The more similar the items, the more likely examinees will respond consistently across items within the test. For CoAlt, coefficient alpha estimates are provided for the overall test as well as for subgroups. The coefficient alpha estimates can be found in Tables 7–27.

Standard Error of Measurement

The SEM is another measure of reliability. This statistic uses the standard deviation of test scores along with a reliability coefficient (such as coefficient alpha) to estimate the number of score points that a student's test score would be expected to vary if the student was tested multiple times with equivalent forms of the assessment. It is calculated as follows:

$$SEM = s_x \sqrt{1 - \rho_{XX'}}$$

where s_x is the standard deviation of test scores and

$\rho_{XX'}$ is the reliability coefficient.

There is an inverse relationship between the reliability coefficient (e.g., alpha) and SEM: the higher the reliability, the lower the SEM. SEM values can be found in Tables 7–27.

Conditional Standard Error of Measurement

While the SEM provides an estimate of precision for an assessment, the CSEM considers how measurement error likely varies across the scale score. In other words, the CSEM provides a measurement error estimate at each score point on an assessment. Because there is typically more information about students with scores in the middle of the score distribution where scores are most frequent, the CSEM is usually smallest, and thus the scores are most reliable, in the middle of the score distribution.

An IRT method for estimating score-level CSEM is used because test- and item-level difficulties for CoAlt: Science and Social Studies were calibrated using the Rasch measurement model. By using CSEMs that are specific to each scale score, a more precise error band can be placed around each student's observed score. CSEM values are provided in Tables 48–53.

Decision Consistency and Accuracy

The CoAlt: Science and Social Studies scales are divided into four performance levels: Novice, Developing, Emerging, and Exploring. Based on a student's scale score, the student is classified into one of the four performance levels. The consistency and accuracy of these performance level classifications is another important aspect of reliability to examine.

The consistency of a decision refers to the extent to which the same classification would result if a student were to take two parallel forms of the same assessment. However, since test-retest data are not available, psychometric models can be used to estimate the decision consistency based on test scores from a single administration. The accuracy of a decision refers to the agreement between a student's observed score classification and a student's true score classification, if a student's true score could be known.

Procedures developed by Livingston and Lewis (1995) were used to estimate the consistency and accuracy of performance level classifications for the CoAlt: Science and Social Studies assessments. The probability of a consistent classification (PC) is the probability that the performance level classification the student received is consistent with the classification that the student would have received on a parallel form. This probability should be a high value. The probability of consistent classification by chance is the probability that the performance level the student received is accurate and occurred by chance. The probability of misclassification (PM) is also provided and is the probability the performance level a student received is incorrect (i.e., 1 minus PC). The probabilities of consistent classification by chance and misclassification should be low. Kappa describes the agreement between classifications on two parallel forms. The kappa value can be interpreted as follows (Altman, 1991):

Value of Kappa	Strength of Agreement
< 0.20	Poor
0.21 – 0.40	Fair
0.41 – 0.60	Moderate
0.61 – 0.80	Good
0.81 – 1.00	Very Good

The probability of an accurate classification (PA) is the probability that the performance level classification a student received is correct and is based on the agreement between the observed classification on the actual test form and true classification. PA values should be high. The probability of false positives (FP) and false negatives (FN) are also provided and these values should be low. Consistency and accuracy estimates are provided in Table 54.

Inter-Rater Agreement

An additional form of reliability, called inter-rater agreement, is also evaluated for CoAlt administrations. Inter-rater agreement examines the extent to which examinees would obtain the same score if scored by different scorers. For this method, two raters simultaneously observe a student taking the CoAlt assessment: a test examiner (i.e., a student's teacher) and a score monitor. Both raters evaluate student performance and enter their scores into the online score entry system. The two independent ratings are then compared to determine the consistency of the ratings. The second set of scores provided by the score monitor is used only to establish the level of consistency in scoring. They are not used for student scoring and reporting.

Procedure

The sampling plan included eight score monitors each conducting observations for both science and social studies which would yield 24 students with second scores for each subject. To identify score monitors, CDE solicited educators who held a Colorado teaching license, were part-time or retired teachers or had teaching experience at the university level, and were familiar with the population of students who take CoAlt. Seven educators were identified who met the selection criteria.

The selected score monitors then received training by CDE and Pearson staff via teleconference. The training was conducted so that the score monitors would be consistent in their approach and scoring when conducting their observations. As part of the training, the meeting facilitators reviewed the purpose of score monitoring, the test materials, the scoring process, and the test administration procedures. Meeting facilitators also reviewed the score monitor observation materials that were to be used to collect the second scores and document a description of the testing environment and test procedures used by the student's teacher.

CDE then recruited schools to participate in score monitoring. The schools solicited were recruited based on the demographic diversity of students, the number of students participating in CoAlt assessment at the schools, and the proximity of the schools to the selected score monitors. The score monitoring was conducted during the Fall 2014 CoAlt operational test window across eight school districts. Due to resource and timing challenges for observations, the target sample of 24 students per subject was not reached. Ultimately, eleven student observations for each subject were used to evaluate inter-rater agreement.

Results

In general, the score monitor observations indicated that test administrators were adhering to the testing processes outlined by the state. When noting their observations of the test administrators, the majority of score monitors indicated that the test administrators seemed comfortable with the students, were well prepared for administering the test, and followed the instructions provided in the test manual and test books. A few challenges in the test administration procedures were noted, and CDE addressed the concerns in test administrator training sessions beginning with the Spring 2015 CoAlt administration.

The metrics used to evaluate inter-rater agreement were the correlation between two independent ratings, perfect agreement, and adjacent agreement. Correlations are used to evaluate the relationship or association between pairs of scores. In this instance, test examiner scores and score monitor scores were the pair of scores used to calculate the correlations. Perfect agreement is when the two independent scorers assign the same score to the same piece of student work. Adjacent agreement is when the two independent scorers assign score points that differ by one (e.g., 1 and 2) to the same piece of student work. Descriptive statistics for each subject and for the samples can be found in Table 4.

Table 4. Descriptive Statistics for CoAlt: Science and Social Studies HS

	Population			Sample		
HS Subject	N	Males	Females	N	Males	Females
Science	446	61%	39%	11	82%	18%
Social Studies	444	61%	39%	11	64%	36%

Correlation coefficients were calculated for each subject and are summarized in Table 5. The correlation of the item-level scores between the first and second scores was 0.97 for science and 0.96 for social studies. The correlation of the test-level scores between the first and second scores was 0.99 for science and 0.99 for social studies.

Table 5. Correlations between First and Second Scores

HS Subject	Item-Level Correlation	Test-Level Correlation
Science	0.97	0.99
Social Studies	0.96	0.99

Perfect and adjacent agreement rates were calculated for each subject and are summarized in Table 6. Perfect agreement rates of item-level scores were 94% for science and 93% for social studies. Adjacent agreement rates of item-level scores were 5% for science and 5% for social studies.

Table 6. Percent Agreement between First and Second Scores

HS Subject	Perfect Agreement	Adjacent Agreement
Science	94%	5%
Social Studies	93%	5%

The correlation coefficients indicate that the first and second scores in this study are highly related. The perfect and adjacent agreement rates also indicate high levels of agreement between scores. When perfect and adjacent agreement rates are combined, 98% to 99% of raters had the same or adjacent item scores. Thus, the results indicate consistency of scoring of the CoAlt assessment for the study participants. Plans are ongoing to establish ways to increase the sample size and the variability in the sample demographics and regional representation so the results can be generalized to the population.

CHAPTER 10: VALIDITY

"Validity refers to the degree to which evidence and theory support the interpretations of test scores for proposed uses of tests" (AERA, APA, NCME, 2014). As such, it is not the CoAlt: Science and Social Studies assessments that are validated but rather the interpretations of the CoAlt scores. The purpose of the CoAlt: Science and Social Studies assessments is to provide information about a student's level of mastery of the EEOs of the CAS. In support of that, the previous chapters of this report describe processes that were implemented throughout the CoAlt: Science and Social Studies assessment cycle with validity and fairness considerations in mind; this chapter provides information regarding specific sources of validity evidence as well as fairness. Furthermore, validation is a process. As the CoAlt: Science and Social Studies assessments mature, validity evidence supporting the assessments' interpretations will continue to be collected and documented.

Sources of Validity Evidence

The following sections describe various sources of validity evidence as outlined in the *Standards for Educational and Psychological Testing* (AERA, APA, NCME, 2014).

Evidence Based on Test Content

It is important to examine the extent to which the items on an assessment measure the intended construct. The CoAlt: Science and Social Studies assessments intend to measure the EEOs of the CAS and steps are put in place throughout the development process with focus on this goal, as outlined in Chapter 2 of this report. For example, there are numerous reviews that an item goes through to confirm that it adequately aligns to the EEO that it is intended to measure. In addition, with the field testing of items, statistical bias analyses (i.e., differential item functioning [DIF] analyses) are conducted to identify any items that may be measuring a dimension unrelated to the intended construct. The test blueprints were carefully developed with specificity at multiple levels in an attempt to most optimally measure the EEOs.

In addition to these aforementioned internal processes, a formal alignment study is being planned which will be conducted by a third party.

Evidence Based on Response Processes

Evidence based on response processes pertains to the cognitive aspect behind how students respond to items and the processes by which judges or observers evaluate student performance. As part of the test administration, test examiners were asked a set of questions about students' instruction, their communication modes, and their item responses. These test validity questions can be used to provide validity evidence. One of the test validity questions asked teachers if they believe that student responses accurately reflect their understanding of the material. This question provides evidence as to whether teachers believe that students are actually using their knowledge of the content when responding to the items. The results from this question indicate that the majority of teachers believe that students are using their content knowledge to answer test questions. These results need to be considered in conjunction with the other data related to

the number of hours of instruction in the content area, teacher's familiarity with the content and the student, and the characteristics of the student population.

The test validity question regarding students' receptive and expressive communication methods provides evidence to support the test design and the types of accommodations provided on the assessment. The results from this question indicate that the majority of students use oral administration or picture communication to receive information, and they use these same methods when responding to others. These results help support the validity of the students' responses on the assessment. The complete results from the test validity questions can be found in Part II of this report.

To evaluate that test examiners were administering and scoring the assessment as expected, an inter-rater agreement study was conducted where external observers, called score monitors, visited schools to observe test examiners administering the assessment. The score monitors collected information such as how teachers administered the assessment and provided additional student-level score information that was used to evaluate the consistency of scoring. The results of the inter-rater agreement study can be found in Chapter 9 of this report.

Evidence Based on Internal Structure

The internal structure of an assessment pertains to the degree to which the items on an assessment measure one underlying construct. When assessments are designed to measure one underlying construct, the internal components of the assessments should exhibit a high degree of homogeneity that can be measured in terms of the internal consistency estimates of reliability. As a result, the internal consistency for the CoAlt: Science and Social Studies assessments is evaluated using reliability coefficients. These internal consistency estimates are described in Chapter 9 and provided for the overall test and various student subgroups in Part II of this report.

Evidence Based on Relations to Other Variables

Another measure of validity evidence is the relationship between test performance and performance on another measure, called criterion-related validity. This can be the relationship between two assessments taken at the same time (i.e., concurrent validity) or the relationship between assessments that measure the same or similar construct (i.e. convergent validity) or unrelated constructs (i.e., discriminant validity). Data sources that can be used for criterion-related validity evidence are currently being evaluated for CoAlt.

Evidence for Validity and Consequences of Testing

As the CAS become more fully integrated into the classroom, and with additional administrations of the CoAlt: Science and Social Studies assessments, it is intended that information around the consequences of the assessment will be collected. Some of the intended consequences of the CoAlt: Science and Social Studies assessments include the appropriate use of the assessment for students with the most significant cognitive disabilities, the inclusion of students with the most significant cognitive disabilities in the state assessment system, and the effective instruction of students with the most significant cognitive disabilities in the EEOs of the

CAS. Data regarding the intended and unintended consequences of the CoAlt: Science and Social Studies assessments will be collected and provided when data become available.

Fairness

Fairness is an important aspect of validity, as it is critical that an assessment provide accurate measurements for **all** students. To that end, fairness considerations have been woven into the development and administration of the CoAlt: Science and Social Studies assessments.

Universal Design

The CoAlt: Science and Social Studies development process adheres to the principles of universal design, as described in Chapter 2, with the goal of avoiding construct-irrelevant aspects of the assessment.

Differential Item Functioning

When sample sizes are sufficient, items are analyzed for DIF in order to identify any items that appear to be unfairly favoring one subgroup over another. All DIF-flagged items are then reviewed by assessment specialists to investigate whether there may be a flaw with the item.

Accessibility and Accommodations

As described in Chapters 3 and 4, the CoAlt: Science and Social Studies assessments were developed to be accessible for students with significant cognitive disabilities. In addition to incorporating accessibility into the assessment, accommodations are also available to those students who need additional changes to the test administration in order to access the assessment. The accommodations include assistive technology, braille, eye gaze, modified objects, three-dimensional objects, translation to another language, and other accommodations approved by the state.

Released Items

Because the CoAlt: Science and Social Studies assessments are new to the field, it was necessary for students and teachers to have an opportunity to experience the assessment items prior to the first operational administration. As a result, items were released so that teachers and students would have the opportunity to become familiar with the test design and scoring of the assessments.

PART II: STATISTICAL SUMMARIES

This section contains an overview of the statistical summaries for the Fall 2014 HS and the Spring 2015 ES/MS administrations. Administration summaries, calibration results, performance results, reliability evidence, and validity evidence are included for the operational items. Test form summaries and item performance review outcomes are provided for the embedded field test items.

CHAPTER 1: OPERATIONAL ITEMS

The following section provides high-level details about the CoAlt: Science and Social Studies assessments.

Administration Summary

Approximately 3,000 students took the CoAlt: Science and Social Studies assessments. Tables 7–27 show descriptive statistics for all students and subgroups. The tables include descriptive statistics for the scale scores and raw scores as well as reliability and SEM estimates. Each grade has a mean scale score near 150 and a standard deviation around 40, as expected based on the scaling methodology. The coefficient alpha for the total group across the science and social studies assessments ranged from 0.93 to 0.97. The SEM values for the total group ranged from 3.62 to 4.36.

Calibration Results

Item Statistics

Tables 28–33 contain the classical item statistics. The "Type" column indicates the item type (i.e., selected response item [SR] or supported performance task [SPT]). Columns "% 0" through "% 6" contain the percentage of students at each score point for each operational item, and the "Mean Score" and "Item-Total Corr" columns contain the average score students earned on the item and the correlation between students' total test score and their item score.

Tables 34–39 contain the item parameter estimates for each grade-level assessment. The "Type" column indicates the item type (i.e., selected response item [SR] or supported performance task [SPT]). The "B" column contains the Rasch item difficulty estimates, columns "D1" through "D7" contain the category estimates, and the "Infit" and "Outfit" columns contain the item fit values.

See Chapter 8 for detailed information about the calibration process.

Performance Results

The cuts scores, percent of students in each performance level, and the scale score ranges are provided in Tables 40–41. The scale score distributions for each assessment are shown in Tables 42–47. Tables 48–53 are provided and include the raw score, scale score, and CSEM values.

Decision Consistency and Accuracy

Table 54 provides statistics related to decision consistency and accuracy. The table shows the consistency and accuracy estimates as well as the probabilities due to chance and kappa for all assessments.

Validity Evidence

Test Validity Questions

Before submitting student scores, test examiners responded to survey questions related to student instruction, communication, and test performance. Table 55 provides the summary of teachers' responses to the test validation questions for each assessment.

CHAPTER 2: EMBEDDED FIELD TEST ITEMS

The following section provides details around the field test items that were embedded within the CoAlt: Science and Social Studies assessments.

Field Test Items

Field test items were included on each operational test form. Forty-two field test items were administered across the science and social studies assessments. For those tests with multiple test forms, each test form was parallel; each student received the same number of each item type and in the same location on the form. Table 56 summarizes the number of field test forms and field test items per grade.

Data Review

Student performance data were obtained for all field test items and reviewed to determine if item performance was acceptable for the item to be used on future operational assessments. If any items were flagged for poor performance during the review process, the items would then go to data review to be reviewed by a committee of educators where they would decide whether to accept or reject the item. Only one item was flagged across the 42 field test items. As a result, the flagged item was reviewed by CDE assessment specialists and a data review meeting was not convened. Table 53 shows the outcomes of the item performance review.

REFERENCES

Altman, D. G. (1991). *Practical Statistics for Medical Research.* London, UK: Chapman and Hall/CRC Press.

American Educational Research Association (AERA), American Psychological Association (APA), and the National Council on Measurement in Education (NCME). (2014). *Standards for educational and psychological testing*. Washington, DC: AERA.

Cizek, G. J. (2012). *Setting performance standards: Foundations, methods, and innovations* (2nd ed.). New York, NY: Routledge.

Cizek, G. J., Bunch, M. B., & Koons, H. (2004). Setting Performance Standards: Contemporary Methods. *Educational Measurement: Issues and Practice*, *23*(4), 31–50.

Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory.* New York, NY: Harcourt Brace Jovanovich College Publishers.

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

Ferguson, G. A., & Takane, Y. (1989). *Statistical analysis in psychology and education* (6th ed.). New York, NY: McGraw-Hill.

Hambleton, R. K. & Plake, B.S. (1995). Using an Extended Angoff Procedure to Set Standards on Complex Performance Assessments. *Applied Measurement in Education*, *8*, 41–56.

Linacre, J.M. (2011). Winsteps® (Version 3.73.0) [computer program]. Beaverton, OR: Winsteps.com.

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

Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, *47*(2), 149–174.

Rasch, G. (1966). An individualistic approach to item analysis. In P. Lazarfeld & N. W. Henry (Eds.), *Readings in Mathematical Social Science* (pp. 89–107). Chicago, IL: Science Research Associates.

Roeber, E. (2002). *Setting standards on alternate assessments* (Synthesis Report 42). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.

Wright, B. D., & Masters, G. N. (1982). *Rating scale analysis*. Chicago, IL: MESA Press.

Wright, B.D., & Stone, M.H. (1979). *Best test design*. Chicago, IL: MESA Press.

COALT: SCIENCE AND SOCIAL STUDIES TABLES 7–56

Table 7. Social Studies Descriptive Statistics by Gender and Race/Ethnicity

Content	Grade	Subgroup	N	%	Scale Score				Raw Score				Alpha	SEM
					Mean	SD	Min	Max	Mean	SD	Min	Max		
SS	4	Total	611	100	149.86	34.39	0	235	50.35	14.95	0	71	0.93	4.00
		Female	205	33.55	152.00	31.04	0	235	51.03	14.45	0	71	0.92	4.03
		Male	406	66.45	148.78	35.95	0	216	50.01	15.20	0	70	0.93	3.99
		American Indian	6	0.98	-	-	-	-	-	-	-	-	-	-
		Asian	14	2.29	-	-	-	-	-	-	-	-	-	-
		Black or African American	48	7.86	144.42	48.90	0	235	48.40	19.72	0	71	0.97	3.68
		Hispanic or Latino	235	38.46	150.04	32.38	0	197	50.45	14.48	0	68	0.93	3.95
		White	273	44.68	150.81	32.54	0	216	50.79	14.19	0	70	0.92	4.08
		Native Hawaiian or other Pacific Islander	5	0.82	-	-	-	-	-	-	-	-	-	-
		Two or More Races	22	3.60	153.27	17.13	119	178	51.32	10.21	29	64	0.83	4.25
		Missing	8	1.31	-	-	-	-	-	-	-	-	-	-
	7	Total	634	100	150.98	35.38	0	250	53.60	14.65	0	72	0.93	3.76
		Female	251	39.59	150.15	37.34	0	250	53.60	14.99	0	72	0.94	3.76
		Male	383	60.41	151.53	34.08	0	250	53.59	14.45	0	72	0.93	3.75
		American Indian	4	0.63	-	-	-	-	-	-	-	-	-	-
		Asian	21	3.31	135.81	52.49	0	201	46.81	21.05	0	69	0.97	3.90
		Black or African American	45	7.10	150.69	37.16	0	212	54.31	13.44	0	70	0.93	3.56
		Hispanic or Latino	247	38.96	151.43	33.55	0	212	54.01	13.92	0	70	0.93	3.62
		White	285	44.95	151.60	36.48	0	250	53.57	15.17	0	72	0.94	3.85
		Native Hawaiian or other Pacific Islander	0	0.00	-	-	-	-	-	-	-	-	-	-
		Two or More Races	26	4.10	155.27	18.12	110	191	55.73	9.47	28	67	0.84	3.85
		Missing	6	0.95	-	-	-	-	-	-	-	-	-	-

Table 8. Social Studies Descriptive Statistics by Gender and Race/Ethnicity (continued)

Content	Grade	Subgroup	N	%	Scale Score				Raw Score				Alpha	SEM
					Mean	SD	Min	Max	Mean	SD	Min	Max		
SS	HS	Total	444	100	150.55	37.71	0	250	82.17	26.01	0	110	0.97	4.36
		Female	175	39.41	147.19	36.12	0	200	80.43	25.86	0	107	0.97	4.45
		Male	269	60.59	152.74	38.61	0	250	83.29	26.09	0	110	0.97	4.29
		American Indian	5	1.13	-	-	-	-	-	-	-	-	-	-
		Asian	18	4.05	146.11	32.10	69	200	76.44	27.82	9	107	0.97	4.77
		Black or African American	42	9.46	150.57	49.79	0	250	81.67	31.03	0	110	0.99	3.82
		Hispanic or Latino	138	31.08	146.20	40.86	0	226	79.29	28.37	0	109	0.98	4.42
		White	221	49.77	152.81	35.01	0	226	83.84	24.40	0	109	0.97	4.35
		Native Hawaiian or other Pacific Islander	1	0.23	-	-	-	-	-	-	-	-	-	-
		Two or More Races	7	1.58	-	-	-	-	-	-	-	-	-	-
		Missing	12	2.70	-	-	-	-	-	-	-	-	-	-

Table 9. Science Descriptive Statistics by Gender and Race/Ethnicity

Content	Grade	Subgroup	N	%	Scale Score				Raw Score				Alpha	SEM
					Mean	SD	Min	Max	Mean	SD	Min	Max		
SC	5	Total	689	100	152.92	36.64	0	250	55.39	16.01	0	72	0.95	3.62
		Female	249	36.14	153.98	35.51	0	250	55.79	15.60	0	72	0.95	3.57
		Male	440	63.86	152.33	37.29	0	250	55.16	16.25	0	72	0.95	3.65
		American Indian	5	0.73	-	-	-	-	-	-	-	-	-	-
		Asian	13	1.89	-	-	-	-	-	-	-	-	-	-
		Black or African American	60	8.71	145.02	40.71	0	213	51.87	19.07	0	71	0.96	3.75
		Hispanic or Latino	265	38.46	150.59	35.69	0	250	55.17	16.18	0	72	0.95	3.58
		White	316	45.86	155.99	37.08	0	250	56.25	15.30	0	72	0.95	3.58
		Native Hawaiian or other Pacific Islander	3	0.44	-	-	-	-	-	-	-	-	-	-
		Two or More Races	26	3.77	156.42	29.35	104	250	56.00	13.17	21	72	0.91	4.00
		Missing	1	0.15	-	-	-	-	-	-	-	-	-	-
	8	Total	607	100	152.15	37.67	0	250	84.89	23.17	0	108	0.97	4.12
		Female	235	38.71	151.25	36.85	0	250	84.15	23.56	0	108	0.97	4.19
		Male	372	61.29	152.72	38.22	0	250	85.36	22.94	0	108	0.97	4.08
		American Indian	10	1.65	-	-	-	-	-	-	-	-	-	-
		Asian	13	2.14	-	-	-	-	-	-	-	-	-	-
		Black or African American	40	6.59	163.08	30.92	91	250	90.68	14.17	28	108	0.93	3.87
		Hispanic or Latino	210	34.60	155.60	35.14	0	250	87.33	21.11	0	108	0.97	3.94
		White	302	49.75	148.66	40.31	0	250	82.54	25.31	0	108	0.97	4.26
		Native Hawaiian or other Pacific Islander	1	0.16	-	-	-	-	-	-	-	-	-	-
		Two or More Races	23	3.79	142.35	46.18	0	211	78.83	30.09	0	106	0.98	4.23
		Missing	8	1.32	-	-	-	-	-	-	-	-	-	-

Table 10. Science Descriptive Statistics by Gender and Race/Ethnicity (continued)

Content	Grade	Subgroup	N	%	Scale Score				Raw Score				Alpha	SEM
					Mean	SD	Min	Max	Mean	SD	Min	Max		
SC	HS	Total	446	100	150.31	38.81	0	250	84.52	26.32	0	110	0.97	4.34
		Female	176	39.46	147.96	37.57	0	218	83.71	26.10	0	109	0.97	4.34
		Male	270	60.54	151.84	39.60	0	250	85.04	26.50	0	110	0.97	4.33
		American Indian	5	1.12	-	-	-	-	-	-	-	-	-	-
		Asian	18	4.04	142.83	31.42	60	193	76.06	29.10	5	107	0.97	4.69
		Black or African American	42	9.42	148.55	55.05	0	250	83.64	31.54	0	110	0.99	3.71
		Hispanic or Latino	141	31.61	145.41	42.23	0	202	81.72	29.30	0	108	0.98	4.32
		White	221	49.55	153.65	34.45	0	218	86.69	23.80	0	109	0.97	4.36
		Native Hawaiian or other Pacific Islander	1	0.22	-	-	-	-	-	-	-	-	-	-
		Two or More Races	6	1.35	-	-	-	-	-	-	-	-	-	-
		Missing	12	2.69	-	-	-	-	-	-	-	-	-	-

Table 11. Descriptive Statistics by Free/Reduced Price Lunch Eligibility

Content	Grade	Subgroup	N	%	Scale Score				Raw Score				Alpha	SEM
					Mean	SD	Min	Max	Mean	SD	Min	Max		
SS	4	Free Lunch Eligible	294	48.12	153.18	32.61	0	235	51.84	14.17	0	71	0.93	3.88
		Reduced Lunch Eligible	56	9.17	148.79	30.70	0	191	49.66	14.60	0	67	0.91	4.27
		Not Eligible	261	42.72	146.35	36.76	0	205	48.82	15.74	0	69	0.93	4.07
		Missing	0	0.00	-	-	-	-	-	-	-	-	-	-
	7	Free Lunch Eligible	307	48.42	154.11	31.52	0	212	55.38	12.76	0	70	0.92	3.61
		Reduced Lunch Eligible	61	9.62	145.44	48.37	0	250	51.03	18.70	0	72	0.96	3.59
		Not Eligible	266	41.96	148.64	35.91	0	250	52.12	15.44	0	72	0.94	3.95
		Missing	0	0.00	-	-	-	-	-	-	-	-	-	-
	HS	Free Lunch Eligible	184	41.44	153.51	35.65	0	226	84.65	24.40	0	109	0.97	4.19
		Reduced Lunch Eligible	38	8.56	143.03	46.09	0	250	75.90	31.01	0	110	0.98	4.28
		Not Eligible	211	47.52	148.93	38.57	0	226	80.73	26.78	0	109	0.97	4.53
		Missing	11	2.48	-	-	-	-	-	-	-	-	-	-
SC	5	Free Lunch Eligible	335	48.62	154.72	34.29	0	250	56.67	15.26	0	72	0.95	3.50
		Reduced Lunch Eligible	56	8.13	144.09	38.16	1	196	52.18	17.36	1	70	0.95	3.89
		Not Eligible	298	43.25	152.56	38.72	0	250	54.55	16.48	0	72	0.95	3.70
		Missing	0	0.00	-	-	-	-	-	-	-	-	-	-
	8	Free Lunch Eligible	282	46.46	154.81	36.80	0	250	86.68	22.21	0	108	0.97	3.92
		Reduced Lunch Eligible	52	8.57	161.87	29.75	118	250	90.12	13.22	56	108	0.91	3.91
		Not Eligible	273	44.98	147.55	39.37	0	230	82.04	25.23	0	107	0.97	4.36
		Missing	0	0.00	-	-	-	-	-	-	-	-	-	-
	HS	Free Lunch Eligible	185	41.48	153.91	37.88	0	250	87.56	24.51	0	110	0.97	4.06
		Reduced Lunch Eligible	38	8.52	144.79	44.48	0	250	78.87	30.45	0	110	0.98	4.53
		Not Eligible	212	47.53	148.16	39.33	0	218	82.86	27.32	0	109	0.97	4.46
		Missing	11	2.47	-	-	-	-	-	-	-	-	-	-

Table 12. Social Studies Descriptive Statistics by English Language Proficiency

Content	Grade	Variable	Subgroup	N	%	Scale Score				Raw Score				Alpha	SEM
						Mean	SD	Min	Max	Mean	SD	Min	Max		
SS	4	Language Proficiency	Not Applicable	476	77.91	150.14	35.08	0	235	50.59	14.90	0	71	0.93	4.00
			NEP	109	17.84	145.91	33.49	0	191	48.21	15.80	0	67	0.93	4.10
			LEP	13	2.13	-	-	-	-	-	-	-	-	-	-
			FEP	7	1.15	-	-	-	-	-	-	-	-	-	-
			PHLOTE	0	0.00	-	-	-	-	-	-	-	-	-	-
			FELL	0	0.00	-	-	-	-	-	-	-	-	-	-
			Missing	6	0.98	-	-	-	-	-	-	-	-	-	-
		ELL Program-Bilingual	No	603	98.69	149.63	34.43	0	235	50.27	14.96	0	71	0.93	4.01
			Yes	2	0.33	-	-	-	-	-	-	-	-	-	-
			Re-designated Monitored Y1	0	0.00	-	-	-	-	-	-	-	-	-	-
			Re-designated Monitored Y2	0	0.00	-	-	-	-	-	-	-	-	-	-
			Exited Y3	0	0.00	-	-	-	-	-	-	-	-	-	-
			Parent Choice	0	0.00	-	-	-	-	-	-	-	-	-	-
			Missing	6	0.98	-	-	-	-	-	-	-	-	-	-
		ELL Program-ESL	No	477	78.07	150.13	35.04	0	235	50.58	14.88	0	71	0.93	4.00
			Yes	115	18.82	147.74	33.56	0	191	49.15	15.75	0	67	0.93	4.03
			Re-designated Monitored Y1	2	0.33	-	-	-	-	-	-	-	-	-	-
			Re-designated Monitored Y2	1	0.16	-	-	-	-	-	-	-	-	-	-
			Exited Y3	2	0.33	-	-	-	-	-	-	-	-	-	-
			Parent Choice	7	1.15	-	-	-	-	-	-	-	-	-	-
			Missing	7	1.15	-	-	-	-	-	-	-	-	-	-

Table 13. Social Studies Descriptive Statistics by English Language Proficiency (continued)

Content	Grade	Variable	Subgroup	N	%	Scale Score				Raw Score				Alpha	SEM
						Mean	SD	Min	Max	Mean	SD	Min	Max		
SS	7	Language Proficiency	Not Applicable	474	74.76	151.10	34.33	0	250	53.58	14.43	0	72	0.93	3.78
			NEP	112	17.67	149.85	40.81	0	231	53.45	15.88	0	71	0.95	3.62
			LEP	9	1.42	-	-	-	-	-	-	-	-	-	-
			FEP	24	3.79	148.08	36.63	0	182	53.21	15.15	0	66	0.94	3.72
			PHLOTE	4	0.63	-	-	-	-	-	-	-	-	-	-
			FELL	0	0.00	-	-	-	-	-	-	-	-	-	-
			Missing	11	1.74	-	-	-	-	-	-	-	-	-	-
		ELL Program-Bilingual	No	620	97.79	151.06	35.52	0	250	53.67	14.63	0	72	0.94	3.74
			Yes	1	0.16	-	-	-	-	-	-	-	-	-	-
			Re-designated Monitored Y1	0	0.00	-	-	-	-	-	-	-	-	-	-
			Re-designated Monitored Y2	0	0.00	-	-	-	-	-	-	-	-	-	-
			Exited Y3	3	0.47	-	-	-	-	-	-	-	-	-	-
			Parent Choice	0	0.00	-	-	-	-	-	-	-	-	-	-
			Missing	10	1.58	-	-	-	-	-	-	-	-	-	-
		ELL Program-ESL	No	488	76.97	151.25	34.02	0	250	53.65	14.27	0	72	0.93	3.78
			Yes	110	17.35	153.42	36.04	0	231	54.98	13.68	0	71	0.93	3.59
			Re-designated Monitored Y1	2	0.32	-	-	-	-	-	-	-	-	-	-
			Re-designated Monitored Y2	0	0.00	-	-	-	-	-	-	-	-	-	-
			Exited Y3	14	2.21	-	-	-	-	-	-	-	-	-	-
			Parent Choice	10	1.58	-	-	-	-	-	-	-	-	-	-
			Missing	10	1.58	-	-	-	-	-	-	-	-	-	-

Table 14. Social Studies Descriptive Statistics by English Language Proficiency (continued)

Content	Grade	Variable	Subgroup	N	%	Scale Score				Raw Score				Alpha	SEM
						Mean	SD	Min	Max	Mean	SD	Min	Max		
SS	HS	Language Proficiency	Not Applicable	365	82.21	151.41	37.40	0	250	82.61	25.71	0	110	0.97	4.39
			NEP	47	10.59	142.49	47.78	0	194	77.49	31.58	0	106	0.98	4.07
			LEP	7	1.58	-	-	-	-	-	-	-	-	-	-
			FEP	11	2.48	-	-	-	-	-	-	-	-	-	-
			PHLOTE	2	0.45	-	-	-	-	-	-	-	-	-	-
			FELL	0	0.00	-	-	-	-	-	-	-	-	-	-
			Missing	12	2.70	-	-	-	-	-	-	-	-	-	-
		ELL Program-Bilingual	No	420	94.59	150.92	37.59	0	250	82.49	25.94	0	110	0.97	4.36
			Yes	0	0.00	-	-	-	-	-	-	-	-	-	-
			Re-designated Monitored Y1	0	0.00	-	-	-	-	-	-	-	-	-	-
			Re-designated Monitored Y2	0	0.00	-	-	-	-	-	-	-	-	-	-
			Exited Y3	1	0.23	-	-	-	-	-	-	-	-	-	-
			Parent Choice	0	0.00	-	-	-	-	-	-	-	-	-	-
			Missing	23	5.18	144.87	40.66	0	226	77.39	27.40	0	109	0.98	4.11
		ELL Program-ESL	No	371	83.56	150.94	37.40	0	250	82.14	25.92	0	110	0.97	4.41
			Yes	45	10.14	143.29	47.84	0	194	79.27	31.31	0	106	0.98	3.97
			Re-designated Monitored Y1	0	0.00	-	-	-	-	-	-	-	-	-	-
			Re-designated Monitored Y2	1	0.23	-	-	-	-	-	-	-	-	-	-
			Exited Y3	6	1.35	-	-	-	-	-	-	-	-	-	-
			Parent Choice	9	2.03	-	-	-	-	-	-	-	-	-	-
			Missing	12	2.70	-	-	-	-	-	-	-	-	-	-

Table 15. Science Descriptive Statistics by English Language Proficiency

Content	Grade	Variable	Subgroup	N	%	Scale Score				Raw Score				Alpha	SEM
						Mean	SD	Min	Max	Mean	SD	Min	Max		
SC	5	Language Proficiency	Not Applicable	557	80.84	154.77	35.65	0	250	55.99	15.54	0	72	0.95	3.61
			NEP	104	15.09	146.60	37.67	0	213	53.44	17.16	0	71	0.96	3.63
			LEP	11	1.60	-	-	-	-	-	-	-	-	-	-
			FEP	9	1.31	-	-	-	-	-	-	-	-	-	-
			PHLOTE	2	0.29	-	-	-	-	-	-	-	-	-	-
			FELL	1	0.15	-	-	-	-	-	-	-	-	-	-
			Missing	5	0.73	-	-	-	-	-	-	-	-	-	-
		ELL Program-Bilingual	No	681	98.84	153.10	36.29	0	250	55.46	15.91	0	72	0.95	3.61
			Yes	1	0.15	-	-	-	-	-	-	-	-	-	-
			Re-designated Monitored Y1	0	0.00	-	-	-	-	-	-	-	-	-	-
			Re-designated Monitored Y2	0	0.00	-	-	-	-	-	-	-	-	-	-
			Exited Y3	1	0.15	-	-	-	-	-	-	-	-	-	-
			Parent Choice	0	0.00	-	-	-	-	-	-	-	-	-	-
			Missing	6	0.87	-	-	-	-	-	-	-	-	-	-
		ELL Program-ESL	No	561	81.42	154.37	36.09	0	250	55.85	15.66	0	72	0.95	3.61
			Yes	111	16.11	148.30	36.45	0	213	54.33	16.27	0	71	0.95	3.60
			Re-designated Monitored Y1	1	0.15	-	-	-	-	-	-	-	-	-	-
			Re-designated Monitored Y2	0	0.00	-	-	-	-	-	-	-	-	-	-
			Exited Y3	7	1.02	-	-	-	-	-	-	-	-	-	-
			Parent Choice	3	0.44	-	-	-	-	-	-	-	-	-	-
			Missing	6	0.87	-	-	-	-	-	-	-	-	-	-

Table 16. Social Studies Descriptive Statistics by English Language Proficiency (continued)

Content	Grade	Variable	Subgroup	N	%	Scale Score				Raw Score				Alpha	SEM
						Mean	SD	Min	Max	Mean	SD	Min	Max		
SC	8	Language Proficiency	Not Applicable	469	77.27	150.37	40.13	0	250	83.52	24.73	0	108	0.97	4.19
			NEP	96	15.82	156.60	27.95	0	230	88.98	16.82	0	107	0.95	3.86
			LEP	12	1.98	-	-	-	-	-	-	-	-	-	-
			FEP	14	2.31	-	-	-	-	-	-	-	-	-	-
			PHLOTE	1	0.16	-	-	-	-	-	-	-	-	-	-
			FELL	1	0.16	-	-	-	-	-	-	-	-	-	-
			Missing	14	2.31	-	-	-	-	-	-	-	-	-	-
		ELL Program-Bilingual	No	591	97.36	151.98	37.88	0	250	84.82	23.33	0	108	0.97	4.11
			Yes	1	0.16	-	-	-	-	-	-	-	-	-	-
			Re-designated Monitored Y1	1	0.16	-	-	-	-	-	-	-	-	-	-
			Re-designated Monitored Y2	0	0.00	-	-	-	-	-	-	-	-	-	-
			Exited Y3	1	0.16	-	-	-	-	-	-	-	-	-	-
			Parent Choice	0	0.00	-	-	-	-	-	-	-	-	-	-
			Missing	13	2.14	-	-	-	-	-	-	-	-	-	-
		ELL Program-ESL	No	476	78.42	150.47	39.94	0	250	83.59	24.60	0	108	0.97	4.19
			Yes	98	16.14	157.52	27.10	0	211	89.66	16.51	0	106	0.95	3.76
			Re-designated Monitored Y1	2	0.33	-	-	-	-	-	-	-	-	-	-
			Re-designated Monitored Y2	3	0.49	-	-	-	-	-	-	-	-	-	-
			Exited Y3	5	0.82	-	-	-	-	-	-	-	-	-	-
			Parent Choice	10	1.65	-	-	-	-	-	-	-	-	-	-
			Missing	13	2.14	-	-	-	-	-	-	-	-	-	-

Table 17. Social Studies Descriptive Statistics by English Language Proficiency (continued)

Content	Grade	Variable	Subgroup	N	%	Scale Score				Raw Score				Alpha	SEM
						Mean	SD	Min	Max	Mean	SD	Min	Max		
SC	HS	Language Proficiency	Not Applicable	366	82.06	151.39	39.04	0	250	85.24	26.04	0	110	0.97	4.29
			NEP	47	10.54	141.21	46.54	0	193	78.62	32.24	0	107	0.98	4.18
			LEP	8	1.79	-	-	-	-	-	-	-	-	-	-
			FEP	11	2.47	-	-	-	-	-	-	-	-	-	-
			PHLOTE	2	0.45	-	-	-	-	-	-	-	-	-	-
			FELL	0	0.00	-	-	-	-	-	-	-	-	-	-
			Missing	12	2.69	-	-	-	-	-	-	-	-	-	-
		ELL Program-Bilingual	No	423	94.84	150.77	38.85	0	250	84.92	26.32	0	110	0.97	4.31
			Yes	0	0.00	-	-	-	-	-	-	-	-	-	-
			Re-designated Monitored Y1	0	0.00	-	-	-	-	-	-	-	-	-	-
			Re-designated Monitored Y2	0	0.00	-	-	-	-	-	-	-	-	-	-
			Exited Y3	1	0.22	-	-	-	-	-	-	-	-	-	-
			Parent Choice	0	0.00	-	-	-	-	-	-	-	-	-	-
			Missing	22	4.93	141.86	38.83	0	218	77.23	26.30	0	109	0.97	4.66
		ELL Program-ESL	No	372	83.41	151.13	38.83	0	250	84.96	26.06	0	110	0.97	4.32
			Yes	46	10.31	142.04	46.92	0	193	79.78	32.46	0	107	0.98	4.15
			Re-designated Monitored Y1	0	0.00	-	-	-	-	-	-	-	-	-	-
			Re-designated Monitored Y2	1	0.22	-	-	-	-	-	-	-	-	-	-
			Exited Y3	6	1.35	-	-	-	-	-	-	-	-	-	-
			Parent Choice	9	2.02	-	-	-	-	-	-	-	-	-	-
			Missing	12	2.69	-	-	-	-	-	-	-	-	-	-

Table 18. Social Studies Descriptive Statistics by Primary Disability

Content	Grade	Primary Disability	N	%	Scale Score				Raw Score				Alpha	SEM
					Mean	SD	Min	Max	Mean	SD	Min	Max		
SS	4	Autism	113	18.49	146.15	29.99	0	216	47.89	13.86	0	70	0.91	4.23
		Deaf-Blindness	0	0.00	-	-	-	-	-	-	-	-	-	-
		Developmental Delay	0	0.00	-	-	-	-	-	-	-	-	-	-
		Hearing Impairment	1	0.16	-	-	-	-	-	-	-	-	-	-
		Intellectual Disability	151	24.71	160.84	20.89	0	205	55.87	9.38	0	69	0.86	3.56
		Multiple Disabilities	240	39.28	142.09	40.75	0	235	46.81	17.32	0	71	0.94	4.17
		Orthopedic Impairment	1	0.16	-	-	-	-	-	-	-	-	-	-
		Other Health Impairment	20	3.27	159.05	18.69	104	188	54.60	10.55	20	66	0.87	3.83
		Physical Disability	49	8.02	145.96	43.40	0	197	48.98	17.23	0	68	0.95	4.03
		Emotional Disability	4	0.65	-	-	-	-	-	-	-	-	-	-
		Specific Learning Disability	11	1.80	-	-	-	-	-	-	-	-	-	-
		Speech Impairment	7	1.15	-	-	-	-	-	-	-	-	-	-
		Traumatic Brain Injury	5	0.82	-	-	-	-	-	-	-	-	-	-
		Visual Impairment	2	0.33	-	-	-	-	-	-	-	-	-	-
		None	7	1.15	-	-	-	-	-	-	-	-	-	-

Table 19. Social Studies Descriptive Statistics by Primary Disability (continued)

Content	Grade	Primary Disability	N	%	Scale Score				Raw Score				Alpha	SEM
					Mean	SD	Min	Max	Mean	SD	Min	Max		
SS	7	Autism	98	15.46	149.35	32.60	0	212	52.26	14.31	0	70	0.93	3.89
		Deaf-Blindness	0	0.00	-	-	-	-	-	-	-	-	-	-
		Developmental Delay	0	0.00	-	-	-	-	-	-	-	-	-	-
		Hearing Impairment	0	0.00	-	-	-	-	-	-	-	-	-	-
		Intellectual Disability	188	29.65	162.81	17.33	105	250	59.03	6.49	25	72	0.72	3.44
		Multiple Disabilities	257	40.54	140.05	44.19	0	231	49.16	18.02	0	71	0.95	3.93
		Orthopedic Impairment	3	0.47	-	-	-	-	-	-	-	-	-	-
		Other Health Impairment	14	2.21	-	-	-	-	-	-	-	-	-	-
		Physical Disability	34	5.36	157.00	31.83	34	212	55.44	13.67	3	70	0.93	3.64
		Emotional Disability	3	0.47	-	-	-	-	-	-	-	-	-	-
		Specific Learning Disability	14	2.21	-	-	-	-	-	-	-	-	-	-
		Speech Impairment	3	0.47	-	-	-	-	-	-	-	-	-	-
		Traumatic Brain Injury	12	1.89	-	-	-	-	-	-	-	-	-	-
		Visual Impairment	2	0.32	-	-	-	-	-	-	-	-	-	-
		None	6	0.95	-	-	-	-	-	-	-	-	-	-

Table 20. Social Studies Descriptive Statistics by Primary Disability (continued)

Content	Grade	Primary Disability	N	%	Scale Score				Raw Score				Alpha	SEM
					Mean	SD	Min	Max	Mean	SD	Min	Max		
SS	HS	Autism	47	10.59	149.85	35.98	0	210	79.53	26.56	0	108	0.97	4.48
		Deaf-Blindness	0	0.00	-	-	-	-	-	-	-	-	-	-
		Developmental Delay	0	0.00	-	-	-	-	-	-	-	-	-	-
		Hearing Impairment	4	0.90	-	-	-	-	-	-	-	-	-	-
		Intellectual Disability	128	28.83	160.86	30.13	0	210	90.93	19.16	0	108	0.96	3.85
		Multiple Disabilities	204	45.95	139.12	42.09	0	200	74.00	29.28	0	107	0.97	4.77
		Orthopedic Impairment	1	0.23	-	-	-	-	-	-	-	-	-	-
		Other Health Impairment	5	1.13	-	-	-	-	-	-	-	-	-	-
		Physical Disability	15	3.38	-	-	-	-	-	-	-	-	-	-
		Emotional Disability	5	1.13	-	-	-	-	-	-	-	-	-	-
		Specific Learning Disability	8	1.80	-	-	-	-	-	-	-	-	-	-
		Speech Impairment	4	0.90	-	-	-	-	-	-	-	-	-	-
		Traumatic Brain Injury	6	1.35	-	-	-	-	-	-	-	-	-	-
		Visual Impairment	0	0.00	-	-	-	-	-	-	-	-	-	-
		None	17	3.83	158.35	23.30	110	226	87.18	16.28	36	109	0.95	3.80

Table 21. Science Descriptive Statistics by Primary Disability

Content	Grade	Primary Disability	N	%	Scale Score				Raw Score				Alpha	SEM
					Mean	SD	Min	Max	Mean	SD	Min	Max		
SC	5	Autism	98	14.22	147.81	32.51	0	250	52.24	16.17	0	72	0.94	4.06
		Deaf-Blindness	0	0.00	-	-	-	-	-	-	-	-	-	-
		Developmental Delay	0	0.00	-	-	-	-	-	-	-	-	-	-
		Hearing Impairment	2	0.29	-	-	-	-	-	-	-	-	-	-
		Intellectual Disability	186	27.00	165.98	30.26	0	250	61.65	10.83	0	72	0.92	3.08
		Multiple Disabilities	306	44.41	144.51	39.06	0	250	51.65	17.49	0	72	0.95	3.83
		Orthopedic Impairment	1	0.15	-	-	-	-	-	-	-	-	-	-
		Other Health Impairment	18	2.61	-	-	-	-	-	-	-	-	-	-
		Physical Disability	45	6.53	148.00	38.98	0	187	54.62	17.29	0	69	0.96	3.66
		Emotional Disability	1	0.15	-	-	-	-	-	-	-	-	-	-
		Specific Learning Disability	15	2.18	-	-	-	-	-	-	-	-	-	-
		Speech Impairment	4	0.58	-	-	-	-	-	-	-	-	-	-
		Traumatic Brain Injury	8	1.16	-	-	-	-	-	-	-	-	-	-
		Visual Impairment	1	0.15	-	-	-	-	-	-	-	-	-	-
		None	4	0.58	-	-	-	-	-	-	-	-	-	-

Table 22. Science Descriptive Statistics by Primary Disability (continued)

Content	Grade	Primary Disability	N	%	Scale Score				Raw Score				Alpha	SEM
					Mean	SD	Min	Max	Mean	SD	Min	Max		
SC	8	Autism	89	14.66	147.51	32.66	1	250	80.97	21.29	1	108	0.96	4.47
		Deaf-Blindness	0	0.00	-	-	-	-	-	-	-	-	-	-
		Developmental Delay	0	0.00	-	-	-	-	-	-	-	-	-	-
		Hearing Impairment	0	0.16	-	-	-	-	-	-	-	-	-	-
		Intellectual Disability	201	33.11	166.48	25.79	0	250	93.98	12.15	0	108	0.91	3.58
		Multiple Disabilities	243	40.03	138.26	43.60	0	230	76.66	28.50	0	107	0.98	4.44
		Orthopedic Impairment	2	0.33	-	-	-	-	-	-	-	-	-	-
		Other Health Impairment	10	1.65	-	-	-	-	-	-	-	-	-	-
		Physical Disability	31	5.11	150.26	37.14	0	200	84.10	22.56	0	105	0.96	4.42
		Emotional Disability	3	0.49	-	-	-	-	-	-	-	-	-	-
		Specific Learning Disability	12	1.98	-	-	-	-	-	-	-	-	-	-
		Speech Impairment	3	0.49	-	-	-	-	-	-	-	-	-	-
		Traumatic Brain Injury	3	0.49	-	-	-	-	-	-	-	-	-	-
		Visual Impairment	0	0.00	-	-	-	-	-	-	-	-	-	-
		None	9	1.48	-	-	-	-	-	-	-	-	-	-

Table 23. Science Descriptive Statistics by Primary Disability (continued)

Content	Grade	Primary Disability	N	%	Scale Score				Raw Score				Alpha	SEM
					Mean	SD	Min	Max	Mean	SD	Min	Max		
SC	HS	Autism	45	10.09	152.27	34.28	0	218	83.62	24.38	0	109	0.97	4.32
		Deaf-Blindness	0	0.00	-	-	-	-	-	-	-	-	-	-
		Developmental Delay	0	0.00	-	-	-	-	-	-	-	-	-	-
		Hearing Impairment	4	0.90	-	-	-	-	-	-	-	-	-	-
		Intellectual Disability	130	29.15	161.38	30.59	0	250	93.58	18.60	0	110	0.96	3.70
		Multiple Disabilities	205	45.96	138.51	44.56	0	218	76.28	30.50	0	109	0.98	4.74
		Orthopedic Impairment	1	0.22	-	-	-	-	-	-	-	-	-	-
		Other Health Impairment	5	1.12	-	-	-	-	-	-	-	-	-	-
		Physical Disability	16	3.59	164.69	22.21	115	218	93.81	16.53	40	109	0.95	3.87
		Emotional Disability	5	1.12	-	-	-	-	-	-	-	-	-	-
		Specific Learning Disability	8	1.79	-	-	-	-	-	-	-	-	-	-
		Speech Impairment	4	0.90	-	-	-	-	-	-	-	-	-	-
		Traumatic Brain Injury	6	1.35	-	-	-	-	-	-	-	-	-	-
		Visual Impairment	0	0.00	-	-	-	-	-	-	-	-	-	-
		None	17	3.81	152.12	23.57	109	218	84.12	19.44	33	109	0.94	4.74

Table 24. Social Studies Descriptive Statistics by Accommodation

Content	Grade	Accommodation	Subgroup	N	%	Scale Score				Raw Score				Alpha	SEM
						Mean	SD	Min	Max	Mean	SD	Min	Max		
SS	4	Assistive Technology	No	577	94.44	150.72	33.85	0	235	50.77	14.73	0	71	0.93	3.99
			Yes	34	5.56	135.18	40.25	0	181	43.29	16.99	0	65	0.94	4.29
		Braille	No	610	99.84	149.90	34.40	0	235	50.38	14.94	0	71	0.93	4.00
			Yes	1	0.16	-	-	-	-	-	-	-	-	-	-
		Eye Gaze	No	593	97.05	151.92	31.23	0	235	51.24	13.89	0	71	0.92	4.00
			Yes	18	2.95	82.06	58.68	0	150	21.17	19.23	0	51	0.95	4.13
		Modified Picture Symbols	No	602	98.53	150.01	34.46	0	235	50.47	14.91	0	71	0.93	3.98
			Yes	9	1.47	-	-	-	-	-	-	-	-	-	-
		Objects	No	593	97.05	150.43	34.19	0	235	50.65	14.86	0	71	0.93	3.98
			Yes	18	2.95	131.06	36.53	0	160	40.50	14.77	0	57	0.90	4.77
		Translation into Native Language	No	610	99.84	149.88	34.41	0	235	50.37	14.95	0	71	0.93	4.00
			Yes	1	0.16	-	-	-	-	-	-	-	-	-	-
		Other	No	551	90.18	150.54	33.71	0	216	50.70	14.63	0	70	0.93	3.98
			Yes	60	9.82	143.57	39.85	0	235	47.17	17.42	0	71	0.94	4.17
	7	Assistive Technology	No	598	94.32	152.07	34.01	0	250	54.07	14.17	0	72	0.93	3.74
			Yes	36	5.68	133.00	50.59	0	201	45.67	19.75	0	69	0.96	4.06
		Braille	No	633	99.84	151.03	35.39	0	250	53.62	14.65	0	72	0.93	3.75
			Yes	1	0.16	-	-	-	-	-	-	-	-	-	-
		Eye Gaze	No	613	96.69	153.37	31.66	0	250	54.63	13.20	0	72	0.92	3.72
			Yes	21	3.31	81.38	60.76	0	170	23.52	21.88	0	63	0.96	4.54
		Modified Picture Symbols	No	625	98.58	151.44	34.41	0	250	53.79	14.36	0	72	0.93	3.75
			Yes	9	1.42	-	-	-	-	-	-	-	-	-	-
		Objects	No	627	98.90	151.88	33.85	0	250	53.96	14.13	0	72	0.93	3.74
			Yes	7	1.10	-	-	-	-	-	-	-	-	-	-
		Translation into Native Language	No	632	99.68	151.05	35.40	0	250	53.63	14.64	0	72	0.93	3.76
			Yes	2	0.32	-	-	-	-	-	-	-	-	-	-
		Other	No	571	90.06	151.88	35.02	0	250	54.03	14.32	0	72	0.93	3.73
			Yes	63	9.94	142.84	37.85	0	212	49.65	17.01	0	70	0.95	3.94

Table 25. Social Studies Descriptive Statistics by Accommodation (continued)

Content	Grade	Accommodation	Subgroup	N	%	Scale Score				Raw Score				Alpha	SEM
						Mean	SD	Min	Max	Mean	SD	Min	Max		
SS	HS	Assistive Technology	No	425	95.72	151.61	36.71	0	250	82.92	25.40	0	110	0.97	4.33
			Yes	19	4.28	126.84	51.17	0	184	65.32	33.74	0	104	0.98	4.92
		Braille	No	444	100	150.55	37.71	0	250	82.17	26.01	0	110	0.97	4.36
			Yes	0	0.00	-	-	-	-	-	-	-	-	-	-
		Eye Gaze	No	424	95.50	154.14	32.27	0	250	84.64	22.79	0	110	0.96	4.34
			Yes	20	4.50	74.50	59.96	0	163	29.75	34.33	0	95	0.98	4.32
		Modified Picture Symbols	No	428	96.40	151.39	37.51	0	250	82.95	25.49	0	110	0.97	4.31
			Yes	16	3.60	128.06	37.01	40	171	61.19	31.58	3	99	0.97	5.33
		Objects	No	441	99.32	150.81	37.70	0	250	82.46	25.85	0	110	0.97	4.32
			Yes	3	0.68	-	-	-	-	-	-	-	-	-	-
		Translation into Native Language	No	444	100	150.55	37.71	0	250	82.17	26.01	0	110	0.97	4.36
			Yes	0	0.00	-	-	-	-	-	-	-	-	-	-
		Other	No	404	90.99	150.90	37.41	0	250	82.44	25.94	0	110	0.97	4.29
			Yes	40	9.01	146.98	40.86	0	194	79.45	26.89	0	106	0.97	4.92

Table 26. Science Descriptive Statistics by Accommodation

Content	Grade	Accommodation	Subgroup	N	%	Scale Score				Raw Score				Alpha	SEM
						Mean	SD	Min	Max	Mean	SD	Min	Max		
SC	5	Assistive Technology	No	646	93.76	154.54	36.49	0	250	56.34	15.42	0	72	0.95	3.54
			Yes	43	6.24	128.65	29.94	49	196	41.02	17.92	3	70	0.93	4.62
		Braille	No	685	99.42	153.08	36.61	0	250	55.50	15.91	0	72	0.95	3.62
			Yes	4	0.58	-	-	-	-	-	-	-	-	-	-
		Eye Gaze	No	655	95.07	156.57	31.44	0	250	57.11	13.78	0	72	0.93	3.58
			Yes	34	4.93	82.62	55.14	0	160	22.15	19.60	0	62	0.96	4.00
		Modified Picture Symbols	No	665	96.52	153.96	35.92	0	250	55.91	15.65	0	72	0.95	3.59
			Yes	24	3.48	124.21	44.86	0	187	40.83	19.24	0	69	0.95	4.39
		Objects	No	657	95.36	154.75	34.73	0	250	56.28	15.07	0	72	0.94	3.60
			Yes	32	4.64	115.47	52.44	0	196	37.16	22.79	0	70	0.97	4.05
		Translation into Native Language	No	688	99.85	152.95	36.66	0	250	55.40	16.02	0	72	0.95	3.61
			Yes	1	0.15	-	-	-	-	-	-	-	-	-	-
		Other	No	617	89.55	154.61	35.46	0	250	56.23	15.36	0	72	0.95	3.57
			Yes	72	10.45	138.47	43.16	0	250	48.15	19.41	0	72	0.96	4.01
	8	Assistive Technology	No	562	92.59	154.41	36.11	0	250	86.59	21.46	0	108	0.96	4.05
			Yes	45	7.41	123.91	45.16	0	192	63.44	32.00	0	104	0.98	4.84
		Braille	No	606	99.84	152.21	37.68	0	250	84.94	23.16	0	108	0.97	4.12
			Yes	1	0.16	-	-	-	-	-	-	-	-	-	-
		Eye Gaze	No	587	96.71	153.96	36.11	0	250	86.27	21.66	0	108	0.97	4.07
			Yes	20	3.29	99.00	44.34	0	161	44.50	29.58	0	95	0.97	5.16
		Modified Picture Symbols	No	592	97.53	153.02	37.05	0	250	85.60	22.48	0	108	0.97	4.10
			Yes	15	2.47	-	-	-	-	-	-	-	-	-	-
		Objects	No	594	97.86	153.26	36.63	0	250	85.66	22.35	0	108	0.97	4.12
			Yes	13	2.14	-	-	-	-	-	-	-	-	-	-
		Translation into Native Language	No	606	99.84	152.16	37.70	0	250	84.89	23.19	0	108	0.97	4.12
			Yes	1	0.16	-	-	-	-	-	-	-	-	-	-
		Other	No	556	91.60	153.14	37.29	0	250	85.46	22.75	0	108	0.97	4.09
			Yes	51	8.40	141.37	40.39	1	211	78.75	26.77	1	106	0.97	4.48

Table 27. Science Descriptive Statistics by Accommodation (continued)

Content	Grade	Accommodation	Subgroup	N	%	Scale Score				Raw Score				Alpha	SEM
						Mean	SD	Min	Max	Mean	SD	Min	Max		
SC	HS	Assistive Technology	No	427	95.74	151.31	38.05	0	250	85.27	25.82	0	110	0.97	4.24
			Yes	19	4.26	127.79	49.24	0	182	67.63	32.23	0	105	0.97	5.29
		Braille	No	446	100	150.31	38.81	0	250	84.52	26.32	0	110	0.97	4.34
			Yes	0	0.00	-	-	-	-	-	-	-	-	-	-
		Eye Gaze	No	426	95.52	153.81	33.75	0	250	87.06	23.07	0	110	0.97	4.31
			Yes	20	4.48	75.80	60.71	0	159	30.45	33.08	0	95	0.98	4.53
		Modified Picture Symbols	No	430	96.41	150.95	38.81	0	250	85.10	26.01	0	110	0.97	4.29
			Yes	16	3.59	133.06	35.96	47	175	69.00	30.56	3	103	0.97	5.39
		Objects	No	444	99.55	150.40	38.87	0	250	84.62	26.32	0	110	0.97	4.31
			Yes	2	0.45	-	-	-	-	-	-	-	-	-	-
		Translation into Native Language	No	446	100	150.31	38.81	0	250	84.52	26.32	0	110	0.97	4.34
			Yes	0	0.00	-	-	-	-	-	-	-	-	-	-
		Other	No	404	90.58	150.67	38.82	0	250	84.77	26.41	0	110	0.97	4.27
			Yes	42	9.42	146.81	39.03	0	202	82.12	25.67	0	108	0.96	4.91

Table 28. Grade 4 Social Studies Classical Statistics

ITEM	TYPE	% 0	% 1	% 2	% 3	% 4	% 5	% 6	MEAN SCORE	ITEM-TOTAL CORR
1	SR	5.6	19.6	30.1	20.9	23.7			2.376	0.603
2	SR	5.7	13.6	18.8	19.0	42.9			2.797	0.712
3	SR	5.7	13.6	17.2	23.4	40.1			2.786	0.651
4	SR	5.1	13.4	13.4	13.7	54.3			2.989	0.730
5	SR	5.2	13.4	19.3	26.4	35.7			2.738	0.629
6	SR	5.4	11.9	15.7	20.5	46.5			2.907	0.643
7	SR	7.5	14.1	17.8	31.4	29.1			2.606	0.733
8	SR	6.9	14.9	19.5	14.1	44.7			2.748	0.700
9	SPT	4.3	1.0	1.6	8.8	25.7	41.2	17.3	4.439	0.705
10	SR	6.2	15.4	11.6	13.6	53.2			2.921	0.764
11	SR	7.4	8.0	16.7	17.3	50.6			2.957	0.627
12	SR	5.6	15.1	20.1	18.0	41.2			2.743	0.727
13	SPT	5.7	1.0	3.1	22.3	31.4	24.4	12.1	3.943	0.713
14	SR	5.6	13.9	12.9	16.5	51.1			2.936	0.737
15	SR	6.2	11.9	23.7	25.5	32.6			2.663	0.573
16	SR	6.7	10.6	10.6	13.4	58.6			3.065	0.765
17	SR	6.7	10.6	19.1	29.0	34.5			2.740	0.569

Table 29. Grade 7 Social Studies Classical Statistics

ITEM	TYPE	% 0	% 1	% 2	% 3	% 4	% 5	% 6	MEAN SCORE	ITEM-TOTAL CORR
1	SR	5.2	5.5	10.6	13.9	64.8			3.276	0.691
2	SR	6.3	12.9	27.3	23.3	30.1			2.580	0.609
3	SR	4.3	10.6	19.9	32.5	32.8			2.790	0.618
4	SR	5.4	5.8	11.7	33.8	43.4			3.039	0.678
5	SR	5.4	9.1	18.9	23.8	42.7			2.894	0.707
6	SR	4.4	6.8	13.1	21.9	53.8			3.139	0.698
7	SR	5.4	10.4	18.8	28.7	36.8			2.811	0.642
8	SR	5.7	9.0	14.2	18.3	52.8			3.036	0.729
9	SPT	5.5	1.7	0.9	27.1	30.8	22.4	11.5	3.891	0.725
10	SR	6.5	6.2	7.9	12.5	67.0			3.274	0.783
11	SR	6.0	7.6	15.6	20.5	50.3			3.016	0.691
12	SR	5.8	9.6	15.3	26.2	43.1			2.910	0.690
13	SPT	5.0	0.8	1.3	6.2	12.0	28.9	45.9	4.894	0.777
14	SR	4.9	8.8	19.2	15.1	51.9			3.003	0.671
15	SR	5.5	7.3	9.5	16.1	61.7			3.211	0.775
16	SR	6.2	9.9	21.6	26.0	36.3			2.763	0.669
17	SR	5.8	6.9	14.5	20.3	52.4			3.065	0.723

Table 30. HS Social Studies Classical Statistics

ITEM	TYPE	% 0	% 1	% 2	% 3	% 4	% 5	% 6	MEAN SCORE	ITEM-TOTAL CORR
1	SR	6.8	7.2	16.2	23.9	45.9			2.950	0.784
2	SR	6.1	5.2	14.9	22.3	51.6			3.081	0.659
3	SR	7.0	8.3	15.1	35.8	33.8			2.811	0.733
4	SPT	9.2	1.6	1.4	17.3	18.7	22.1	29.7	4.198	0.806
5	SR	6.8	7.2	17.1	23.2	45.7			2.939	0.773
6	SR	6.8	7.7	21.6	21.2	42.8			2.856	0.705
7	SR	6.3	4.3	6.5	9.7	73.2			3.392	0.851
8	SR	6.3	11.0	27.7	28.6	26.4			2.577	0.694
9	SR	6.3	6.1	16.9	24.1	46.6			2.986	0.815
10	SPT	7.0	2.5	1.6	14.4	27.0	23.0	24.5	4.191	0.810
11	SR	7.0	12.8	27.0	18.9	34.2			2.606	0.610
12	SR	5.9	7.4	12.4	21.6	52.7			3.079	0.756
13	SR	6.3	10.8	27.0	27.9	27.9			2.604	0.695
14	SR	5.9	7.0	11.7	18.5	57.0			3.137	0.809
15	SR	6.1	5.2	9.2	9.7	69.8			3.320	0.819
16	SR	6.3	6.5	11.3	12.2	63.7			3.205	0.805
17	SR	6.5	7.7	18.5	14.6	52.7			2.993	0.801
18	SR	7.0	8.8	16.7	16.2	51.4			2.962	0.794
19	SR	7.0	7.4	24.1	20.3	41.2			2.813	0.663
20	SR	7.2	8.6	16.9	14.2	53.2			2.975	0.809
21	SPT	7.4	0.7	2.3	5.4	8.1	20.0	56.1	4.905	0.875
22	SR	7.9	6.3	9.0	15.3	61.5			3.162	0.824
23	SR	8.1	4.1	13.1	14.4	60.4			3.149	0.725
24	SR	8.6	5.6	8.8	14.6	62.4			3.167	0.831
25	SR	8.3	5.0	16.9	20.7	49.1			2.973	0.744
26	SR	8.1	5.9	11.0	14.4	60.6			3.135	0.823

Table 31. Grade 5 Science Classical Statistics

ITEM	TYPE	% 0	% 1	% 2	% 3	% 4	% 5	% 6	MEAN SCORE	ITEM-TOTAL CORR
1	SR	4.9	10.2	10.7	12.9	61.2			3.154	0.742
2	SR	5.1	15.1	33.2	19.0	27.6			2.489	0.596
3	SR	5.5	4.9	8.7	10.2	70.7			3.356	0.746
4	SPT	4.4	0.7	1.3	6.4	9.6	15.4	62.3	5.113	0.819
5	SR	5.1	8.9	9.0	10.2	66.9			3.250	0.778
6	SR	5.7	13.1	14.7	20.9	45.7			2.880	0.722
7	SR	6.1	10.7	16.5	18.3	48.3			2.920	0.744
8	SR	5.5	13.1	23.5	25.4	32.5			2.663	0.717
9	SR	4.6	11.6	12.0	10.4	61.2			3.120	0.806
10	SR	6.4	14.4	27.9	16.5	34.8			2.591	0.647
11	SR	5.4	12.6	17.3	21.8	43.0			2.843	0.711
12	SR	5.1	6.1	9.9	8.7	70.2			3.329	0.759
13	SR	5.7	6.4	7.5	7.0	73.4			3.361	0.822
14	SPT	4.9	0.9	1.3	6.4	16.7	28.3	41.5	4.800	0.763
15	SR	5.4	7.4	13.9	14.8	58.5			3.136	0.726
16	SR	6.5	9.0	14.5	12.3	57.6			3.055	0.760
17	SR	6.2	6.2	7.7	8.3	71.6			3.327	0.764

Table 32. Grade 8 Science Classical Statistics

ITEM	TYPE	% 0	% 1	% 2	% 3	% 4	% 5	% 6	MEAN SCORE	ITEM-TOTAL CORR
1	SR	5.1	8.9	17.5	21.7	46.8			2.962	0.697
2	SR	5.1	9.6	13.2	13.7	58.5			3.109	0.752
3	SR	4.9	3.3	5.8	6.3	79.7			3.526	0.813
4	SR	4.3	7.1	11.2	16.6	60.8			3.226	0.778
5	SPT	4.8	1.0	1.5	3.8	9.6	19.9	59.5	5.100	0.809
6	SR	4.3	12.4	23.6	27.3	32.5			2.713	0.607
7	SR	4.4	7.6	14.0	27.7	46.3			3.038	0.694
8	SR	4.6	3.5	5.1	6.1	80.7			3.549	0.793
9	SR	4.4	4.4	7.1	11.9	72.2			3.428	0.755
10	SR	4.9	10.7	20.3	27.2	36.9			2.804	0.674
11	SR	5.3	8.1	10.4	8.2	68.0			3.257	0.822
12	SR	5.1	10.9	18.5	27.8	37.7			2.822	0.745
13	SR	4.6	8.9	12.7	10.7	63.1			3.188	0.802
14	SR	4.6	6.9	27.5	25.0	35.9			2.807	0.530
15	SR	4.3	9.4	12.4	21.1	52.9			3.089	0.798
16	SR	4.3	7.2	28.2	36.6	23.7			2.682	0.580
17	SR	4.6	4.8	8.9	12.0	69.7			3.374	0.727
18	SR	4.9	4.9	6.3	8.1	75.8			3.448	0.841
19	SR	4.9	7.9	10.4	14.7	62.1			3.211	0.766
20	SR	4.8	11.2	25.0	21.4	37.6			2.758	0.732
21	SR	5.6	9.4	19.8	26.2	39.0			2.837	0.723
22	SPT	5.6	0.3	0.8	8.6	17.6	25.4	41.7	4.751	0.784
23	SR	4.1	7.4	10.4	13.7	64.4			3.269	0.820
24	SR	4.6	3.8	7.2	10.9	73.5			3.448	0.814
25	SR	4.4	7.7	13.8	15.2	58.8			3.161	0.805
26	SR	4.4	7.2	8.7	9.6	70.0			3.334	0.808

Table 33. HS Science Classical Statistics

ITEM	TYPE	% 0	% 1	% 2	% 3	% 4	% 5	% 6	MEAN SCORE	ITEM-TOTAL CORR
1	SR	4.9	4.5	7.6	7.6	75.3			3.439	0.811
2	SR	6.5	4.0	9.6	8.3	71.5			3.343	0.781
3	SR	6.5	5.4	7.2	8.5	72.4			3.350	0.860
4	SPT	7.0	0.9	1.3	9.4	10.1	30.5	40.4	4.668	0.828
5	SR	5.8	9.6	22.2	18.8	43.5			2.845	0.717
6	SR	6.5	4.0	8.3	11.4	69.7			3.339	0.786
7	SR	7.0	9.4	17.5	26.0	40.1			2.830	0.747
8	SR	7.0	7.8	21.3	21.1	42.8			2.850	0.761
9	SR	7.6	3.4	10.3	14.3	64.3			3.244	0.762
10	SPT	8.5	0.9	1.3	15.5	19.1	19.5	34.8	4.323	0.781
11	SR	7.0	6.1	7.6	9.4	70.0			3.294	0.823
12	SR	7.0	6.1	18.4	27.1	41.5			2.901	0.735
13	SR	7.2	5.4	21.5	31.2	34.8			2.809	0.722
14	SR	6.3	5.4	10.1	13.2	65.0			3.253	0.823
15	SR	6.7	5.8	12.8	20.9	53.8			3.092	0.748
16	SR	7.8	8.5	20.6	19.7	43.3			2.821	0.753
17	SR	6.3	5.8	13.2	24.2	50.4			3.067	0.779
18	SR	7.4	6.7	13.0	20.2	52.7			3.040	0.796
19	SR	6.7	5.8	11.9	13.7	61.9			3.182	0.786
20	SR	8.7	8.5	24.7	24.0	34.1			2.661	0.710
21	SPT	7.0	0.7	1.3	6.3	10.3	15.0	59.0	4.924	0.827
22	SR	8.3	8.7	24.4	22.0	36.5			2.697	0.689
23	SR	8.5	7.2	13.2	13.7	57.4			3.043	0.795
24	SR	9.0	6.3	9.2	7.8	67.7			3.191	0.831
25	SR	8.3	5.2	6.3	7.2	73.1			3.316	0.830
26	SR	8.7	7.0	15.2	14.1	54.9			2.996	0.719

Table 34. Grade 4 Social Studies Item Parameter Estimates

ITEM	TYPE	B	D1	D2	D3	D4	D5	D6	D7	INFIT	OUTFIT
1	SR	0.3001	0	-1.9965	-0.1209	1.0781	1.0393			1.06	1.09
2	SR	-0.0896	0	-1.5816	0.2833	0.8469	0.4514			0.91	0.88
3	SR	-0.0745	0	-1.7552	0.3614	0.6682	0.7256			1.05	1.07
4	SR	-0.2883	0	-1.5743	0.3035	1.3227	-0.0519			0.94	0.83
5	SR	-0.1462	0	-1.8403	-0.1310	0.8772	1.0941			1.16	1.08
6	SR	-0.1537	0	-1.4792	-0.1084	1.1001	0.4874			1.12	1.03
7	SR	0.1440	0	-1.4442	0.0042	0.3238	1.1162			0.87	0.85
8	SR	0.0360	0	-1.4833	0.2732	1.2088	0.0012			0.91	0.90
9	SPT	-0.3582	0	-0.9166	-1.0555	-1.3048	-0.0936	0.8590	2.5115	1.17	1.19
10	SR	-0.1134	0	-1.6260	0.9086	0.8261	-0.1087			0.78	0.68
11	SR	-0.0570	0	-1.0456	-0.0434	0.7401	0.3490			1.19	1.30
12	SR	-0.0156	0	-1.7108	0.2341	1.5167	-0.0399			0.78	0.75
13	SPT	0.3472	0	-0.1932	-0.2591	-2.5586	-0.0418	1.2165	1.8363	1.13	1.27
14	SR	-0.1037	0	-1.1420	0.2453	0.7699	0.1268			0.91	0.81
15	SR	0.0543	0	-1.5601	-0.1821	0.8104	0.9318			1.22	1.18
16	SR	-0.1663	0	-1.1550	0.6082	0.7516	-0.2047			0.84	0.75
17	SR	0.0549	0	-1.3119	-0.1083	0.4443	0.9758			1.27	1.21

Table 35. Grade 7 Social Studies Item Parameter Estimates

ITEM	TYPE	B	D1	D2	D3	D4	D5	D6	D7	INFIT	OUTFIT
1	SR	-0.2904	0	-0.5379	-0.9258	1.1054	0.3583			1.27	1.12
2	SR	0.4368	0	-1.6058	-0.2780	0.9952	0.8886			1.12	1.14
3	SR	-0.1260	0	-2.2184	-0.1154	0.8068	1.5270			1.21	1.24
4	SR	-0.1419	0	-1.3578	-0.2732	0.3866	1.2444			1.14	1.04
5	SR	0.1143	0	-1.5146	-0.0957	0.8249	0.7855			0.92	0.89
6	SR	-0.3251	0	-1.5864	-0.0575	1.0020	0.6418			1.11	1.03
7	SR	0.1861	0	-1.6718	0.0063	0.5858	1.0797			1.10	1.11
8	SR	0.0097	0	-1.2601	-0.2543	1.1077	0.4066			0.97	0.92
9	SPT	0.7132	0	0.8171	-0.8777	-3.0613	0.4725	1.2452	1.4041	0.95	1.71
10	SR	-0.0265	0	-0.6665	0.3661	0.5970	-0.2966			0.85	0.83
11	SR	0.0657	0	-0.9411	-0.5618	0.9948	0.5081			1.08	0.99
12	SR	0.1286	0	-1.0107	-0.4662	0.7297	0.7472			1.06	1.06
13	SPT	-0.1091	0	0.1479	-0.6266	-1.1881	0.2070	0.3651	1.0947	1.12	1.29
14	SR	0.1696	0	-1.2830	-0.6076	1.4417	0.4489			1.06	1.01
15	SR	-0.0864	0	-1.1458	0.4154	0.6056	0.1248			0.85	0.77
16	SR	0.3000	0	-1.3652	-0.2564	0.7290	0.8927			1.01	1.07
17	SR	-0.1001	0	-1.2131	-0.1457	0.6764	0.6824			1.08	1.04

Table 36. HS Social Studies Item Parameter Estimates

ITEM	TYPE	B	D1	D2	D3	D4	D5	D6	D7	INFIT	OUTFIT
1	SR	0.0073	0	-1.3871	-0.2899	0.7515	0.9256			0.90	0.91
2	SR	-0.2230	0	-1.3103	-0.4863	0.8774	0.9192			1.41	1.51
3	SR	0.1752	0	-1.5444	-0.1628	0.1691	1.5381			1.06	1.18
4	SPT	0.4505	0	0.1316	-0.1874	-2.3311	0.5476	0.8034	1.0359	1.08	1.08
5	SR	0.0151	0	-1.3889	-0.3429	0.8331	0.8988			0.95	0.89
6	SR	0.0796	0	-1.4641	-0.5195	1.1390	0.8447			1.12	1.04
7	SR	-0.4774	0	-0.9100	0.1461	0.9326	-0.1687			0.85	0.66
8	SR	0.2922	0	-2.0855	-0.4684	1.0039	1.5500			1.06	1.08
9	SR	-0.0962	0	-1.3965	-0.4757	0.8596	1.0125			0.81	0.78
10	SPT	0.2746	0	-1.0186	0.0003	-1.9334	0.1454	1.3240	1.4823	1.03	1.02
11	SR	0.3211	0	-1.9028	-0.2690	1.3586	0.8132			1.43	1.60
12	SR	-0.2125	0	-1.6937	0.1029	0.7200	0.8708			1.06	0.92
13	SR	0.2650	0	-2.0494	-0.4504	1.0181	1.4816			1.07	1.03
14	SR	-0.2686	0	-1.6104	0.1082	0.8542	0.6480			0.88	0.88
15	SR	-0.4163	0	-1.1851	0.0334	1.2669	-0.1152			0.95	0.77
16	SR	-0.2568	0	-1.3354	0.0462	1.2123	0.0769			0.96	0.82
17	SR	-0.0546	0	-1.4626	-0.2977	1.4216	0.3388			0.83	0.72
18	SR	0.0377	0	-1.4428	-0.0924	1.1312	0.4040			0.85	0.76
19	SR	0.1334	0	-1.3810	-0.6800	1.2585	0.8025			1.29	1.30
20	SR	0.0499	0	-1.3417	-0.1509	1.2896	0.2029			0.81	0.70
21	SPT	-0.1385	0	0.6949	-1.5436	-0.5612	0.4266	0.3501	0.6332	0.98	1.02
22	SR	-0.0427	0	-0.8419	0.1475	0.5393	0.1551			0.88	0.79
23	SR	-0.0419	0	-0.3937	-0.6908	0.9840	0.1005			1.28	1.31
24	SR	0.0071	0	-0.5726	0.0288	0.5224	0.0214			0.84	0.75
25	SR	0.1272	0	-0.5952	-0.7900	0.8048	0.5803			1.10	1.03
26	SR	-0.0075	0	-0.7237	-0.1406	0.7980	0.0663			0.87	0.66

Table 37. Grade 5 Science Item Parameter Estimates

ITEM	TYPE	B	D1	D2	D3	D4	D5	D6	D7	INFIT	OUTFIT
1	SR	-0.2195	0	-1.5522	0.5599	0.9397	0.0526			1.00	0.83
2	SR	0.4215	0	-1.8743	-0.5880	1.3082	1.1541			1.25	1.26
3	SR	-0.2893	0	-0.2463	-0.2520	1.2401	-0.7418			1.04	1.17
4	SPT	-0.4915	0	0.1111	-0.7997	-1.1649	0.5111	0.8888	0.4537	1.07	1.01
5	SR	-0.2553	0	-0.9204	0.1694	0.9878	-0.2369			0.97	0.71
6	SR	0.1763	0	-1.5482	0.1721	0.7220	0.6541			0.98	0.96
7	SR	0.0775	0	-1.3482	0.0350	0.8534	0.4598			0.94	0.91
8	SR	0.2139	0	-1.4884	-0.4438	0.8203	1.1118			0.97	0.99
9	SR	-0.0989	0	-1.3749	0.3629	1.0028	0.0092			0.74	0.64
10	SR	0.2718	0	-1.6661	-0.1360	1.2025	0.5996			1.14	1.20
11	SR	0.1026	0	-1.5396	-0.1958	0.8134	0.9220			1.05	1.04
12	SR	-0.3697	0	-0.9984	0.1352	1.2835	-0.4203			1.01	0.83
13	SR	-0.2707	0	-0.5728	0.0422	1.4977	-0.9670			0.73	0.71
14	SPT	-0.0044	0	0.8846	0.0625	-2.5743	0.0117	0.4382	1.1773	1.13	1.60
15	SR	-0.1735	0	-1.1487	-0.0741	1.0205	0.2024			1.05	1.13
16	SR	0.0040	0	-1.0130	0.0038	1.1205	-0.1113			0.93	0.88
17	SR	-0.2365	0	-0.6600	0.3277	0.9608	-0.6285			1.02	0.73

Table 38. Grade 8 Science Item Parameter Estimates

ITEM	TYPE	B	D1	D2	D3	D4	D5	D6	D7	INFIT	OUTFIT
1	SR	0.1529	0	-1.5952	-0.1173	0.8843	0.8282			1.11	1.08
2	SR	-0.0474	0	-1.7711	0.3571	1.1451	0.2689			0.97	1.00
3	SR	-0.3603	0	-0.8141	0.0398	1.3200	-0.5457			0.80	0.47
4	SR	-0.2696	0	-1.2724	-0.1535	0.9375	0.4884			1.03	0.94
5	SPT	-0.1791	0	-0.0477	-0.5233	-0.5743	-0.1095	0.5550	0.6998	1.16	1.42
6	SR	0.4834	0	-1.8358	-0.6399	1.1522	1.3235			1.27	1.26
7	SR	-0.0138	0	-1.7565	0.0068	0.5319	1.2178			1.13	1.08
8	SR	-0.4870	0	-0.7415	0.2923	1.1318	-0.6826			0.97	1.05
9	SR	-0.3877	0	-1.1085	0.2298	0.8021	0.0766			1.07	0.95
10	SR	0.3747	0	-1.4519	-0.0941	0.7219	0.8241			1.03	1.05
11	SR	-0.0618	0	-1.2927	0.3729	1.3843	-0.4645			0.76	0.62
12	SR	0.3385	0	-1.4944	-0.3164	0.5783	1.2326			0.91	0.86
13	SR	-0.0994	0	-1.6860	0.3318	1.4028	-0.0486			0.79	0.68
14	SR	0.2510	0	-1.8418	-0.7948	1.2617	1.3750			1.52	1.51
15	SR	-0.0101	0	-1.5539	0.3304	0.5312	0.6924			0.80	0.76
16	SR	0.4331	0	-1.6492	-1.0314	0.5309	2.1497			1.35	1.29
17	SR	-0.2993	0	-1.1008	0.0537	0.9777	0.0694			1.14	1.40
18	SR	-0.2992	0	-1.2220	0.5239	1.1157	-0.4176			0.71	0.42
19	SR	-0.1110	0	-1.1627	-0.0987	1.2320	0.0294			1.01	0.84
20	SR	0.3557	0	-1.4858	-0.5260	1.0255	0.9862			0.90	0.87
21	SR	0.3207	0	-1.5338	-0.2799	0.7203	1.0935			0.99	0.96
22	SPT	0.1280	0	1.2751	-1.1672	-2.0967	0.0007	0.8219	1.1661	1.14	1.63
23	SR	-0.2471	0	-1.4268	0.1288	0.9212	0.3767			0.83	0.63
24	SR	-0.1956	0	-0.1981	-0.1417	1.0274	-0.6876			0.81	0.65
25	SR	0.0809	0	-1.0553	-0.0694	0.8903	0.2344			0.74	0.71
26	SR	-0.2608	0	-1.5439	0.5297	1.1976	-0.1835			0.85	0.79

Table 39. HS Science Item Parameter Estimates

ITEM	TYPE	B	D1	D2	D3	D4	D5	D6	D7	INFIT	OUTFIT
1	SR	-0.7904	0	-1.8221	0.3446	1.6074	-0.1299			1.02	0.99
2	SR	-0.3335	0	-0.7567	-0.2373	1.3985	-0.4045			1.11	0.91
3	SR	-0.3202	0	-1.0107	0.3539	1.0652	-0.4084			0.77	0.61
4	SPT	0.1065	0	0.4143	-0.6617	-1.6972	0.6486	0.0209	1.2751	1.10	1.15
5	SR	0.0002	0	-2.0325	-0.1651	1.3847	0.8130			1.02	0.94
6	SR	-0.3319	0	-0.7692	-0.0994	0.9026	-0.0341			1.11	0.88
7	SR	0.1875	0	-1.5626	-0.1097	0.6384	1.0339			1.00	1.02
8	SR	0.1575	0	-1.3815	-0.4844	1.0796	0.7863			0.94	1.01
9	SR	-0.1131	0	-0.2374	-0.6019	0.7643	0.0750			1.13	1.30
10	SPT	0.4144	0	0.7966	-0.6899	-2.2420	0.3977	0.9633	0.7744	1.16	1.23
11	SR	-0.1972	0	-0.9637	0.3438	0.9841	-0.3642			0.91	0.80
12	SR	0.1118	0	-1.1583	-0.6103	0.6755	1.0930			1.06	1.05
13	SR	0.2169	0	-1.0368	-0.9519	0.6237	1.3649			1.06	1.02
14	SR	-0.2761	0	-1.1544	0.0227	0.9848	0.1469			0.89	0.72
15	SR	-0.0731	0	-1.1134	-0.2119	0.6606	0.6648			1.10	1.03
16	SR	0.2705	0	-1.1791	-0.4226	1.0066	0.5951			0.96	0.92
17	SR	-0.1199	0	-1.3246	-0.2161	0.5995	0.9412			0.96	0.90
18	SR	0.0565	0	-1.0219	-0.1391	0.6189	0.5421			0.90	0.86
19	SR	-0.1422	0	-1.0612	-0.1077	1.0196	0.1493			1.00	0.93
20	SR	0.4691	0	-1.0469	-0.7150	0.8605	0.9014			1.09	1.13
21	SPT	-0.0563	0	0.8246	-0.8635	-1.2421	0.2564	0.8023	0.2223	1.23	1.06
22	SR	0.4037	0	-1.1485	-0.6345	0.9901	0.7929			1.11	1.12
23	SR	0.1579	0	-0.7580	-0.1354	0.9368	-0.0434			0.87	0.71
24	SR	0.0783	0	-0.4765	0.1000	1.1304	-0.7540			0.81	0.68
25	SR	-0.0846	0	-0.4107	0.3177	0.9085	-0.8155			0.96	0.54
26	SR	0.2075	0	-0.6948	-0.3295	1.0211	0.0032			1.17	1.26

Table 40. Cut Scores and Students in Each Performance Level

Content	Grade	Cut Scores			Performance Levels									
		Emerging	Developing	Novice	Exploring		Emerging		Developing		Novice		Developing and Novice Combined	
					N	%	N	%	N	%	N	%	N	%
SS	4	46	58	66	145	24	247	40	177	29	42	7	219	36
	7	45	60	67	90	14	288	45	211	33	45	7	256	40
	HS	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
SC	5	47	62	68	138	20	220	32	216	31	115	17	331	48
	8	68	96	103	92	15	268	44	171	28	76	13	247	41
	HS	77	98	107	96	22	185	42	144	32	21	5	165	37

Table 41. Scale Score Ranges for Each Performance Level

	Exploring Level	Emerging Level	Developing Level	Novice Level
Grade 4 Social Studies	0–142	143–162	163–187	188–250
Grade 7 Social Studies	0–133	134–162	163–190	191–250
HS Social Studies	NA	NA	NA	NA
Grade 5 Science	0–134	135–159	160–183	184–250
Grade 8 Science	0–127	128–163	164–189	190–250
HS Science	0–139	140–163	164–192	193–250

Table 42. Grade 4 Social Studies Scale Score Frequency Distributions

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	13	2.13	13	2.13
1	4	0.65	17	2.78
38	2	0.33	19	3.11
47	1	0.16	20	3.27
61	3	0.49	23	3.76
70	1	0.16	24	3.93
74	2	0.33	26	4.26
84	4	0.65	30	4.91
87	1	0.16	31	5.07
90	2	0.33	33	5.40
92	1	0.16	34	5.56
95	2	0.33	36	5.89
97	3	0.49	39	6.38
101	3	0.49	42	6.87
104	1	0.16	43	7.04
107	1	0.16	44	7.20
113	1	0.16	45	7.36
115	2	0.33	47	7.69
116	3	0.49	50	8.18
119	3	0.49	53	8.67
123	2	0.33	55	9.00
124	6	0.98	61	9.98
125	5	0.82	66	10.80
127	6	0.98	72	11.78
128	2	0.33	74	12.11
130	4	0.65	78	12.77
131	6	0.98	84	13.75
133	3	0.49	87	14.24
134	7	1.15	94	15.38

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
135	6	0.98	100	16.37
137	10	1.64	110	18.00
138	8	1.31	118	19.31
139	8	1.31	126	20.62
141	10	1.64	136	22.26
142	9	1.47	145	23.73
143	16	2.62	161	26.35
145	9	1.47	170	27.82
146	17	2.78	187	30.61
147	20	3.27	207	33.88
149	12	1.96	219	35.84
150	15	2.45	234	38.30
152	25	4.09	259	42.39
154	32	5.24	291	47.63
155	17	2.78	308	50.41
157	31	5.07	339	55.48
159	17	2.78	356	58.27
160	36	5.89	392	64.16
163	31	5.07	423	69.23
164	26	4.26	449	73.49
167	25	4.09	474	77.58
169	23	3.76	497	81.34
172	23	3.76	520	85.11
175	21	3.44	541	88.54

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
178	15	2.45	556	91.00
181	13	2.13	569	93.13
188	14	2.29	583	95.42
191	13	2.13	596	97.55
197	9	1.47	605	99.02
205	4	0.65	609	99.67
216	1	0.16	610	99.84
235	1	0.16	611	100.00

Table 43. Grade 7 Social Studies Scale Score Frequency Distributions

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	17	2.68	17	2.68
34	4	0.63	21	3.31
43	2	0.32	23	3.63
56	1	0.16	24	3.79
66	1	0.16	25	3.94
79	1	0.16	26	4.10
85	1	0.16	27	4.26
89	2	0.32	29	4.57
91	1	0.16	30	4.73
93	1	0.16	31	4.89
95	1	0.16	32	5.05
97	1	0.16	33	5.21
100	1	0.16	34	5.36
102	3	0.47	37	5.84
104	3	0.47	40	6.31
105	2	0.32	42	6.62
107	2	0.32	44	6.94
108	1	0.16	45	7.10
110	3	0.47	48	7.57
111	3	0.47	51	8.04
113	2	0.32	53	8.36
114	1	0.16	54	8.52
115	2	0.32	56	8.83
117	2	0.32	58	9.15
118	3	0.47	61	9.62
120	2	0.32	63	9.94
121	1	0.16	64	10.09
123	2	0.32	66	10.41
124	4	0.63	70	11.04

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
125	5	0.79	75	11.83
127	3	0.47	78	12.30
128	2	0.32	80	12.62
130	2	0.32	82	12.93
131	4	0.63	86	13.56
133	4	0.63	90	14.20
134	9	1.42	99	15.62
136	8	1.26	107	16.88
137	9	1.42	116	18.30
139	10	1.58	126	19.87
140	10	1.58	136	21.45
142	15	2.37	151	23.82
144	8	1.26	159	25.08
145	16	2.52	175	27.60
147	20	3.15	195	30.76
149	23	3.63	218	34.38
151	29	4.57	247	38.96
153	31	4.89	278	43.85
155	36	5.68	314	49.53
157	42	6.62	356	56.15
159	22	3.47	378	59.62
163	39	6.15	417	65.77
164	31	4.89	448	70.66
167	32	5.05	480	75.71
170	36	5.68	516	81.39

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
174	28	4.42	544	85.80
177	28	4.42	572	90.22
182	17	2.68	589	92.90
191	20	3.15	609	96.06
193	6	0.95	615	97.00
201	9	1.42	624	98.42
212	7	1.10	631	99.53
231	1	0.16	632	99.68
250	2	0.32	634	100.00

Table 44. HS Social Studies Scale Score Frequency Distributions

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	14	3.15	14	3.15
29	1	0.23	15	3.38
40	1	0.23	16	3.60
53	1	0.23	17	3.83
58	1	0.23	18	4.05
62	1	0.23	19	4.28
66	1	0.23	20	4.50
69	1	0.23	21	4.73
72	1	0.23	22	4.95
79	2	0.45	24	5.41
85	2	0.45	26	5.86
87	1	0.23	27	6.08
89	1	0.23	28	6.31
93	2	0.45	30	6.76
96	1	0.23	31	6.98
101	1	0.23	32	7.21
102	2	0.45	34	7.66
103	1	0.23	35	7.88
104	2	0.45	37	8.33
106	1	0.23	38	8.56
110	1	0.23	39	8.78
111	1	0.23	40	9.01
114	2	0.45	42	9.46
116	1	0.23	43	9.68
118	1	0.23	44	9.91
119	1	0.23	45	10.14
120	3	0.68	48	10.81
121	2	0.45	50	11.26
123	5	1.13	55	12.39
124	2	0.45	57	12.84

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
126	1	0.23	58	13.06
127	1	0.23	59	13.29
128	1	0.23	60	13.51
129	1	0.23	61	13.74
130	2	0.45	63	14.19
131	1	0.23	64	14.41
132	2	0.45	66	14.86
133	2	0.45	68	15.32
134	3	0.68	71	15.99
135	2	0.45	73	16.44
136	3	0.68	76	17.12
137	6	1.35	82	18.47
138	7	1.58	89	20.05
139	8	1.80	97	21.85
140	3	0.68	100	22.52
141	3	0.68	103	23.20
142	3	0.68	106	23.87
143	4	0.90	110	24.77
144	5	1.13	115	25.90
145	10	2.25	125	28.15
146	6	1.35	131	29.50
147	13	2.93	144	32.43
148	5	1.13	149	33.56
150	11	2.48	160	36.04
151	7	1.58	167	37.61
152	9	2.03	176	39.64

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
153	8	1.80	184	41.44
154	18	4.05	202	45.50
155	8	1.80	210	47.30
156	9	2.03	219	49.32
158	12	2.70	231	52.03
159	11	2.48	242	54.50
160	14	3.15	256	57.66
162	14	3.15	270	60.81
163	7	1.58	277	62.39
165	18	4.05	295	66.44
167	17	3.83	312	70.27
169	15	3.38	327	73.65
171	24	5.41	351	79.05
173	15	3.38	366	82.43
175	12	2.70	378	85.14
178	17	3.83	395	88.96
181	6	1.35	401	90.32
184	10	2.25	411	92.57
189	7	1.58	418	94.14
194	13	2.93	431	97.07
200	8	1.80	439	98.87
210	2	0.45	441	99.32
226	2	0.45	443	99.77
250	1	0.23	444	100.00

Table 45. Grade 5 Science Scale Score Frequency Distributions

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	12	1.74	12	1.74
1	4	0.58	16	2.32
38	2	0.29	18	2.61
49	1	0.15	19	2.76
57	2	0.29	21	3.05
63	1	0.15	22	3.19
73	1	0.15	23	3.34
77	1	0.15	24	3.48
80	2	0.29	26	3.77
86	3	0.44	29	4.21
88	3	0.44	32	4.64
90	2	0.29	34	4.93
98	1	0.15	35	5.08
100	1	0.15	36	5.22
101	1	0.15	37	5.37
103	3	0.44	40	5.81
104	2	0.29	42	6.10
107	2	0.29	44	6.39
108	1	0.15	45	6.53
110	1	0.15	46	6.68
111	4	0.58	50	7.26
114	1	0.15	51	7.40
115	3	0.44	54	7.84
116	2	0.29	56	8.13
117	4	0.58	60	8.71
118	2	0.29	62	9.00
119	1	0.15	63	9.14
122	6	0.87	69	10.01
123	5	0.73	74	10.74
124	6	0.87	80	11.61

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
125	3	0.44	83	12.05
126	2	0.29	85	12.34
127	9	1.31	94	13.64
128	7	1.02	101	14.66
130	5	0.73	106	15.38
131	4	0.58	110	15.97
132	8	1.16	118	17.13
133	9	1.31	127	18.43
134	11	1.60	138	20.03
135	7	1.02	145	21.04
137	10	1.45	155	22.50
138	11	1.60	166	24.09
139	10	1.45	176	25.54
140	9	1.31	185	26.85
142	6	0.87	191	27.72
143	10	1.45	201	29.17
145	18	2.61	219	31.79
146	16	2.32	235	34.11
148	11	1.60	246	35.70
149	16	2.32	262	38.03
151	26	3.77	288	41.80
153	25	3.63	313	45.43
155	22	3.19	335	48.62
157	23	3.34	358	51.96
160	38	5.52	396	57.47
162	32	4.64	428	62.12

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
164	39	5.66	467	67.78
167	32	4.64	499	72.42
171	32	4.64	531	77.07
175	43	6.24	574	83.31
184	34	4.93	608	88.24
187	31	4.50	639	92.74
196	25	3.63	664	96.37
213	14	2.03	678	98.40
250	11	1.60	689	100.00

Table 46. Grade 8 Science Scale Score Frequency Distributions

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	13	2.14	13	2.14
1	1	0.16	14	2.31
8	2	0.33	16	2.64
29	1	0.16	17	2.80
35	1	0.16	18	2.97
41	1	0.16	19	3.13
46	1	0.16	20	3.29
60	2	0.33	22	3.62
66	1	0.16	23	3.79
68	1	0.16	24	3.95
83	1	0.16	25	4.12
85	1	0.16	26	4.28
88	1	0.16	27	4.45
91	1	0.16	28	4.61
93	1	0.16	29	4.78
94	1	0.16	30	4.94
96	1	0.16	31	5.11
98	1	0.16	32	5.27
101	1	0.16	33	5.44
102	3	0.49	36	5.93
103	1	0.16	37	6.10
104	1	0.16	38	6.26
106	2	0.33	40	6.59
107	3	0.49	43	7.08
108	4	0.66	47	7.74
110	1	0.16	48	7.91
111	1	0.16	49	8.07
112	3	0.49	52	8.57
113	1	0.16	53	8.73
114	2	0.33	55	9.06

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
116	1	0.16	56	9.23
117	2	0.33	58	9.56
118	5	0.82	63	10.38
119	2	0.33	65	10.71
120	5	0.82	70	11.53
121	3	0.49	73	12.03
122	3	0.49	76	12.52
123	2	0.33	78	12.85
124	1	0.16	79	13.01
125	1	0.16	80	13.18
126	7	1.15	87	14.33
127	5	0.82	92	15.16
128	1	0.16	93	15.32
129	6	0.99	99	16.31
130	8	1.32	107	17.63
131	4	0.66	111	18.29
132	7	1.15	118	19.44
133	4	0.66	122	20.10
134	2	0.33	124	20.43
135	3	0.49	127	20.92
136	9	1.48	136	22.41
137	5	0.82	141	23.23
138	4	0.66	145	23.89
139	11	1.81	156	25.70
140	8	1.32	164	27.02
142	7	1.15	171	28.17
143	9	1.48	180	29.65
144	7	1.15	187	30.81
145	11	1.81	198	32.62
146	8	1.32	206	33.94
148	8	1.32	214	35.26

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
149	11	1.81	225	37.07
151	17	2.80	242	39.87
152	16	2.64	258	42.50
154	22	3.62	280	46.13
155	17	2.80	297	48.93
157	15	2.47	312	51.40
159	20	3.29	332	54.70
161	28	4.61	360	59.31
164	27	4.45	387	63.76
165	33	5.44	420	69.19
168	19	3.13	439	72.32
171	20	3.29	459	75.62
174	27	4.45	486	80.07
178	25	4.12	511	84.18
182	20	3.29	531	87.48
190	26	4.28	557	91.76
192	19	3.13	576	94.89
200	8	1.32	584	96.21
211	14	2.31	598	98.52
230	5	0.82	603	99.34
250	4	0.66	607	100.00

Table 47. HS Science Scale Score Frequency Distributions

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	17	3.81	17	3.81
1	1	0.22	18	4.04
47	1	0.22	19	4.26
55	1	0.22	20	4.48
60	1	0.22	21	4.71
78	2	0.45	23	5.16
89	2	0.45	25	5.61
91	1	0.22	26	5.83
100	2	0.45	28	6.28
102	1	0.22	29	6.50
105	1	0.22	30	6.73
106	1	0.22	31	6.95
109	1	0.22	32	7.17
111	2	0.45	34	7.62
112	2	0.45	36	8.07
114	1	0.22	37	8.30
115	3	0.67	40	8.97
116	2	0.45	42	9.42
119	1	0.22	43	9.64
120	2	0.45	45	10.09
121	1	0.22	46	10.31
122	4	0.90	50	11.21
124	3	0.67	53	11.88
125	4	0.90	57	12.78
126	2	0.45	59	13.23
127	1	0.22	60	13.45
128	1	0.22	61	13.68
129	1	0.22	62	13.90
130	2	0.45	64	14.35
131	3	0.67	67	15.02

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
132	1	0.22	68	15.25
133	2	0.45	70	15.70
134	4	0.90	74	16.59
135	1	0.22	75	16.82
136	9	2.02	84	18.83
137	3	0.67	87	19.51
138	3	0.67	90	20.18
139	6	1.35	96	21.52
140	2	0.45	98	21.97
141	2	0.45	100	22.42
142	4	0.90	104	23.32
143	10	2.24	114	25.56
144	3	0.67	117	26.23
145	7	1.57	124	27.80
146	3	0.67	127	28.48
147	9	2.02	136	30.49
148	14	3.14	150	33.63
149	1	0.22	151	33.86
150	8	1.79	159	35.65
151	12	2.69	171	38.34
152	11	2.47	182	40.81
154	8	1.79	190	42.60
155	19	4.26	209	46.86
156	6	1.35	215	48.21
157	11	2.47	226	50.67
159	17	3.81	243	54.48
160	20	4.48	263	58.97
162	18	4.04	281	63.00
164	15	3.36	296	66.37
166	23	5.16	319	71.52
168	15	3.36	334	74.89

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
170	21	4.71	355	79.60
172	14	3.14	369	82.74
175	16	3.59	385	86.32
178	9	2.02	394	88.34
182	12	2.69	406	91.03
187	19	4.26	425	95.29
193	8	1.79	433	97.09
202	5	1.12	438	98.21
218	6	1.35	444	99.55
250	2	0.45	446	100.00

Table 48. Grade 4 Social Studies Scale Scores and Conditional Standard Error of Measurement (CSEM)

Raw Score	Scale Score	CSEM
0	0	52
1	1	29
2	26	21
3	38	17
4	47	15
5	55	14
6	61	12
7	66	12
8	70	11
9	74	10
10	78	10
11	81	10
12	84	9
13	87	9
14	90	9
15	92	8
16	95	8
17	97	8
18	99	8
19	101	8
20	104	8
21	106	7
22	107	7
23	109	7
24	111	7
25	113	7
26	115	7

27	116	7
28	118	7
29	119	7
30	121	7
31	123	7
32	124	6
33	125	6
34	127	6
35	128	6
36	130	6
37	131	6
38	133	6
39	134	6
40	135	6
41	137	6
42	138	6
43	139	6
44	141	6
45	142	6
46	143	6
47	145	6
48	146	6
49	147	6
50	149	6
51	150	7
52	152	7
53	154	7
54	155	7
55	157	7
56	159	7

57	160	7
58	163	8
59	164	8
60	167	8
61	169	8
62	172	9
63	175	9
64	178	10
65	181	10
66	188	11
67	191	12
68	197	14
69	205	16
70	216	20
71	235	28
72	250	52

Table 49. Grade 7 Social Studies Scale Scores and Conditional Standard Error of Measurement (CSEM)

Raw Score	Scale Score	CSEM
0	0	54
1	1	30
2	21	21
3	34	18
4	43	15
5	50	14
6	56	13
7	61	12
8	66	11
9	70	10
10	73	10
11	76	9
12	79	9
13	82	9
14	85	8
15	87	8
16	89	8
17	91	8
18	93	8
19	95	7
20	97	7
21	99	7
22	100	7
23	102	7
24	104	7
25	105	7
26	107	7

27	108	7
28	110	7
29	111	7
30	113	7
31	114	6
32	115	6
33	117	6
34	118	6
35	120	6
36	121	6
37	123	6
38	124	6
39	125	6
40	127	6
41	128	6
42	130	7
43	131	7
44	133	7
45	134	7
46	136	7
47	137	7
48	139	7
49	140	7
50	142	7
51	144	7
52	145	7
53	147	7
54	149	7
55	151	8
56	153	8

57	155	8
58	157	8
59	159	8
60	163	9
61	164	9
62	167	9
63	170	10
64	174	10
65	177	11
66	182	12
67	191	13
68	193	14
69	201	16
70	212	20
71	231	29
72	250	53

Table 50. HS Social Studies Scale Scores and Conditional Standard Error of Measurement (CSEM)

Raw Score	Scale Score	CSEM
0	0	45
1	1	25
2	29	18
3	40	15
4	47	13
5	53	11
6	58	10
7	62	10
8	66	9
9	69	9
10	72	8
11	75	8
12	77	8
13	79	7
14	81	7
15	83	7
16	85	7
17	87	6
18	89	6
19	90	6
20	92	6
21	93	6
22	95	6
23	96	6
24	97	6
25	98	5
26	100	5

27	101	5
28	102	5
29	103	5
30	104	5
31	105	5
32	106	5
33	107	5
34	108	5
35	109	5
36	110	5
37	111	5
38	111	5
39	112	5
40	113	5
41	114	5
42	115	4
43	116	4
44	116	4
45	117	4
46	118	4
47	119	4
48	120	4
49	120	4
50	121	4
51	122	4
52	123	4
53	123	4
54	124	4
55	125	4
56	126	4

57	127	4
58	127	4
59	128	4
60	129	4
61	130	4
62	130	4
63	131	4
64	132	4
65	133	4
66	134	4
67	134	4
68	135	4
69	136	4
70	137	4
71	138	5
72	138	5
73	139	5
74	140	5
75	141	5
76	142	5
77	143	5
78	144	5
79	145	5
80	146	5
81	147	5
82	147	5
83	148	5
84	150	5
85	151	5
86	152	5

87	153	5
88	154	5
89	155	5
90	156	6
91	158	6
92	159	6
93	160	6
94	162	6
95	163	6
96	165	6
97	167	7
98	169	7
99	171	7
100	173	7
101	175	8
102	178	8
103	181	9
104	184	10
105	189	11
106	194	12
107	200	14
108	210	17
109	226	24
110	250	44

Table 51. Grade 5 Science Scale Scores and Conditional Standard Error of Measurement (CSEM)

Raw Score	Scale Score	CSEM
0	0	48
1	1	26
2	38	19
3	49	16
4	57	14
5	63	12
6	68	11
7	73	10
8	77	10
9	80	9
10	83	9
11	86	8
12	88	8
13	90	8
14	92	7
15	94	7
16	96	7
17	98	7
18	100	7
19	101	6
20	103	6
21	104	6
22	106	6
23	107	6
24	108	6
25	110	6
26	111	6

27	112	6
28	114	6
29	115	6
30	116	6
31	117	6
32	118	6
33	119	5
34	121	5
35	122	5
36	123	5
37	124	5
38	125	5
39	126	5
40	127	5
41	128	5
42	130	5
43	131	5
44	132	5
45	133	6
46	134	6
47	135	6
48	137	6
49	138	6
50	139	6
51	140	6
52	142	6
53	143	6
54	145	6
55	146	6
56	148	6

57	149	7
58	151	7
59	153	7
60	155	7
61	157	8
62	160	8
63	162	8
64	164	9
65	167	9
66	171	10
67	175	11
68	184	12
69	187	14
70	196	17
71	213	25
72	250	47

Table 52. Grade 8 Science Scale Scores and Conditional Standard Error of Measurement (CSEM)

Raw Score	Scale Score	CSEM
0	0	50
1	1	28
2	8	20
3	20	16
4	29	14
5	35	13
6	41	12
7	46	11
8	50	10
9	54	10
10	57	9
11	60	9
12	63	9
13	66	8
14	68	8
15	70	8
16	73	8
17	75	7
18	77	7
19	78	7
20	80	7
21	82	7
22	83	6
23	85	6
24	86	6
25	88	6
26	89	6

27	90	6
28	91	6
29	93	6
30	94	6
31	95	6
32	96	5
33	97	5
34	98	5
35	99	5
36	100	5
37	101	5
38	102	5
39	103	5
40	104	5
41	105	5
42	106	5
43	107	5
44	107	5
45	108	5
46	109	5
47	110	5
48	111	5
49	112	5
50	113	5
51	113	5
52	114	5
53	115	5
54	116	5
55	117	5
56	118	5

57	118	5
58	119	5
59	120	5
60	121	5
61	122	5
62	123	5
63	124	5
64	124	5
65	125	5
66	126	5
67	127	5
68	128	5
69	129	5
70	130	5
71	130	5
72	131	5
73	132	5
74	133	5
75	134	5
76	135	5
77	136	5
78	137	5
79	138	5
80	139	5
81	140	6
82	142	6
83	143	6
84	144	6
85	145	6
86	146	6

87	148	6
88	149	6
89	151	6
90	152	7
91	154	7
92	155	7
93	157	7
94	159	7
95	161	8
96	164	8
97	165	8
98	168	9
99	171	9
100	174	10
101	178	10
102	182	11
103	190	12
104	192	14
105	200	16
106	211	19
107	230	27
108	250	50

Table 53. HS Science Scale Scores and Conditional Standard Error of Measurement (CSEM)

Raw Score	Scale Score	CSEM
0	0	44
1	1	24
2	37	17
3	47	14
4	55	12
5	60	11
6	65	10
7	69	9
8	73	9
9	76	8
10	78	8
11	81	8
12	83	7
13	85	7
14	87	7
15	89	6
16	91	6
17	92	6
18	94	6
19	95	6
20	96	6
21	98	5
22	99	5
23	100	5
24	101	5
25	102	5
26	103	5

27	104	5
28	105	5
29	106	5
30	107	5
31	108	4
32	109	4
33	109	4
34	110	4
35	111	4
36	112	4
37	112	4
38	113	4
39	114	4
40	115	4
41	115	4
42	116	4
43	117	4
44	117	4
45	118	4
46	119	4
47	120	4
48	120	4
49	121	4
50	122	4
51	122	4
52	123	4
53	123	4
54	124	4
55	125	4
56	125	4

57	126	4
58	127	4
59	127	4
60	128	4
61	129	4
62	129	4
63	130	4
64	131	4
65	131	4
66	132	4
67	133	4
68	134	4
69	134	4
70	135	4
71	136	4
72	136	4
73	137	4
74	138	4
75	139	4
76	139	4
77	140	4
78	141	4
79	142	4
80	143	4
81	143	5
82	144	5
83	145	5
84	146	5
85	147	5
86	148	5

87	149	5
88	150	5
89	151	5
90	152	5
91	154	5
92	155	5
93	156	6
94	157	6
95	159	6
96	160	6
97	162	6
98	164	7
99	166	7
100	168	7
101	170	8
102	172	8
103	175	9
104	178	9
105	182	10
106	187	11
107	193	13
108	202	16
109	218	23
110	250	43

Table 54. Classification Consistency and Accuracy

		Consistency				Accuracy		
Content	Grade	Prob of Consistent Classification (PC)	Prob of Consistent Classification by Chance (Chance)	Kappa	Prob of Misclassification (PM)	Prob of Accurate Classification (PA)	Prob of False Positive Error (FP)	Prob of False Negative Error (FN)
SS	4	0.58	0.32	0.38	0.42	0.63	0.14	0.23
	7	0.55	0.36	0.30	0.45	0.59	0.13	0.28
	HS	NA	NA	NA	NA	NA	NA	NA
SC	5	0.56	0.28	0.38	0.44	0.63	0.21	0.17
	8	0.56	0.33	0.35	0.44	0.60	0.17	0.22
	HS	0.61	0.34	0.40	0.39	0.61	0.36	0.03

Table 55. Test Validity Questions Summary

Question	Subject	Grade	Very Familiar	Somewhat Familiar	Familiar	Somewhat Unfamiliar	Unfamiliar			
How familiar are you with this student?	SS	4	89.20%	4.91%	3.27%	1.64%	0.98%			
		7	91.17%	4.10%	1.26%	2.21%	1.26%			
		HS	78.38%	11.94%	6.98%	2.03%	0.68%			
	SC	5	90.86%	4.06%	2.90%	1.45%	0.73%			
		8	91.76 %	3.79%	1.15%	1.98%	1.32%			
		HS	77.80%	11.88%	7.62%	2.02%	0.67%			
Question		Grade	<1 Hr	1 to <2 Hrs	2 to <3 Hrs	3 to <4 Hrs	4 to<5 Hrs	>=5 Hrs	Do Not Know	
How many hours per week does this student spend in instruction on this content area?	SS	4	32.41%	30.61%	16.86%	8.02%	5.40%	5.07%	1.64%	
		7	23.19%	10.57%	8.36%	19.24%	29.34%	7.57%	1.74%	
		HS	31.08%	13.29%	8.78%	17.12%	17.57%	8.33%	3.83%	
	SC	5	23.22%	27.43%	17.71%	13.06%	11.32%	5.81%	1.45%	
		8	18.45%	9.23%	11.86%	20.43%	31.47%	7.25%	1.32%	
		HS	32.29%	11.21%	8.30%	18.16%	17.94%	8.07%	4.04%	
Question		Grade	25%	50%	75%	100%	None	Missing		
Approximately how much instructional time for this content area is in the general education classroom?	SS	4	21.44%	6.55%	9.82%	30.77%	31.42%	0.00%		
		7	9.46%	5.84%	9.78%	25.71%	49.21%	0.00%		
		HS	7.88%	5.18%	4.05%	12.61%	70.05%	0.23%		
	SC	5	20.17%	7.11%	12.63%	34.54%	25.54%	0.00%		
		8	7.25%	5.27%	11.37%	31.96%	44.15%	0.00%		
		HS	8.07%	4.48%	5.38%	13.00%	68.83%	0.22%		
Question		Grade	Oral Language	Reading	Picture Communication	Tactile	Other	Do Not Know		
This student's primary receptive communication is:	SS	4	92.96%	1.31%	3.11%	0.16%	1.96%	0.49%		
		7	92.59%	1.74%	2.21%	0.32%	2.52%	0.63%		
		HS	88.74%	4.05%	2.93%	0.00%	2.25%	2.03%		
	SC	5	92.02%	1.60%	3.05%	0.44%	2.76%	0.15%		
		8	94.23%	0.66%	1.98%	0.16%	2.64%	0.33%		
		HS	88.34%	4.04%	3.36%	0.00%	2.24%	2.02%		

Question		Grade	Oral Language	Writing	Picture Communication	Tactile	Other	Do Not Know		
This student's primary expressive communication is:	SS	4	84.29%	0.82%	7.36%	0.00%	6.71%	0.82%		
		7	86.12%	1.10%	5.36%	0.32%	6.47%	0.63%		
		HS	82.43%	2.03%	6.76%	0.45%	6.31%	2.03%		
	SC	5	85.20%	0.73%	6.53%	0.15%	7.26%	0.15%		
		8	85.67%	0.82%	5.60%	0.00%	7.74%	0.16%		
		HS	82.29%	2.24%	6.05%	0.67%	6.73%	2.02%		
Question		Grade	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree	Do Not Know		
I feel that the student's responses accurately reflect their understanding of the material.	SS	4	32.57%	39.77%	16.20%	6.55%	3.60%	1.31%		
		7	36.12%	41.32%	14.04%	4.10%	3.00%	1.42%		
		HS	36.71%	39.64%	12.16%	5.41%	3.38%	2.70%		
	SC	5	41.51%	39.62%	11.32%	3.63%	2.32%	1.60%		
		8	43.33%	39.37%	10.71%	4.45%	1.48%	0.66%		
		HS	36.77%	39.69%	14.13%	4.71%	2.69%	2.02%		
Question		Grade	0–15 min	16–30 min	31–60 min	61–90 min	91–120 min	121–150 min	151–180 min	>=181 min
How much time did this student take on the assessment?	SS	4	2.78%	44.03%	46.81%	4.75%	1.15%	0.49%	0.00%	0.00%
		7	2.21%	34.23%	55.99%	5.68%	1.58%	0.16%	0.16%	0.00%
		HS	0.68%	24.10%	60.59%	10.81%	2.93%	0.45%	0.23%	0.23%
	SC	5	5.08%	49.49%	39.91%	4.35%	1.02%	0.00%	0.15%	0.00%
		8	3.13%	43.00%	47.28%	4.61%	1.48%	0.49%	0.00%	0.00%
		HS	1.35%	26.01%	58.07%	10.76%	3.14%	0.22%	0.22%	0.22%

Table 56. Items Field Tested and Item Performance Review Outcomes

	Social Studies			Science		
	Grade 4	Grade 7	HS	Grade 5	Grade 8	HS
Number of field test forms	1	1	2	1	2	2
Number of items field tested	6	6	8	6	8	8
Item performance review outcome						
• Flagged Items	0	0	1	0	0	0

APPENDICES

APPENDIX A: COALT: SCIENCE AND SOCIAL STUDIES ELIGIBILITY GUIDELINES

Alternate Academic Achievement Standards and Alternate Assessment Participation Guidelines Worksheet

***For further clarification of terms used in this worksheet, please refer to the companion document**
Participation Guidelines: Alternate Academic Achievement Standards for Instruction and Alternate Assessment

Criterion #1: **The student has been evaluated and determined to be eligible to receive special education services and has an IEP.**	**Response:**
☐ Has the student been determined to be a student with a disability eligible to receive special education services under the Individuals with Disabilities Education Act (IDEA)? ☐ Has a current Individualized Education Program (IEP) been developed for the student?	☐ **No**. **Stop here.** The student must meet Special Education *Determination of Eligibility* criteria in one or more disability categories defined in ECEA Rules http://www.cde.state.co.us/cdesped/IEP_Forms.asp ☐ **Yes**. If both elements can be affirmed, **continue to Criterion #2.**
Criterion #2: **The student has documented evidence of a cognitive disability.**	**Response:**
☐ During the process of determining eligibility for a student, did the IEP Team review a body of evidence that supports the existence of a cognitive disability?	☐ **No**. **Stop here.** The student must have documented evidence of the existence of a cognitive disability, regardless of the special education disability category. ☐ **Yes**. Empirical evidence of a cognitive disability is documented in the IEP. **Continue to Criterion #3.**
Criterion #3: **The student has a <u>significant</u> cognitive disability.**	**Response:**
☐ The student's demonstrated cognitive functioning and adaptive behavior in the home, school, and community environments are **significantly below age expectations, even with program modifications, adaptations and accommodations** <u>and</u> ☐ the School Psychologist (or other personnel trained in administering psychometric evaluation) presents evidence that the student's cognitive and adaptive functioning is consistent with that of a student with a **significant cognitive disability*.** ***Empirical evidence includes, but is not limited to, formal testing results, multi-disciplinary team evaluations, and other evaluative data.***	☐ **No.** The documented evidence supporting the existence of a cognitive disability **does not fall into the "significant cognitive disability" range**. With appropriate adaptations (supports and accommodations), the student receives daily instruction based on the Colorado Academic Standards enrolled grade-level expectations. The student does not qualify for instruction on alternate academic achievement standards or to take alternate assessment based on alternate academic achievement standards. **Continue to 4A to select <u>Grade-level</u> standards-based instruction and appropriate grade-level assessment.** ☐ **No.** The documented evidence supporting the existence of a cognitive disability **does not fall into the "significant cognitive disability" range**. **However**, the IEP Team has considered the impact of the disability and other related factors in order to determine that the student qualifies to receive daily instruction based on the Colorado Academic Standards Extended Evidence Outcomes (alternate academic achievement standards) and participate in alternate assessment based on alternate academic achievement standards. **Continue to 4B to select <u>Alternate</u> standards-based instruction and appropriate alternate assessment.** ☐ **Yes**. Both elements affirm that the student meets the qualifications as a student with a **significant cognitive disability**. The student (a) requires extensive, repeated individualized instruction and support that is not of a temporary or transient nature and (b) uses substantially adapted and modified materials and individualized methods of accessing information in alternative ways to acquire, maintain, generalize, demonstrate and transfer academic and functional skills necessary for application in school, work, home and community environments. Daily modified instruction is linked to the enrolled grade level **Colorado Academic Standards Extended Evidence Outcomes (EEOs)**. For students receiving instruction on alternate standards and taking alternate assessment, the IEP must contain measurable annual goals **and** objectives for content areas. **Continue to 4B to select alternate standards-based instruction and alternate assessment.**

For questions related to this optional worksheet and companion guidance, please contact:
Linda Lamirande Lamirande_L@cde.state.co.us Exceptional Student Services Unit Colorado Dept. of Education
Rev. 8/13

Alternate Academic Achievement Standards and Alternate Assessment Participation Guidelines Worksheet

Tested Content Areas	4A Instruction and Assessment based on Grade-Level Academic Achievement Standards (Grade-level Expectations / Evidence Outcomes)	4B Instruction based on Extended Evidence Outcomes (EEOs) and Alternate Assessment based on Alternate Academic Achievement Standards (AA-AAS)
Reading **Writing** **Math** **Science** **Social Studies**	☐ Grade-level classroom/ district assessments ☐ with accommodation ☐ without accommodation ☐ State Summative Assessment ☐ with accommodations allowed for use on state assessment ☐ without accommodation ☐ nonstandard request- pending approval by CDE Assessment Unit	☐ Alternate classroom/ district assessments based on alternate standards ☐ Alternate State Summative Assessments **Note**: With the passage of IDEA in 1997 and its reauthorization in 2004, it is required that both state and districts provide an alternate assessment for students who cannot participate in general state and district assessments.
Dual Assessment	Typically, if a student meets participation guidelines for alternate standards for instruction, the alternate assessment will be taken for all content areas tested in the student's enrolled grade level. However, in a few rare instances, a student may demonstrate specific academic strength in a particular content area. The IEP Team may determine that a student receive grade-level instruction and participate in grade-level assessment in one or more content areas, but receive instruction under alternate standards and take an alternate assessment in another content area. In such cases, a **Request for Dual Assessment** form must be submitted to the CDE Assessment Unit. (See Assessment Appendix in the Colorado Accommodation Manual)	
Other	☐ ACCESS for ELLs (K-12) ☐ with allowable accommodations ☐ Colorado ACT ☐ with allowable accommodations for use on the ACT assessment	☐ Alternate ACCESS for ELLs (Gr. 1-12) ☐ 11th Grade Alternate Assessment for Colorado ACT

Exclusionary Factors:

The IEP Team affirms

- ☐ that annual assessment data was reviewed for each content area and
- ☐ the decision for participation in the Alternate Assessment is **NOT** based on:
  1. A disability category or label
  2. Poor attendance or extended absences
  3. Native language/social/cultural or economic difference
  4. Expected poor performance on the grade-level assessment
  5. Services student receives
  6. Educational environment or instructional setting
  7. Percent of time receiving special education
  8. English Language Learner (ELL) status
  9. Low reading level/academic level
  10. Anticipated student's disruptive behavior
  11. Impact of student scores on accountability system
  12. Administrator decision
  13. Anticipated student's emotional duress

IEP Team Consensus: (Record decision on IEP Form)

☐ **Student meets participation guidelines as a student with a significant cognitive disability and will receive instruction based upon alternate academic standards and participate in alternate assessment as indicated above.**

***For further clarification of terms used in this worksheet, please refer to the companion document**

Participation Guidelines: Alternate Academic Achievement Standards for Instruction and Alternate Assessment

For questions related to this optional worksheet and companion guidance, please contact:
Linda Lamirande Lamirande_L@cde.state.co.us Exceptional Student Services Unit Colorado Dept. of Education
Rev. 8/13

APPENDIX B: COALT: SCIENCE AND SOCIAL STUDIES TEST BLUEPRINTS

CoAlt Blueprint – Grade 4 Social Studies

TEST BLUEPRINT CoAlt Social Studies Grade 4		SRs	SPTs	Total Points	Total Items	% of Score Points
1	History	4	0	16	4	22%
	GLE 1	2	0	8		
	GLE 2	2	0	8		
2	Geography	4	0 or 1	16 or 22	4 or 5	22% or 31%
	GLE 1	2	0 or 1	8 or 14		
	GLE 2	2	0	8		
3	Economics	4	0 or 1	16 or 22	4 or 5	22% or 31%
	GLE 1	2	0	8		
	GLE 2	2	0 or 1	8 or 14		
4	Civics	3	1	18	4	25%
	GLE 1	2	0	8		
	GLE 2	1	1	10		
	TOTAL	**15**	**2**	**72**	**17**	**100%**

Note: SRs=selected response items, SPTs=supported performance task items, and GLE=grade level expectation

CoAlt Blueprint – Grade 5 Science

TEST BLUEPRINT CoAlt Science Grade 5		SRs	SPTs	Total Points	Total Items	% of Score Points
1	Physical Science	3	0	12	3	17%
	GLE 1	3	0	12		
2	Life Science	6	1	30	7	42%
	GLE 1	3	0 or 1	12 or 18		
	GLE 2	3	0 or 1	12 or 18		
3	Earth Systems Science	6	1	30	7	42%
	GLE 1	2	0 or 1	8 or 14		
	GLE 2	2	0 or 1	8 or 14		
	GLE 3	2	0 or 1	8 or 14		
	TOTAL	**15**	**2**	**72**	**17**	**100%**

Note: SRs=selected response items, SPTs=supported performance task items, and GLE=grade level expectation

CoAlt Blueprint – Grade 7 Social Studies

TEST BLUEPRINT CoAlt Social Studies Grade 7		SRs	SPTs	Total Points	Total Items	% of Score Points
1	History	4	0 or 1	16 or 22	4 or 5	22% or 31%
	GLE 1	2	0 or 1	8 or 14		
	GLE 2	2	0	8		
2	Geography	4	0 or 1	16 or 22	4 or 5	22% or 31%
	GLE 1	2	0 or 1	8 or 14		
	GLE 2	2	0	8		
3	Economics	3	0	12	3	17%
	GLE 1	2	0	8		
	GLE 2	1	0	4		
4	Civics	4	1	22	5	31%
	GLE 1	2	1	14		
	GLE 2	2	0	8		
	TOTAL	**15**	**2**	**72**	**17**	**100%**

Note: SRs=selected response items, SPTs=supported performance task items, and GLE=grade level expectation

CoAlt Blueprint – Grade 8 Science

TEST BLUEPRINT CoAlt Science Grade 8		SRs	SPTs	Total Points	Total Items	% of Score Points
1	Physical Science	6 or 7	0 or 1	28 or 30	7	26% or 28%
	GLE 1	0	0 or 1	0 or 6		
	GLE 2	1 or 2	0	4 or 8		
	GLE 3	2	0	8		
	GLE 4	3	0	12		
2	Life Science	6 or 7	0 or 1	28 or 30	7	26% or 28%
	GLE 1	1 or 2	0 or 1	4 to 14		
	GLE 2	4 to 6	0	16 to 24		
3	Earth Systems Science	11	1	50	12	46%
	GLE 1	2	0 or 1	8 or 14		
	GLE 2	3	0	12		
	GLE 3	3	0 or 1	12 or 18		
	GLE 4	3	0 or 1	12 or 18		
	TOTAL	**24**	**2**	**108**	**26**	**100%**

Note: SRs=selected response items, SPTs=supported performance task items, and GLE=grade level expectation

CoAlt Blueprint – HS Social Studies

TEST BLUEPRINT CoAlt Social Studies High School		SRs	SPTs	Total Points	Total Items	% of Score Points
1	History	6	1	30	7	27%
	GLE 1	1	1	10		
	GLE 2	4	0	16		
	GLE 3	1	0	4		
2	Geography	6	1	30	7	27%
	GLE 1	2	1	14		
	GLE 2	2	0	8		
	GLE 3	2	0	8		
3	Economics	6	1	30	7	27%
	GLE 1	1	0	4		
	GLE 2	1	0	4		
	GLE 3	1	0	4		
	GLE 4	0	1	6		
	GLE 5	1	0	4		
	GLE 6	1	0	4		
	GLE 7	1	0	4		
4	Civics	5	0	20	5	18%
	GLE 1	1	0	4		
	GLE 2	2	0	8		
	GLE 3	2	0	8		
	TOTAL	**23**	**3**	**110**	**26**	**100%**

Note: SRs=selected response items, SPTs=supported performance task items, and GLE=grade level expectation

CoAlt Blueprint – HS Science

TEST BLUEPRINT CoAlt Science High School		SRs	SPTs	Total Points	Total Items	% of Score Points
1	Physical Science	6	1	30	7	27%
	GLE 1	1	0 or 1	4 or 10		
	GLE 2	1	0 or 1	4 or 10		
	GLE 3	1	0	4		
	GLE 4	1	0 or 1	4 or 10		
	GLE 5	1	0 or 1	4 or 10		
	GLE 6	1	0	4		
2	Life Science	10	1	46	11	42%
	GLE 1	1	0 or 1	4 or 10		
	GLE 2	1	0 or 1	4 or 10		
	GLE 3	1	0 or 1	4 or 10		
	GLE 4	1	0 or 1	4 or 10		
	GLE 5	1 or 2	0	4 or 8		
	GLE 6	1	0 or 1	4 or 10		
	GLE 7	1 or 2	0	4 or 8		
	GLE 8	1	0	4		
	GLE 9	1	0	4		
3	Earth Systems Science	7	1	34	8	31%
	GLE 1	1	0	4		
	GLE 2	1	0 or 1	4 or 10		
	GLE 3	1	0 or 1	4 or 10		
	GLE 4	1	0 or 1	4 or 10		
	GLE 5	1	0	4		
	GLE 6	1	0 or 1	4 or 10		
	GLE 7	1	0	4		
	TOTAL	**23**	**3**	**110**	**26**	**100%**

Note: SRs=selected response items, SPTs=supported performance task items, and GLE=grade level expectation

APPENDIX C: COALT: SCIENCE AND SOCIAL STUDIES STANDARD-SETTING REPORT

Colorado Alternate Assessment (CoAlt)
Fall 2014 Standard Setting Report

COLORADO DEPARTMENT *of* EDUCATION

&

March 19, 2015

Table of Contents

OVERVIEW

In February 2015, the Colorado Department of Education (CDE) convened two standard setting committees to recommend three cut scores that would define four performance levels—*Novice Level*, *Developing Level*, *Emerging Level*, and *Exploring Level*—for the new Colorado Alternate Assessment (CoAlt): Science and Social Studies High School assessments. The purpose of this document is to provide a detailed report of the standard setting process for the Fall 2014 administration of the new CoAlt: Science and Social Studies assessments at high school.

CoAlt is a standards-based assessment designed for students with a significant cognitive disability who are unable to participate in the Colorado Measures of Academic Success (CMAS) assessments, even with accommodations. The CoAlt: Science and Social Studies assessments are aligned to the Extended Evidence Outcomes (EEOs) of the Colorado Academic Standards (CAS) in the content areas of science and social studies, which can be located at http://www.cde.state.co.us/coextendedeo/statestandards.

The CoAlt: Science and Social Studies assessments have test books that are used by a Test Examiner to administer test items to a student. The test book is oriented so that the Test Examiner administers the test while facing the student. The test book includes scripted text for the Test Examiner to read test questions and answer choices to the student. During the course of the administration, the Test Examiner scores each item; and at the conclusion of the administration, the Test Examiner enters the student's scores into an online score entry system.

Each CoAlt: Science and Social Studies assessment contains selected response (SR) items and supported performance task (SPT) items. SR items contain a primary prompt with a question and three answer options from which the student selects an answer. If the student responds incorrectly or does not respond to the primary prompt after it is repeated, an additional prompt is presented to the student to provide the student with an example that is similar to the primary prompt and answer options. The additional prompt is used to engage the student with the item. If the student responds incorrectly or does not respond to the additional prompt, the student is presented with the correct answer and is presented with the primary prompt again to have another opportunity to respond. In essence, the student will work with the test item until he or she provides the correct answer or the maximum number of attempts is reached. Each selected response item is scored using a four-point rubric (see Appendix A for the Selected Response Scoring Rubric).

SPT items are composed of three prompts that are related to one overall task. This item type requires students to manipulate option cards by placing them on a designated response page (e.g., placing option cards in designated boxes within a chart or diagram). Each of the three prompts is scored using a two-point rubric. The points for the three prompts are then added together to provide one score for the SPT item (see Appendix A for the Supported Performance Task Scoring Rubric).

The new CoAlt: Science and Social Studies assessments are available in elementary (ES), middle (MS), and high school (HS). The subject and grade combinations for CoAlt are shown in Table 1. The first operational administration of the grades 4, 5, 7, and 8 assessments was in April 2014, and the performance standards, or cut scores, for the ES and MS assessments were

subsequently set in July 2014. The first operational administration of the HS assessments was from November 3–21, 2014. The cut scores for the HS assessments were recommended in February 2015 to aid in the interpretability of the test scores.

Table 1: CoAlt Subjects and Grades

	Grade				
	4	**5**	**7**	**8**	**HS**
Science		X		X	X
Social Studies	X		X		X

To support the interpretation of student results of the new CoAlt: Science and Social Studies HS assessments, student performance is described in terms of four performance levels—*Novice Level*, *Developing Level*, *Emerging Level*, and *Exploring Level*. The CoAlt standard setting meeting held in February 2015 was convened to obtain cut score recommendations to assist the state in delineating thresholds for each of the four performance levels. When student performance is not evident across all the items on the assessment (i.e., an overall test score of zero), students will receive an indicator of *Inconclusive*.

The Modified Extended Angoff approach (Cizek, 2012; Cizek, Bunch, & Koons, 2004; Hambleton & Plake, 1995) was used to set performance standards on the CoAlt: Science and Social Studies HS assessments. With this methodology, standard setting panelists review the content of each test item, and considering the content the item is measuring and the content knowledge of the students right at the cut scores (i.e., threshold students), the panelists make a judgment about what score a threshold student should receive on the item to be considered "just-barely" in a performance level. Panelists use performance level descriptors (PLDs) to conceptualize "threshold" students (those students just barely in a particular performance level) in order to determine the score the threshold student should obtain on each item. The individual item-level cut scores for each particular performance level are then summed for each panelist to obtain the recommended test-level cut scores that are used to define the performance levels.

The Reasoned Judgment approach (Roeber, 2002) was also used in this methodology to help panelists think about the level of content knowledge that may be needed for a student to earn a specific rubric score, the patterns of performance (i.e., combinations of item scores) that lead to overall test scores, and whether various scoring patterns make sense for a given performance level. Different patterns of student performance, called score profiles, were presented to panelists with this approach. The score profile is a graphical representation of how a student could achieve a specific test score (i.e., the score points a student earned on each item to obtain a specific test score). In essence, the score profiles were used to help panelists understand what the overall group of item scores for a test can tell us about what a student knows and can do.

PREPARATION FOR STANDARD SETTING

Preparation for standard setting started months before the actual meeting. This section provides details about the selection of panelists, the development of the PLDs, the various materials that were gathered or created for the meeting, and the training of those who facilitated the meeting and analyzed the data.

PANELIST SELECTION AND COMPOSITION

The CoAlt: Science and Social Studies HS standard setting meeting included ten panelists for each subject-area committee. Panelists were grouped into tables of three within each meeting room. CDE selected the panelists for each committee to represent the state in terms of gender and ethnicity as well as relevant demographic characteristics (e.g., school size, geographic location). The CoAlt panelists included educators who taught at the high school level, including special educators with experience working with students with significant cognitive disabilities, special educators with experience working with students with other types of disabilities, and content experts with knowledge of the subject-area curriculum. In addition to classroom teachers, special education administrators and higher education representatives also participated in the meeting. Panelists from the CMAS Science and Social Studies HS standard setting meeting were also recruited to participate in the CoAlt HS standard setting meeting. Including panelists from the prior CMAS standard setting meeting helped provide context to the CoAlt panelists regarding how the earlier recommended performance standards were determined. Appendix B describes panel composition for each grade-level committee.

DEVELOPMENT OF PLDS

PLDs are an important tool for recommending cut scores. PLDs outline the expectations of student performance at each performance level of a test. The CoAlt: Science and Social Studies HS PLDs were written prior to the standard setting meeting and were developed by CDE and Pearson content experts and reviewed and edited by a committee of Colorado educators, comprised of both general education teachers and special education teachers. The educators reviewed the PLDs for each HS subject-area individually, and following the PLD educator committee meeting, CDE and Pearson staff reviewed the feedback from the educators and incorporated their feedback into the PLDs, where appropriate. During the standard setting meetings, the standard setting panelists were offered the opportunity to provide additional clarifications to the PLDs. Following the standard setting meetings, CDE incorporated panelists' feedback into the PLDs where appropriate. The final CoAlt PLDs are provided in Appendix C.

CREATION OF MATERIALS

A standard setting meeting requires a myriad of materials. Documents were obtained from several different sources for the meeting. Some documents were uniquely created for panelists, while other documents were obtained from the materials distributed from the Fall 2014 CoAlt: Science and Social Studies HS test administration or downloaded and printed from the CDE website. CDE reviewed and edited all documents, as needed, prior to the standard setting meeting. This section outlines the primary materials for the meeting and lists where the documents can be found. A description of how the preceding documents were used during the standard setting meeting can be found later in the report.

Agendas

There were two main components of the meeting: a general session (a large-group setting) and a breakout session (a small-group setting). A general agenda, which contained an outline of the standard setting tasks that all the panelists would be completing during the meeting, was created and provided to the panelists at the beginning of the general session. A specific agenda was also created, and it was provided to CDE and Pearson staff. This agenda outlined the same tasks listed in the general agenda, but with more detail regarding each task and the specific times each task was to begin and end.

Slides and Script

For the general session, a PowerPoint presentation was created to provide a general overview of the standard setting meeting. For the breakout session, an additional PowerPoint presentation and an accompanying detailed script were developed. The slides and the script allowed for the breakout sessions to be standardized for each grade-level committee.

CoAlt Test Book

To allow the panelists the opportunity to become familiar with the items and the scoring of the CoAlt HS assessment, the Fall 2014 CoAlt test book corresponding to each subject-area committee was provided to panelists to use as part of the standard setting process. All operational items that appeared on the Fall 2014 assessment were included in the test book. The field test item pages in the test book were covered as those items were not part of the standard setting process because the determination of whether the items would be eligible for future operational tests had yet to be determined. In addition to the test book, the Assessment Frameworks for each grade-level meeting was provided to the panelists. The Assessment Frameworks for each subject-area can be found in Appendix D.

Rubrics

The SR item and the SPT item rubrics were provided to the panelists to refer to as needed as they participated in the standard setting process.

Score Profiles

Examples of several different patterns of student performance, or score profiles, were presented to the panelists for review during the general session and breakout sessions. When reviewing the score profiles, panelists were asked to think about what students should know and be able to do to achieve a certain rubric score, what the group of scores in the score profile can tell us about what a student knows and can do, and whether the score profiles make sense for a given performance level.

Item Ordered Book

During the rounds of rating, the panelists in each subject-area meeting reviewed the operational items in an item ordered book. The item ordered book contained the operational items ordered by item difficulty using the item mean score (i.e., the average rubric score obtained by all students who took an item). Ordering items by the item mean scores allowed panelists insight into how easy or hard the items were for all the students taking the items and allowed them to consider this information when determining their score recommendations. Each item in the item ordered book

was labeled with the item number based on its sequence in the item ordered book and the item number based on its sequence in the CoAlt operational test.

Item Mean Reports

Item means were provided to panelists as part of the feedback provided after Round 1 recommendations. The item means were presented with the highest to the lowest item mean score, which corresponded to how the operational items were presented in the item ordered book.

Forms

Numerous forms were created for panelists to complete and include the following:

- Panelist Information Form: While some demographic information was already included in the database of Colorado educators, the panelist information sheet was used to collect additional demographic information.

- Performance Level Content Match Form: To help panelists better consider the content of each item when providing their item-level cut score recommendations, panelists completed a performance level content matching activity. As part of this activity, panelists matched each item to a performance level based on the concepts and skills the item is measuring and used the Performance Level Content Match form to indicate the performance level they chose for an item.

- Readiness Survey: A brief questionnaire was provided to panelists before each round of the standard setting process in which panelists are asked to verify that they understand the task at hand and are ready to move forward with providing their recommendations.

- Ratings Recommendation Forms: The ratings forms were used to collect panelists' item ratings for each round.

- Standard Setting Evaluation: An evaluation was administered after the standard setting to gather information on panelists' perceptions of the meeting.

TRAINING OF FACILITATORS AND DATA ANALYSTS

Several meetings were held with the facilitators and data analysts to properly train and prepare them for the meeting. For the facilitator training, an overview of the new CoAlt assessments were provided and the breakout session slides and script were reviewed and discussed in detail to ensure that all facilitators were in sync in terms of how to lead the panelists through the standard setting process and the logistics of the meeting. In addition to reviewing the slides and script, the facilitators also reviewed their facilitator materials and the materials to be distributed to the panelists during the meeting.

For the data analysts, it was important that the analysis spreadsheets be set up properly to ensure accurate and rapid analysis of panelists' recommendations. All the analysis code and spreadsheets created for the meetings were tested and verified before the meetings. Although not specifically trained for the meeting, it should be noted that the Pearson CoAlt content specialists met with the lead facilitator to discuss the standard setting process and meeting logistics and

were available throughout the standard setting meeting to answer any content-related questions posed by panelists.

STANDARD SETTING MEETING ACTIVITIES

The standard setting for the CoAlt: Science and Social Studies HS assessments was held on February 18–19, 2015. During the two-day meeting, panelists from each of the two subject-area standard setting committees received training on the assessment and the standard setting process, reviewed the grade-level PLDs, reviewed the Fall 2014 operational items, reviewed the threshold student descriptors, and applied the Modified Extended Angoff method to establish cut score recommendations across three rounds of rating. During the process of establishing cut score recommendations, panelists also reviewed the content assessed by the CoAlt items and matched the items to performance levels based on the concepts and skills found in the PLDs, engaged in table and committee-level discussions, and considered the impact of their cut scores on student performance when making their cut score recommendations. The specific procedures involved in the CoAlt standard setting are described in the sections that follow.

GENERAL SESSION

The standard setting meeting began with a general session in which panelists from both subject-areas convened to listen to introductory comments and receive directions for the meeting. To begin the general session, a representative from CDE welcomed the panelists to the meeting and provided the context for the meeting by presenting details describing the CoAlt assessments and the importance of standard setting in the assessment development process. This information helped the panelists understand what standard setting is and the reason they were asked to be part of a standard setting committee.

Next, a member of the Pearson Psychometric Services staff provided an overview of how the CoAlt items are scored and how to understand the scores (i.e., what scores may indicate a student has content knowledge versus what scores indicate a student has pre-content, or entry-level, skills). A brief overview of the standard setting process and a description of the Modified Extended Angoff method, including the rationale behind the procedure and the types of decisions panelists would be asked to make during the meeting, was also discussed. Important considerations, such as the number of score points a student could earn without demonstrating content knowledge and the number of points a student could earn by guessing alone, were also discussed in relation to CoAlt assessments. Score profiles were shown to panelists to help them understand these considerations. A high-level agenda containing the tasks the panelists would complete over the two-day meeting was also provided to the panelists. Once the general session was completed, panelists were dismissed to their designated breakout session rooms.

THE STANDARD SETTING PROCESS

The standard setting specific tasks took place over the course of two days as outlined in this section of the report. Each content committee was facilitated independently, but the same standardized process was used across both content areas.

Review and Discuss Performance Level Descriptors

In the breakout session rooms, each grade-level meeting began with the facilitator welcoming the panelists to the meeting and thanking them for their participation. CDE staff observed each breakout room for the remainder of the meeting to observe the process and to answer any assessment, content, or policy related questions asked by the panelists. Trained Pearson facilitators then followed with formal introductions of all participants, a review of the meeting agenda, and answered any panelist's questions regarding meeting logistics and the standard setting process.

After introductions and general housekeeping tasks were completed, each panelist was provided with a document listing the grade-level PLDs for the committee meeting. Panelists use the PLDs to obtain a common understanding of the knowledge, skills, and abilities possessed by a student clearly in the middle of each performance level for a grade and subject. After being given the specific grade-level PLDs, panelists were then asked to review the performance labels and the PLDs in conjunction with the content frameworks and write down any comments they had regarding the PLDs. Pearson content specialists recorded educator comments and suggestions for CDE to review and consider for incorporation into the final PLDs.

After providing comments regarding the PLDs, the meeting facilitator led the panelists in a discussion of the characteristics that most differentiate the four adjacent performance levels until they could clearly distinguish between each level. The panelists were instructed to refer to these characteristics as they moved through the standard setting process.

Review Assessment Items

To become more familiar with the test for which they would be setting performance standards, the panelists reviewed the CoAlt HS assessment items. Before having the panelists review the assessment items, the facilitator provided the panelists with another presentation of the CoAlt scoring rubrics that included information on what rubric scores may indicate content knowledge versus entry-level skills. After the refresher, panelists were asked to review the assessment items. After reviewing the test items, panelists discussed the types of knowledge and skills the students are asked to demonstrate for each item and the amount of support they believed students would need to complete each item. In addition, panelists discussed the test itself in terms of content, difficulty, and the construct being measured.

Performance Level Content Matching

After reviewing and discussing the CoAlt test items, panelists completed the performance level content matching activity. To help panelists better consider the content of each item when providing their item-level cut score recommendations later in the standard setting process, panelists were asked to think about the concepts and skills the operational items measure and think about the concepts and skills within each performance level. Panelists were then asked to match each item to a performance level based on the concepts and skills the item measures. The panelists used the Performance Level Content Match form found in Appendix E to write a performance level next to each item. After completing the performance level content matching individually, the entire room shared and discussed their performance level content matches and agreed on performance level labels for each item. This information was then transferred to the

Practice Exercise Ratings sheet and the Ratings Recommendation Forms that would be used to record the item-level cut score for each item.

Development of Threshold Descriptors

Panelists were reminded that the main purpose for reviewing and discussing the PLDs was to operationalize the performance levels to *support the standard setting task*. The focus of this next activity was on the threshold students—those students who "just barely" make it into a particular performance level. These students are the focus of standard setting because it is these students the panelists must consider when recommending the cut scores that define the four performance levels. The goal of this activity was to have the panelists develop threshold student descriptors as a whole group to gain a common understanding of these students so that when panelists were asked to think about a threshold student, they were all in agreement regarding what such a student knows and can do.

To develop the threshold student descriptors, panelists were asked to identify concepts and skills in a given PLD that should describe the threshold student. Questions that helped guide the discussion included:

- Do any concepts and skills listed in the PLD do this outright?
- How could you modify or constrain the PLD to better reflect the limited capabilities of the "just-barely" student?
- What should the "threshold" student be able to do relative to these particular skills?

Each of the three table groups worked together to create specific descriptions that would separate students who are just barely in a particular performance level (threshold students) from students who are at the top of the previous performance level. At this point, the concept of table leaders was introduced to the committee. Table leaders were identified early in the breakout sessions and helped keep each table group on track with tasks and discussions. Once the threshold student descriptors were drafted at the table level, the entire room shared and discussed their threshold descriptors and agreed on a final set of threshold student descriptors for their specific grade. Once final, the threshold student descriptors were printed for each panelist to use throughout the remainder of the standard setting activity.

Standard Setting Training and Practice Round

After the development of the threshold student descriptors, panelists were introduced to the Modified Extended Angoff standard-setting method. The meeting facilitators introduced the method to the panelists and then explained the steps that the panelists would need to complete as part of the method. Following the training session, panelists engaged in a practice round of standard setting using a small set of items. The purpose of this exercise was to have panelists practice evaluating and rating items to make sure they were comfortable with the task.

For the practice exercise, an item ordered book with a set of six items and a Practice Exercise Ratings sheet with the performance level for each item were presented to the panelists (see Appendix F for an example). Panelists were asked to review each item in conjunction with the performance level content label, the PLDs, and the threshold student descriptors. Panelists were then asked to think about the threshold student that just barely makes it into a performance level and determine what rubric score a threshold student would receive on the item to be considered

just barely in each performance level. The rubric score available for panelists to consider was based on the expectation that threshold students can model the correct answer when instructed by the teacher. In addition, panelists also considered the assumption of performance found in the PLDs (i.e., students who can demonstrate the concepts and skills at higher levels should be able to demonstrate the concepts and skills within the lower levels). The following outlines the specific steps that were to be followed for the "Emerging Level" cut.

1. Review the items listed on the rating sheet.
2. Identify the skills required for the item or task.
    - Think about the performance level content the item is measuring.
3. Decide: Should threshold Emerging Level students be able to demonstrate the skills assessed by the item?
4. Decide: How should performance appear for the threshold students?
    - Think about the difficulty of the item, whether the content of the item is measuring is below the performance level cut you are recommending, whether the threshold student should be able to answer the item correctly, and what level of support should they need.
5. On the ratings sheet, indicate the item-level score you feel describes what a threshold student should be able to obtain.
    - Score range available for each item is based on the expectation that threshold students can model the correct answer when instructed by the teacher.

The same steps were repeated for the "Developing Level" and "Novice Level" cuts. Panelists were reminded that because the content standards are new, they may not yet be fully implemented, so it was important that panelists consider threshold students who have been instructed in the new standards when determining their ratings. Before beginning their practice ratings, panelists were asked to complete a practice round readiness form that indicated they understood the steps of the process and were ready to provide the item-level cut scores for each performance level. After the panelists provided their ratings on their Practice Exercise Ratings sheet, the facilitator asked the panelists to share their rating results with the whole group, leading to a group discussion where panelists discussed their ratings and the general process employed. Based on the panelists' discussion, facilitators provided additional instructions and guidance as needed.

Readiness Survey

To evaluate whether the training activities successfully helped panelists understand the task, a readiness survey was completed by each panelist prior to each round of recommendations (see Appendix G). The readiness survey asked panelists to report if they understood the task asked of them as well as any feedback data provided. Results of the readiness survey indicated that panelists unanimously understood their tasks for each round and understood the data presented.

Round 1

After completing the readiness survey, the panelists were ready to begin Round 1 of the standard setting. Prior to beginning Round 1, panelists were reminded to think about the performance level content each item measures, the PLDs, the scoring rubrics, and the threshold student descriptors. During Round 1, panelists received a round readiness form and a Round 1 Ratings form to complete (see Appendix H for an example). Panelists worked independently to make

their item-level cut score ratings for each performance level, and when they were finished providing their ratings, the meeting facilitator collected each panelist's ratings form and the panelists were dismissed for the day.

Round 1 Feedback

To begin day 2, panelists were provided with several pieces of feedback information. With each piece of data, the panelists were reminded that the data were intended to inform their decisions, but not to dictate them.

Panelists were presented with feedback showing their individual test-level cut scores and the committee-level test-level cut scores. The committee-level feedback included the minimum, maximum, mean, and median test-level cut scores for the Emerging, Developing, and Novice Levels, as well as a bar chart reflecting the panelists' cut score agreement for the performance levels. Panelists also received test-level cut scores for their table, which included the same type of statistics shown for the committee-level cut scores, and a summary of the frequency distribution of item scores for each item at each performance level. The panelists' Round 1 Ratings form was redistributed with the Round 1 feedback, so the panelists could refer to their initial ratings as they reviewed the summary of the frequency distribution of the item scores as a table group.

Item mean scores and score profiles were also presented to the panelists. The item mean scores were provided for each operational item and showed the average rubric score obtained by all the students who took the item. The item means were intended to be used to validate panelists' perceptions of item difficulty. The score profiles showed examples of how students achieved total test scores at the recommended whole group cut scores. The profiles were intended to help the panelists think about whether the performance shown in the profile is acceptable for each performance level.

Panelists were instructed to consider how close their recommendations were to those of others in their table group as well as the whole group and discuss why they may have had different ratings for certain items. During the table-level and committee-level discussions, the group tried to determine the factors underlying the variability in recommendations. While panelists were encouraged to reassess their cut recommendations based on these discussions, the main purpose of this activity was to allow panelists to think through and discuss the recommendation process; it was not to arrive at a consensus.

Round 2

After discussing Round 1 feedback and completing the readiness survey for Round 2, panelists worked independently to re-evaluate their recommendations and decide whether they wanted to revise them. During Round 2, the panelists continued to consider the performance level content each item measures, the PLDs, scoring rubrics, and the threshold student descriptors before providing their item-level cut score ratings. As before, panelists were reminded that their recommendations should be grounded in content and what students should know and be able to do, not what they can do or are currently doing. Panelists recorded their Round 2 recommendations on their Round 2 Ratings form (see Appendix H for an example) and submitted it to the facilitator.

Round 2 Feedback

As done previously, several pieces of feedback data were provided based on Round 2 recommendations. Panelists received the same summary statistics as in Round 1, but based on their Round 2 recommendations. Table-level and whole group-level discussions were again had around these data.

For this round, impact data were also provided. Based on Round 2 recommendations, graphs indicating the percentage of students who would score in each of the performance levels were displayed, and the impact data were based on the median test-level cut scores. Fall 2014 test-taker impact was provided, but it was also disaggregated by gender, ethnicity, and socio-economic status. Panelists were asked to discuss whether the percentage of students in each performance level met their expectations given what they know about the population of students tested and the test content.

Impact data corresponding to the CoAlt MS subject-area assessment and CMAS HS subject-area assessment were also shown to panelists during this round. Before being shown the CoAlt MS impact data and CMAS HS impact data, panelists were asked about their expectations regarding the impact data in relation to the CoAlt HS impact data. Both sets of impact data were intended to provide a reasonableness check, but panelists were reminded that any modifications to cut score recommendations should be based on content and not driven by impact data alone.

Round 3

After discussing Round 2 feedback and completing the readiness survey for Round 3, panelists worked independently to again re-evaluate their recommendations. During Round 3, panelists provided their final recommendation as to what the test-level cut score should be for each performance level. Panelists completed their round readiness form for this last round and then recorded their final ratings on the Round 3 Ratings form (see Appendix H for an example) and submitted their completed ratings sheet to the facilitator.

Round 3 Feedback

After completing their Round 3 ratings, panelists were shown their Round 3 feedback. They were shown the committee-level cut score recommendations for each performance level and panelist agreement data. Impact data and score profiles based on their Round 3 ratings were also shown to the panelists and were based on the median test-level cut scores.

Evaluation

After all panelists were finished and final results were determined, panelists were asked to complete a short evaluation. The evaluation asked about panelists' level of comfort with the standard setting procedure, their understanding of the performance levels, and their satisfaction with final cut scores. The standard setting evaluation and results can be found in Appendix I. Upon completing the evaluations, panelists were thanked for their time and participation and dismissed.

ROUND 3 RECOMMENDED CUT SCORES

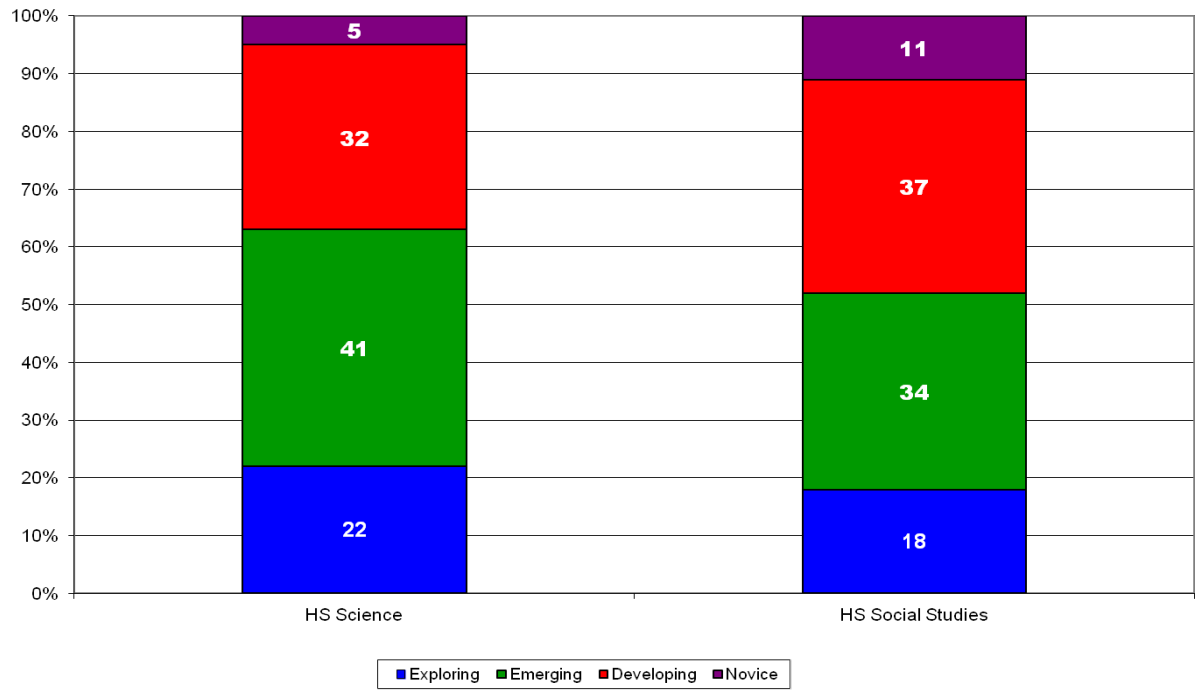
This section provides results from the standard setting portion of the meeting. Table 2 shows the median of panelists' recommendations by round.

Table 2. Panelist Recommendations by Round

		Emerging Level	Developing Level	Novice Level
HS Science	**Round 1**	71	90	106
	Round 2	74	94	107
	Round 3	77	98	107
HS Social Studies	**Round 1**	71	91	101
	Round 2	71	91	103
	Round 3	72	92	103

Based on Round 3 recommendations, Table 3 shows the percentages of students who are estimated to fall into each performance level based on the Fall 2014 administration.

Table 3. Round 3 Estimated Impact Data for Science and Social Studies



APPROVAL OF THE FINAL PERFORMANCE STANDARDS

The proposed recommended cut scores from the CoAlt: Science and Social Studies HS standard setting were presented to the Colorado State Board of Education in February 2015 for review. The Colorado State Board of Education approved the HS science recommended cut scores for one year. The approval of the HS social studies recommended cut scores are currently pending. Table 4 shows the scale score ranges resulting from the approved HS science cut scores and the scale score ranges for the proposed HS social studies cut scores.

Table 4. CoAlt: Science and Social Studies HS Scale Score Ranges

	Exploring Level	Emerging Level	Developing Level	Novice Level
HS Science	0–139	140–163	164–192	193–250
HS Social Studies*	0–137	138–158	159–180	181–250

* These cuts have not been approved by the Colorado State Board of Education

REFERENCES

Cizek, G. J. (2012). *Setting performance standards: Foundations, methods, and innovations* (2nd ed.). New York: Routledge.

Cizek, G. J., Bunch, M. B., & Koons, H. (2004). Setting Performance Standards: Contemporary Methods. *Educational Measurement: Issues and Practice*, 23(4), 31–50.

Hambleton, R. K. & Plake, B.S. (1995). Using an Extended Angoff Procedure to Set Standards on Complex Performance Assessments. *Applied Measurement in Education*, 8, 41–56.

Roeber, E. (2002). *Setting standards on alternate assessments* (Synthesis Report 42). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.

APPENDIX A: SCORING RUBRICS

Selected Response Scoring Rubric

Score Point Selected Response Scoring Rubric	
4	Student responds correctly, independently
3	Student responds correctly after being presented with an additional prompt
2	Student responds correctly after being presented with the correct response
1	Student responds incorrectly
NR	Student does not respond

Supported Performance Task Scoring Rubric

Score Point Supported Performance Task Scoring Rubric (utilized for each of three prompts within each task)	
2	Student responds correctly
1	Student responds incorrectly
NR	Student does not respond

APPENDIX B: PANEL COMPOSITION

Panelist Breakdown by Expertise

	Sig Support Teacher	Special Ed Teacher	Content Expert	Administrator	Higher Ed	Total
HS Science	6	0	2	1	1	**10**
HS Social Studies	5	2	1	2	0	**10**
Total	**11**	**2**	**3**	**3**	**1**	**20**

Panelists Breakdown by School Type

	Charter/Innovation School	Neither Charter nor Innovation School	District Level	Other	Total
HS Science	0	10	0	0	**10**
HS Social Studies	0	7	2	1	**10**
Total	**0**	**17**	**2**	**1**	**20**

Panelists Breakdown by Region

	Total
Denver Metro	**2**
North Central	**1**
Northeast	**1**
Northwest	**0**
Pikes Peak	**3**
Southeast	**1**
Southwest	**0**
West Central	**1**
Omit	**10**
Total	**19**

*Higher Ed participants are not included in this table.

APPENDIX C: PERFORMANCE LEVEL DESCRIPTORS

Colorado Alternate Assessment: High School Science Performance Level Descriptors (PLDs)

Students demonstrate science concepts and skills aligned to the Grade Level Expectations and Extended Evidence Outcomes contained in the Colorado Academic Standards at various performance levels. The performance level descriptors are organized in a manner that assumes students in higher performance levels also demonstrate the concepts and skills within the lower levels. For example, a student at the Emerging Level also demonstrates the concepts and skills included within the Exploring Level.

At Novice Level, with appropriate support, a student can typically:

- Predict the direction or relative speed of an object as a result of force
- Predict action and reaction relationships between moving objects
- Group items based on chemical or physical properties
- Identify causes of observable changes due to chemical reactions
- Identify how like cells form body systems and how body systems work together
- Identify the functions of parts of plant and animal cells
- Identify how plants and organisms convert energy
- Recognize how respiration and digestion within organisms occur
- Identify how natural factors affect a location's climate
- Identify Earth's layers
- Recognize how Earth has changed over time through events such as earthquakes and volcanoes

At Developing Level, with appropriate support, a student can typically:

- Recognize that force acting on an object determines its speed and direction
- Identify elements in common compounds and common compounds created by elements
- Identify examples of energy transformations
- Recognize causes of when a body system is not functioning properly
- Identify parts of plant and animal cells
- Distinguish between inherited and acquired traits and learned and instinctual behaviors
- Identify the relationship between environment and plant-based or meat-based diets
- Recognize impacts of using renewable and non-renewable resources
- Identify how climates affect human behavior and activity
- Identify how human needs are met in space
- Identify required components for space vehicles and how aerospace design impacts space travel

At Emerging Level, with appropriate support, a student can typically:

- Recognize that objects require force to change motion and that the mass of an object is related to the force
- Identify physical and chemical characteristics in common items and processes
- Recognize that atoms, molecules, elements, and compounds are related
- Identify nutritional groups, lifestyle choices, and environmental factors that are healthy and unhealthy to humans
- Identify the basic components needed by plants and organisms to convert energy
- Identify the diet of carnivores, herbivores and omnivores
- Recognize how populations may adapt to environmental changes and identify how human activity can create changes in the ecosystem
- Recognize the effects of using natural resources
- List basic human needs for space travel
- Match scientific tools to their use in weather and space exploration

At Exploring Level, with explicit modeling, a student can typically:

- Identify the fastest object in a group
- Identify common chemical reactions in household items and processes
- Identify sources of different forms of energy
- Recognize symptoms of when a body system is not functioning properly
- Identify similarities and differences in parents and children
- Recognize that ecosystems are affected by human activities
- Recognize that humans use a variety of resources
- Select basic methods of preparing for severe weather conditions or natural hazards
- Identify objects in the Solar System

An Inconclusive designation is given to students who did not respond to any items on the assessment.

Colorado Alternate Assessment: High School Social Studies Performance Level Descriptors (PLDs)

Students demonstrate social studies concepts and skills aligned to the Grade Level Expectations and Extended Evidence Outcomes contained in the Colorado Academic Standards at various performance levels. The performance level descriptors are organized in a manner that assumes students in higher performance levels also demonstrate the concepts and skills within the lower levels. For example, a student at the Emerging Level also demonstrates the concepts and skills included within the Exploring Level.

At Novice Level, with appropriate support, a student can typically:

- Obtain information about historical events and social movements from reliable sources
- Sequence historical events
- Create a graph based on geographical data
- Create a budget
- Differentiate between embargos, tariffs, and subsidies
- Compare multiple perspectives on a current public issue
- Compare information from multiple sources related to a current event
- Identify the leaders and responsibilities of the different branches of the US Government

At Developing Level, with appropriate support, a student can typically:

- Identify sources to answer historical questions
- Identify examples of movements and groups that influenced history
- Identify the correct display and/or draw conclusions about geographical data
- Recognize the purpose of insurance and identify common types of insurance
- Identify the economic outcomes of a surplus or shortage of goods
- Make consumer choices within a fixed dollar amount
- Locate information about current events in reliable sources
- Identify citizen rights and responsibilities in, and characteristics of, a democratic government

At Emerging Level, with appropriate support, a student can typically:

- Identify reliable historical sources
- Recognize symbols, artifacts, and resources related to major historical events
- Contrast activities and lifestyles in regions based on geographic differences
- Recognize displays of geographical data
- Identify ways to save money and responsibilities related to debt
- Recognize the impact of competition in an economic market (e.g., high competition can lead to lower prices)
- Identify individual rights and responsibilities in a society
- Recognize ways individuals can influence public policy

At Exploring Level, with explicit modeling, a student can typically:

- Distinguish between historical and current events
- Match activities to and select appropriate clothing choices for geographic regions and climates
- Identify what should be insured
- Differentiate between a want and a need in relation to a budget
- Identify individual rights and responsibilities within home and school environments

An Inconclusive designation is given to students who did not respond to any items on the assessment.

APPENDIX D: ASSESSMENT FRAMEWORKS

Colorado Academic Standards with Extended Evidence Outcomes Alternate Assessment (CoAlt) Science High School	Approximate % of Score Points	Approximate Score Points
1 Physical Science	**27%**	**30**
1. Newton's laws of motion and gravitation describe the relationships among forces acting on and between objects, their masses, and changes in their motion – but have limitations		
I. Gather, record and interpret data about speed and direction of moving objects		
II. Demonstrate that objects with greater mass require greater force to initiate or change movement		
III. Predict action-reaction relationships between moving objects		
2. Matter has definite structure that determines characteristic physical and chemical properties		
I. Gather data and justify grouping of objects or materials based on chemical and physical properties (e.g. melting point, boiling, conductivity)		
II. Demonstrate that different ratios of substances can be combined to create unique mixtures		
III. Explore the relationship between atoms, molecules, elements and compounds		
3. Matter can change form through chemical or nuclear reactions abiding by the laws of conservation of mass and energy		
I. Identify common household products or processes that use chemical reactions		
4. Atoms bond in different ways to form molecules and compounds that have definite properties		
I. Demonstrate how two or more objects can be connected together to create a different product or outcome		
II. Demonstrate that different bonding agents have different properties		
5. Energy exists in many forms such as mechanical, chemical, electrical, radiant, thermal, and nuclear, that can be quantified and experimentally determined		
I. Describe ways in which nonliving objects get energy		
II. Identify a source for each type of energy (heat, sound, light, mechanical, electrical)		
III. Predict types of energy associated with objects (heat, sound, light, mechanical, electrical)		
6. When energy changes form, it is neither created not destroyed; however, because some is necessarily lost as heat, the amount of energy available to do work decreases		
I. Predict and experiment with energy transformations		
II. Select examples of energy transformations		
2 Life Science	**42%**	**46**
1. Matter tends to be cycled within an ecosystem, while energy is transformed and eventually exits an ecosystem		
I. Compare and contrast carnivores, herbivores and omnivores		
2. The size and persistence of populations depend on their interactions with each other and on the abiotic factors in an ecosystem		
I. Compare and contrast positive and/or negative impacts humans have on our ecosystem		
II. Describe what happens when an organisms area is destroyed or disturbed		

Colorado Academic Standards with Extended Evidence Outcomes Alternate Assessment (CoAlt) Science High School	Approximate % of Score Points	Approximate Score Points
3. Cellular metabolic activities are carried out by biomolecules produced by organisms		
I. Demonstrate how like cells group together to make a structure		
II. Identify common food sources of fats, carbohydrates and proteins		
III. Explain how lifestyle choices impact the body		
4. The energy for life primarily derives from the interrelated processes of photosynthesis and cellular respiration. Photosynthesis transforms the sun's light energy into the chemical energy of molecular bonds. Cellular respiration allows cells to utilize chemical energy when these bonds are		
I. Identify chemical reactions within organisms (respiration and digestion ensure survival)		
II. Describe three components in the environment that are necessary for photosynthesis (sunlight, water, nutrients) and what occurs when one component is lacking		
5. Cells use the passive and active transport of substances across membranes to maintain relatively stable intracellular environments		
I. Identify common symptoms that show when a body system is not functioning properly		
II. Identify the three major components of a plant or animal cell (nucleus, cell membrane/cell wall and cytoplasm)		
6. Cells, tissues, organs, and organ systems maintain relatively stable internal environments, even in the face of changing external environments		
I. Describe how two organ systems work together to promote health		
II. Identify two or more health decisions influencing organ health		
7. Physical and behavioral characteristics of an organism are influenced to varying degrees by heritable genes, many of which encode instructions for the production of proteins		
I. Compare and contrast the inheritable traits between parents and their offspring (single allele such as tongue rolling, ear lobes, hitchhikers thumb, widows peak, long second toe)		
II. Identify learned versus instinctual behaviors		
8. Multicellularity makes possible a division of labor at the cellular level through the expression of select genes, but not the entire genome		
I. Indentify environmental toxins that are harmful to humans		
9. Evolution occurs as the heritable characteristics of populations change across generations and can lead populations to become better adapted to their environment		
I. Identify changes in the environment over time that have driven adaptations of living things		
3 Earth Systems Science	**31%**	**34**
1. The history of the universe, solar system and Earth can be inferred from evidence left from past events		
I. Identify ways how the Earth has changed over time to accommodate a variety of life forms (sea life, dinosaurs, land animals, mammals)		
II. Identify the Solar System as having formed around the sun		
2. As part of the solar system, Earth interacts with various extraterrestrial forces and energies such as gravity, solar phenomena, electromagnetic radiation, and impact events that influence the planet's geosphere, atmosphere, and biosphere in a variety of ways		

Colorado Academic Standards with Extended Evidence Outcomes Alternate Assessment (CoAlt) Science High School	Approximate % of Score Points	Approximate Score Points
I. Identify how aerospace design impacts space travel (e.g. Where you can go on an airplane vs where you can go on a space shuttle)		
II. Describe ways in which basic needs can be met in space compared to needs on Earth (e.g. air, water, heat, food)		
3. The theory of plate tectonics helps to explain geological, physical, and geographical features of Earth		
I. Identify and locate places on Earth where earthquakes and volcanoes occur		
II. Label the layers of the Earth (inner core, core, mantle and crust)		
4. Climate is the result of energy transfer among interactions of the atmosphere, hydrosphere, geosphere, and biosphere		
I. Describe how climate affects humans		
II. Explain how human behavior affect climate		
III. Identify Earth's tilt, seasons, elevation, proximity to oceans as factors that determine a location's climate		
IV. Use tools to measure temperature, wind, precipitation and then analyze information from the sources about climate change		
5. There are costs, benefits, and consequences of exploration, development, and consumption of renewable and nonrenewable resources		
I. Determine the effects of using natural resources		
II. Compare the advantages and disadvantages of renewable and non-renewable resources		
6. The interaction of Earth's surface with water, air, gravity, and biological activity causes physical and chemical changes		
I. Identify the properties of gravity		
II. Investigate how human activity can cause physical and chemical changes in water and air		
7. Natural hazards have local, national and global impacts such as volcanoes, earthquakes, tsunamis, hurricanes, and thunderstorms		
I. Identify impacts of natural hazards (blizzard, tornado, flood, volcanoes, fire and earthquakes)		
II. Select appropriate ways to prepare for natural hazards (blizzards, tornadoes, floods)		
TOTAL	**100%**	**110**

Colorado Academic Standards with Extended Evidence Outcomes Alternate Assessment (CoAlt) Social Studies High School	Approximate % of Score Points	Approximate Score Points
1 History	**27.3%**	**30**
1. Use the historical method of inquiry to ask questions, evaluate primary and secondary sources, critically analyze and interpret data, and develop interpretations defended by evidence		
I. Obtain information about a historical event using multiple reliable sources		
2. The key concepts of continuity and change, cause and effect, complexity, unity and diversity over time		
I. Create a timeline using symbols including four major events in modern world history (Industrial Revolution to present)		
II. Explore different movements and groups that influenced world history		
III. Create a timeline using symbols including four major events in United States history (Industrial Revolution to present)		
3. The significance of ideas as powerful forces throughout history		
I. Explore historical ideas as related to religion, social movements, and civil rights		
2 Geography	**27.3%**	**30**
1. Use different types of maps and geographic tools to analyze features on Earth to investigate and solve geographic questions		
I. Create a simple graph based on geographical data (e.g., population, geographical features, etc.)		
2. Explain and interpret geographic variables that influence the interactions of people, places and environments		
I. Identify regional differences in lifestyle in parts of the United States due to geographic differences		
3. The interconnected nature of the world, its people and places		
I. Describe ways groups of people are the same and different across regions of the world		
3 Economics	**27.3%**	**30**
1. Productive resources – natural, human, capital – are scarce; therefore, choices are made about how individuals, businesses, governments, and societies allocate these resources		
I. Explain what happens when there is not enough of a good or service		
2. Economic policies affect markets		
I. Explore embargos, tariffs and subsidies		
3. Government and competition affect markets		
I. Identify competition in the local economic market (e.g., cars, groceries, apartments, etc)		
4. Design, analyze, and apply a financial plan based on short- and long-term financial goals (PFL)		
I. Develop a simple monthly budget		
5. Analyze strategic spending, saving, and investment options to achieve the objectives of diversification, liquidity, income, and growth (PFL)		
I. Identify ways to save money		
6. The components of personal credit to manage credit and debt (PFL)		

Colorado Academic Standards with Extended Evidence Outcomes Alternate Assessment (CoAlt) Social Studies High School	Approximate % of Score Points	Approximate Score Points
I. Demonstrate an understanding of personal responsibility related to debt		
7. Identify, develop, and evaluate risk-management strategies (PFL)		
I. Explore types and purposes of insurance		
4 Civics	**18%**	**20**
1. Research, formulate positions, and engage in appropriate civic participation to address local, state, and national issues or policies		
I. Compare information from multiple sources related to a current event (local, state or national)		
II. Engage in activities as a responsible public citizen		
2. Purposes of and limitations on the foundations, structures and functions of government		
I. Identify personal rights and responsibilities of self and others		
II. Identify the responsibilities of the three branches of government (i.e. judicial = interprets, legislative = makes, executive = enforces)		
3. Analyze how public policy - domestic and foreign - is developed at the local, state, and national levels and compare how policy-making occurs in other forms of government		
I. Identify ways individuals can influence public policy		
II. Identify the multiple perspectives of a current public issue		
TOTAL	**100%**	**110**

APPENDIX E: PERFORMANCE LEVEL CONTENT MATCH FORM

Colorado Alternate Assessment (CoAlt)

Assessment: _____

Panelist ID: _____

Table Number: _____

Performance Level Content Match

Instructions: Review each item/task in the test book after reviewing the performance level descriptors (PLDs). For each item/task, write whether the item/task is measuring **Novice**, **Developing**, **Emerging**, or **Exploring** level content based on the concepts and skills found in the PLDs.

Item/Task	Item Type	Performance Level Content
Item 1	SR	
Item 2	SR	
Item 4	SR	
Task 5	SPT	
Item 6	SR	
Item 7	SR	
Item 8	SR	
Item 9	SR	
Item 10	SR	
Task 12	SPT	
Item 13	SR	
Item 14	SR	
Item 15	SR	
Item 16	SR	
Item 17	SR	
Item 18	SR	
Item 20	SR	
Item 21	SR	
Item 22	SR	
Item 23	SR	
Task 24	SPT	
Item 25	SR	
Item 26	SR	
Item 28	SR	
Item 29	SR	
Item 30	SR	

APPENDIX F: PRACTICE EXERCISE RATINGS SHEET

Colorado Alternate Assessment (CoAlt)

Assessment: _____

Panelist ID: _____

Table Number: _____

Practice Exercise Ratings

Instructions: For each item/task, write your item-level cut score recommendation for each performance level in the appropriate box.

If a SR item measures content below the performance level cut you are recommending, the item should receive a rating of 4 based on the assumption of performance described in the PLDs.

If a SPT item measures content below the performance level cut you are recommending, the item should receive a rating of 6 based on the assumption of performance described in the PLDs.

			Practice Exercise		
Item/Task	**Item Type**	**Performance Level Content**	**Emerging**	**Developing**	**Novice**
			SR score range:2–4 SPT score range: 3–6	SR score range:2–4 SPT score range:3–6	SR score range:2–4 SPT score range:3–6
Item 1	SR	Exploring			
Item 2	SR	Exploring			
Task 3	SPT	Emerging			
Item 4	SR	Emerging			
Item 5	SR	Developing			
Item 6	SR	Novice			

APPENDIX G: READINESS SURVEY

Panelist ID: _____

Instructions: Please circle your response to the following questions.

Round 1		
I understand that my task for Round 1 is to use the assessed content, my experience with CoAlt students, the scoring rubrics, and the threshold student descriptors to make item-level cut score recommendations. To make my recommendations, I will write my item-level scores on the ratings sheet.	**No**	**Yes**
I am ready to begin Round 1.	**No**	**Yes**

Round 2		
I understand that my task for Round 2 is to use the assessed content, my experience with CoAlt students, the scoring rubrics, and the threshold student descriptors to make item-level cut score recommendations. To make my recommendations, I will write my item-level scores on the ratings sheet.	**No**	**Yes**
I understand the panelist feedback data that were presented from Round 1.	**No**	**Yes**
I understand the item mean scores that were provided.	**No**	**Yes**
I understand the score profiles that were provided.	**No**	**Yes**
I am ready to begin Round 2.	**No**	**Yes**

Round 3		
I understand that my task for Round 3 is to use the assessed content, my experience with CoAlt students, the scoring rubrics, and the threshold student descriptors to make test-level cut score recommendations. To make my recommendations, I will write my test-level scores on the ratings sheet.	**No**	**Yes**
I understand the impact data that were presented from Round 2.	**No**	**Yes**
I understand the score profiles that were provided.	**No**	**Yes**
I am ready to begin Round 3.	**No**	**Yes**

APPENDIX H: SAMPLE RATINGS FORMS

Assessment: _____

Panelist ID: _____

Table Number: _____

Round 1 Ratings

Instructions: For each item/task, write your item-level cut score recommendation for each performance level in the appropriate box.

If a SR item measures content below the performance level cut you are recommending, the item should receive a rating of 4 based on the assumption of performance described in the PLDs.

If a SPT item measures content below the performance level cut you are recommending, the item should receive a rating of 6 based on the assumption of performance described in the PLDs.

			Round 1		
Item/Task	**Item Type**	**Performance Level Content**	**Emerging**	**Developing**	**Novice**
			SR score range:2–4 SPT score range: 3–6	SR score range:2–4 SPT score range:3–6	SR score range:2–4 SPT score range:3–6
Item 1	SR	Exploring			
Item 2	SR	Emerging			
Task 3	SPT	Exploring			
Item 4	SR	Exploring			
Item 5	SR	Emerging			
Item 6	SR	Exploring			
Item 7	SR	Emerging			
Item 8	SR	Emerging			
Item 9	SR	Emerging			
Item 10	SR	Emerging			
Item 11	SR	Emerging			
Item 12	SR	Emerging			
Item 13	SR	Emerging			
Item 14	SR	Emerging			
Item 15	SR	Emerging			
Item 16	SR	Developing			
Item 17	SR	Developing			
Item 18	SR	Developing			
Item 19	SR	Developing			
Item 20	SR	Novice			
Item 21	SR	Developing			
Task 22	SPT	Novice			
Task 23	SPT	Developing			

Item 24	SR	Novice			
Item 25	SR	Novice			
Item 26	SR	Novice			

Assessment: _____

Panelist ID: _____

Table Number: _____

Round 2 Ratings

Instructions: For each item/task, write your item-level cut score recommendation for each performance level in the appropriate box.

If a SR item measures content below the performance level cut you are recommending, the item should receive a rating of 4 based on the assumption of performance described in the PLDs.

If a SPT item measures content below the performance level cut you are recommending, the item should receive a rating of 6 based on the assumption of performance described in the PLDs.

			Round 2		
Item/Task	**Item Type**	**Performance Level Content**	**Emerging**	**Developing**	**Novice**
			SR score range:2–4 SPT score range: 3–6	SR score range:2–4 SPT score range:3–6	SR score range:2–4 SPT score range:3–6
Item 1	SR	Exploring			
Item 2	SR	Emerging			
Task 3	SPT	Exploring			
Item 4	SR	Exploring			
Item 5	SR	Emerging			
Item 6	SR	Exploring			
Item 7	SR	Emerging			
Item 8	SR	Emerging			
Item 9	SR	Emerging			
Item 10	SR	Emerging			
Item 11	SR	Emerging			
Item 12	SR	Emerging			
Item 13	SR	Emerging			
Item 14	SR	Emerging			
Item 15	SR	Emerging			
Item 16	SR	Developing			
Item 17	SR	Developing			
Item 18	SR	Developing			
Item 19	SR	Developing			
Item 20	SR	Novice			
Item 21	SR	Developing			
Task 22	SPT	Novice			
Task 23	SPT	Developing			
Item 24	SR	Novice			
Item 25	SR	Novice			
Item 26	SR	Novice			

Assessment: _____

Panelist ID: _____

Table Number: _____

Round 3 Ratings

Instructions: Please write your test-level Emerging cut score, Developing cut score, and Novice cut score recommendations in the appropriate box.

My Cut Score Recommendations		
Emerging Cut Score Recommendation	**Developing Cut Score Recommendation**	**Novice Cut Score Recommendation**

APPENDIX I: STANDARD SETTING EVALUATION

Colorado Alternate Assessment (CoAlt)
Standard Setting Evaluation Form

The purpose of this evaluation form is to collect information about your experience in recommending performance cut scores for CoAlt. Your opinions provide an important part of our evaluation of this meeting. Please do not write your name on this evaluation form as we want your comments to be anonymous. Thank you for your willingness to participate in this survey.

In which standard setting meeting did you participate?

- ☐ High School Science
- ☐ High School Social Studies

Indicate your response by checking the appropriate box.

	Do not support	Support with some reservation	Moderately support	Strongly support
1. To what degree do you support the recommended cut score for "Emerging Level?"	☐	☐	☐	☐
HS science	0%	10%	0%	90%
HS social studies	0%	10%	40%	50%
If you cannot support, please explain why not:				
2. To what degree do you support the recommended cut score for "Developing Level?"	☐	☐	☐	☐
HS science	0%	0%	10%	90%
HS social studies	0%	0%	40%	60%
If you cannot support, please explain why not:				
3. To what degree do you support the recommended cut score for "Novice Level?"	☐	☐	☐	☐
HS science	0%	0%	0%	100%
HS social studies	0%	0%	10%	90%
If you cannot support, please explain why not:				

	Way too low	A bit low	Appropriate	A bit high	Way too high
4. The recommended cut score for "Emerging Level" is:	☐	☐	☐	☐	☐
HS science	0%	10%	90%	0%	0%
HS social studies	10%	40%	50%	0%	0%
5. The recommended cut score for "Developing Level" is:	☐	☐	☐	☐	☐
HS science	0%	10%	90%	0%	0%
HS social studies	0%	10%	80%	10%	0%
6. The recommended cut score for "Novice Level" is:	☐	☐	☐	☐	☐
HS science	0%	0%	100%	0%	0%
HS social studies	0%	10%	80%	10%	0%

	Strongly Disagree	Disagree	Agree	Strongly Agree
7. The Modified Extended Angoff Method was explained clearly by the group facilitator.	☐	☐	☐	☐
HS science	0%	0%	40%	60%
HS social studies	0%	0%	80%	20%
8. I had a solid understanding of what the test was intended to measure.	☐	☐	☐	☐
HS science	0%	0%	30%	70%
HS social studies	0%	0%	50%	50%
9. I could clearly distinguish between performance levels.	☐	☐	☐	☐
HS science	0%	0%	50%	50%
HS social studies	0%	0%	60%	40%
10. After the first round of recommendations, I felt comfortable with the standard setting procedure.	☐	☐	☐	☐
HS science	0%	0%	30%	70%
HS social studies	0%	0%	40%	60%
11. I found the feedback on the comparison of all panelists' recommendations to be useful in standard setting.	☐	☐	☐	☐
HS science	0%	0%	40%	60%
HS social studies	0%	0%	30%	70%
12. I found the item mean score information to be useful in standard setting.	☐	☐	☐	☐
HS science	0%	0%	50%	50%
HS social studies	0%	0%	50%	50%
13. I found the score profile information to be useful in standard setting.	☐	☐	☐	☐
HS science	0%	10%	30%	60%
HS social studies	0%	0%	40%	60%
14. I found the feedback on the percentage of the students tested that would be classified at each performance level to be useful in standard setting.	☐	☐	☐	☐
HS science	0%	0%	40%	60%
HS social studies	0%	0%	20%	80%
15. Table and group discussions were open and honest.	☐	☐	☐	☐
HS science	0%	0%	10%	90%
HS social studies	0%	0%	20%	80%

		Strongly Disagree	Disagree	Agree	Strongly Agree
16. I believe that my opinions were considered and valued by my group.		☐	☐	☐	☐
	HS science	0%	0%	20%	80%
	HS social studies	0%	0%	20%	80%
17. The facilitator led the group through the standard setting process without imposing ideas about where cut scores should be.		☐	☐	☐	☐
	HS science	0%	0%	30%	70%
	HS social studies	0%	0%	10%	90%
18. I am confident that the final cut score recommendations reflect the performance level descriptors associated with CoAlt.		☐	☐	☐	☐
	HS science	0%	0%	20%	80%
	HS social studies	0%	0%	60%	40%
19. I am confident that the final cut score recommendations reflect high expectations consistent with the Extended Evidence Outcomes of the Colorado Academic Standards.		☐	☐	☐	☐
	HS science	0%	0%	20%	80%
	HS social studies	0%	0%	50%	50%

Please use the back of this page to provide any additional comments.

Additional evaluation questions asked during standard setting:

I found the performance level matching activity useful in standard setting.		Strongly Disagree	Disagree	Agree	Strongly Agree
	HS science	0%	0%	20%	80%
	HS social studies	0%	0%	10%	90%

Familiarity with science content	Not Familiar	Below Average	Average	Above Average	Greatly Familiar
Physical science	0%	0%	70%	10%	20%
Life science	0%	0%	40%	30%	30%
Earth science	0%	10%	30%	30%	30%

Familiarity with social studies content	Not Familiar	Below Average	Average	Above Average	Greatly Familiar
History	0%	0%	60%	10%	30%
Geography	0%	0%	30%	60%	10%
Economics	0%	10%	50%	30%	10%
Civics	0%	0%	70%	30%	0%

APPENDIX D: COALT: SCIENCE AND SOCIAL STUDIES SAMPLE SCORE REPORTS

Student Performance Report

Colorado Alternate Assessment

Student: **FIRSTNAM16 G. LASTNAME16**

SASID: **1660000007** Birthdate: **07/07/2006**
School: **SAMPLE001 SCHOOL ONE (7103)**
District: **SAMPLE001 DISTRICT (7003)**

Spring 2015

Social Studies — Grade 7

This score report provides information about your student's performance on the Colorado Alternate (CoAlt) Social Studies Assessment.

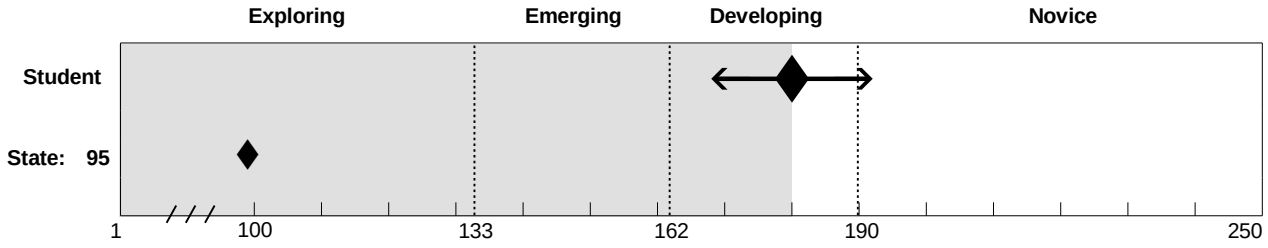
- Your student's performance is represented by a scale score. Scores are placed on a scale so that student performance can be compared across years.
- State averages are provided so that you can compare your student's performance to the performance of others. The percentage of students in each performance level across the state is reported below the graph.
- Scores are represented by diamonds. The arrows around the student's diamond show the range of scores that your student would likely receive if the assessment was taken multiple times.
- Dotted lines show where the range of scores is divided into performance levels. Descriptions of the performance levels can be found at the end of this report.

Social Studies

Your Student's Score

180

Developing

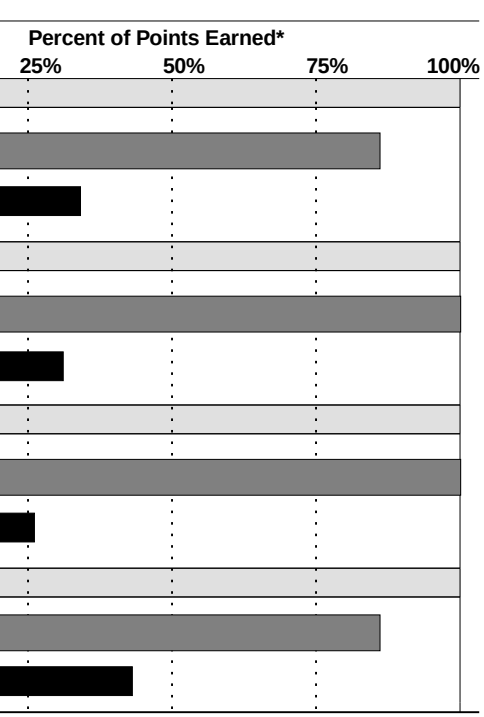


Percent of CO students by Performance Level: 0.0% 0.0% 0.0% 0.0%

The Extended Evidence Outcomes of the Colorado Academic Standards include expectations for student performance. Your student demonstrates a developing understanding of 7th grade level concepts and skills in social studies.

Content Standard Performance

Content Standard Description	Points Earned	Points Possible	Percent of Points Earned*
History			
History develops moral understanding, defines identity and creates an appreciation of how things change while building skills in judgment and decision-making. History enhances the ability to read varied sources and develop the skills to analyze, interpret and communicate.	19	22	86% / 34%
Geography			
Geography provides students with an understanding of spatial perspectives and technologies for spatial analysis, awareness of interdependence of world regions and resources and how places are connected at local, national and global scales.	16	16	100% / 31%
Economics			
Economics teaches how society manages its scarce resources, how people make decisions, how people interact in the domestic and international markets, and how forces and trends affect the economy as a whole. Personal financial literacy applies the economic way of thinking to help individuals understand how to manage their own scarce resources.	12	12	100% / 26%
Civics			
Civics teaches the complexity of the origins, structure, and functions of governments; the rights, roles and responsibilities of ethical citizenship; the importance of law; and the skills necessary to participate in all levels of government.	19	22	86% / 43%



*The percent of points earned cannot be compared across years because individual items change from year to year. They also cannot be compared across Standards because the number of items and the difficulty of items may not be the same.

Student's Score State Average

Purpose

This report describes your student's mastery of the Extended Evidence Outcomes of the Colorado Academic Standards in Social Studies.

For more information on the CoAlt assessment program, visit:
www.cde.state.co.us/assessment/coaltassess

Social Studies Performance Level Descriptions

Students demonstrate social studies concepts and skills aligned to the Grade Level Expectations and Extended Evidence Outcomes contained in the Colorado Academic Standards.

At Novice Level, with appropriate support, a student can typically:

- Determine appropriate questions to ask in order to learn about specific historical events
- Compare information from multiple sources related to a significant historical event
- Identify the best source of information regarding a historical event and use a historical event to match a source with a particular perspective
- Match natural resources with ancient communities and their dwellings
- Use a map to determine where to go for a specific purpose and to determine the direction in which to travel from one point to another
- Estimate the total purchase price of an item with sales tax included
- Recognize how supply and demand can affect price
- Recognize rights and responsibilities of citizens

At Developing Level, with appropriate support, a student can typically:

- Match artifacts with their ancient culture or location within the Eastern Hemisphere
- Select the appropriate source of information to answer questions surrounding historical events
- Recognize that sources have different purposes
- Use map symbols and directionality words to locate places on a map
- Recognize that communities were built near natural resources
- Identify the environmental resources that influenced settlement in the Eastern Hemisphere
- Recognize that the total purchase price of an item will increase because of sales tax
- Identify community needs or services that are paid for by taxes
- Differentiate between laws and rules
- Identify the positive and negative consequences of obeying laws and rules

At Emerging Level, with appropriate support, a student can typically:

- Recognize significant artifacts related to ancient civilizations of the Eastern Hemisphere
- Select the appropriate source of information to answer social studies questions
- Identify the appropriate questions to ask in order to learn more about an event or era
- Use symbols to identify a location on a map
- Identify reasons goods and services might go on sale
- Identify ways in which countries and nations resolve differences
- Recognize local laws, state laws, and federal laws and identify examples of following these laws/rules

At Exploring Level, with explicit modeling, a student can typically:

- Recognize artifacts
- Identify part(s) of a map (e.g., title, key, compass rose, scale)
- Recognize there are different types of informational resources
- Recognize that areas have different natural resources
- Recognize that many items have a sales tax
- Recognize that all countries have laws

An Inconclusive designation is given to students who did not respond to any items on the assessment.

For more information about the standards included in this assessment, please visit the Colorado Department of Education's website at
www.cde.state.co.us/standardsandinstruction

State Performance Level Summary

Colorado Alternate Assessment

Spring 2015

Science

Grade 8

Purpose: This report describes group achievement in terms of performance levels.

	Number of Valid Scores	Average Scale Score	Performance Levels								Developing and Novice		No Scores Reported	Total Number of Students
			Exploring		Emerging		Developing		Novice					
			#	%	#	%	#	%	#	%	#	%	#	#
State	**607**	**152**	**92**	**15.2%**	**268**	**44.2%**	**171**	**28.2%**	**76**	**12.5%**	**247**	**40.7%**	**40**	**647**
Gender														
Female	235	151	35	14.9%	103	43.8%	72	30.6%	25	10.6%	97	41.3%	9	244
Male	372	153	57	15.3%	165	44.4%	99	26.6%	51	13.7%	150	40.3%	31	403
Ethnicity/Race														
Hispanic or Latino	210	156	25	11.9%	90	42.9%	63	30.0%	32	15.2%	95	45.2%	4	214
American Indian or Alaska Native	<16	*	*		*		*		*		*		*	*
Asian	<16	*	*		*		*		*		*		*	*
Black or African-American	40	163	1	2.5%	22	55.0%	10	25.0%	7	17.5%	17	42.5%	5	45
White	302	149	58	19.2%	131	43.4%	80	26.5%	33	10.9%	113	37.4%	26	328
Native Hawaiian or Other Pacific Islander	<16	*	*		*		*		*		*		*	*
Two or more races	23	142	5	21.7%	10	43.5%	6	26.1%	2	8.7%	8	34.8%	1	24
Not Indicated	<16	*	*		*		*		*		*		*	*
Language Background														
English	470	150	84	17.9%	199	42.3%	130	27.7%	57	12.1%	187	39.8%	33	503
Spanish	104	158	6	5.8%	53	51.0%	30	28.8%	15	14.4%	45	43.3%	3	107
Other	18	157	0	0.0%	10	55.6%	7	38.9%	1	5.6%	8	44.4%	0	18
Not Indicated	<16	*	*		*		*		*		*		*	*
Language Proficiency														
Not Applicable	469	150	84	17.9%	199	42.4%	130	27.7%	56	11.9%	186	39.7%	34	503
NEP	96	157	5	5.2%	50	52.1%	29	30.2%	12	12.5%	41	42.7%	2	98
LEP	<16	*	*		*		*		*		*		*	*
NEP and LEP	108	158	6	5.6%	53	49.1%	34	31.5%	15	13.9%	49	45.4%	2	110
FEP	<16	*	*		*		*		*		*		*	*
PHLOTE	<16	*	*		*		*		*		*		*	*
FELL	<16	*	*		*		*		*		*		*	*
Not in ELL Program	499	151	86	17.2%	215	43.1%	137	27.5%	61	12.2%	198	39.7%	38	537
Not Indicated	<16	*	*		*		*		*		*		*	*

*No data are reported when number of valid scores is less than 16 per group.

Content Standards Roster

Colorado Alternate Assessment

Spring 2015

School: **SAMPLE001 SCHOOL ONE (7103)**
District: **SAMPLE001 DISTRICT (7003)**

Science

Grade 8

CONFIDENTIAL - DO NOT DISTRIBUTE

Purpose: This report presents each student's performance on the overall test and content standards for your school or district.

Performance Level	Scale Score Ranges
Novice	190 - 250
Developing	164 - 189
Emerging	128 - 163
Exploring	1 - 127

	Overall Scale Score	Content Standards Performance		
		Physical Science	Life Science	Earth Systems Science
		Points Possible		
		28	30	50
		Percent of Points Earned		
State Average	90	38%	28%	36%
District Average	116	41%	45%	43%
School Average	56	9%	9%	10%

	STUDENT NAME	Overall Performance Level				
1	LASTNAME20, FIRSTNAM20 R.	Exploring	49	7%	7%	8%
2	LASTNAME21, FIRSTNAM21 Z.	Exploring	62	11%	10%	12%

Note: Students with no scores are not included in summary calculations.