Colorado Alternate Assessment:
Science and Social Studies

Technical Report

Spring
2014

Table of Contents

Table of Contents i

Part I: Historical Overview and Summary of Processes 3

CHAPTER 1: INTRODUCTION AND BACKGROUND 3
Purpose of the Document 3
Overview of CoAlt 3
Assessment Development Partners 7
CHAPTER 2: ITEM DEVELOPMENT AND ITEM BANKING 9
Item-Writing Process 9
Item Reviews 10
Item Banking Process 11
CHAPTER 3: TEST CONSTRUCTION 13
CHAPTER 4: TEST ADMINISTRATION PROCEDURES 15
Manuals 15
Training 15
Accessibility and Accommodations 15
Test Security 16
CHAPTER 5: SCORING THE ASSESSMENTS 17
CHAPTER 6: STANDARD SETTING 20
CHAPTER 7: REPORTING 22
Description of Scores 22
Score Reports 23
CHAPTER 8: CALIBRATION, EQUATING, AND SCALING 24
Calibration 24
Equating and Scaling 24
Steps in the Calibration and Scaling Process 25
CHAPTER 9: RELIABILITY 27
Cronbach's Alpha 27
Standard Error of Measurement 28
Conditional Standard Error of Measurement 28
Decision Consistency and Accuracy 28
Inter-Rater Agreement 29
CHAPTER 10: VALIDITY 30
Sources of Validity Evidence 30
Fairness 32

Part II: Statistical Summaries for Spring 2014 33

CHAPTER 1: SPRING 2014 OPERATIONAL ITEMS 34
Administration Summary 34
Calibration Results 34
Performance Results 34
Validity Evidence 35
CHAPTER 2: SPRING 2014 EMBEDDED FIELD-TEST ITEMS 36
Field-Test Items 36
Data Review 36

References 37

CoAlt: Science and Social Studies Tables 4–33 38

Appendices 89

APPENDIX A: COALT: SCIENCE AND SOCIAL STUDIES ELIGIBILITY GUIDELINES 90
APPENDIX B: COALT: SCIENCE AND SOCIAL STUDIES TEST BLUEPRINTS 93

APPENDIX C: COALT: SCIENCE AND SOCIAL STUDIES STANDARD-SETTING REPORT.....95
APPENDIX D: COALT: SCIENCE AND SOCIAL STUDIES SAMPLE SCORE REPORTS.....152

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 Spring 2014 CoAlt: Science and Social Studies assessments, including content, assessment development, administration, scoring, and technical attributes.

The Spring 2014 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 Spring 2014 administration, 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 response 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 first operational administration of CoAlt: Science and Social Studies was April 14, 2014 to May 2, 2014. The following assessments were administered during the assessment window:

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

High school CoAlt: Science and Social Studies assessments were administered for the first time in November 2014. Therefore, technical attributes for those assessments will be provided on a delayed timeline. This technical report only pertains to the Spring 2014 CoAlt: Science and Social Studies Elementary School and Middle School (ES/MS) administration.

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 ES/MS 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 ES/MS assessments began in Summer 2012. The newly-developed items were then administered in a stand-alone field test in Spring 2013. The goal of the stand-alone field test 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. In addition, input and advice is 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 ES/MS 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 ES/MS 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 ES/MS items, selected items were administered in a stand-alone field test in Spring 2013. 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 Spring 2013 CoAlt: Science and Social Studies ES/MS 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 is 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 includes 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 for the Spring 2014 administration is shown in Table 1.

Table 1. Spring 2014 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

CHAPTER 4: TEST ADMINISTRATION PROCEDURES

This chapter provides information related to the Spring 2014 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 Spring 2014 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 eight 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 which 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, 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 ES/MS assessments is described in terms of four performance levels: Novice, Developing, Emerging, and Exploring. After the first operational administration of the CoAlt: Science and Social Studies ES/MS assessments in Spring 2014, a standard setting meeting was held with Colorado educators to determine the 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 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 during standard setting to help panelists think about whether the student performance at and around the recommended cut scores made sense for the performance levels.

The standard setting meeting included approximately nine panelists for each grade-level 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 K–12 educators, such as educators with experience working with students with significant cognitive disabilities, educators with experience working with students with other types of disabilities, and content experts with knowledge of the grade-level curriculum. In addition to classroom teachers, special education administrators also participated in the meeting.

The standard setting for the CoAlt: Science and Social Studies ES/MS assessments was held on July 17–18, 2014. During the two-day meeting, panelists from each of the four standard setting committees received training on the assessment and the standard setting process, reviewed the grade-level PLDs, reviewed the Spring 2014 operational items, developed and reviewed threshold student descriptors, and applied the Modified Extended Angoff method to establish cut score recommendations across three rounds of rating. On the afternoon of the second day, a vertical articulation meeting was held. During the meeting, panelists were allowed to review the cut scores set by each grade-level committee and make adjustments if necessary. After the completion of vertical articulation, the recommendations were reviewed by CDE.

As part of this review, CDE reviewed the cut score recommendations from vertical articulation and considered additional information when evaluating the cut scores. The review and consideration given to this additional information was used to determine the Department recommended cut scores. On August 11, 2014, CDE convened a half-day meeting with each subject-area standard setting committee to discuss the Department's adjustments to the cut scores and the rationales for the adjustments. The proposed recommended cut scores from this meeting

were presented to the State Board of Education for review. On August 13, 2014, the Colorado State Board of Education reviewed the recommendations and approved the Department recommended cut scores for the CoAlt: Science and Social Studies ES/MS assessments.

For grades 4 and 7 social studies, an estimated 41% and 38% of students, respectively, were in the top two performance levels (Novice Level and Developing Level). For grades 5 and 8 science, an estimated 44% and 41% of students, respectively, were in the top two performance levels. More details about the CoAlt: Science and Social Studies 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 2014 Score Interpretive Guide at http://www.cde.state.co.us/assessment/newassess-coaltsss.

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
- State 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 ES/MS 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 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 Spring 2014 administration of the CoAlt: Science and Social Studies assessments represents the first operational tests on the newly developed Rasch scale. In the following years, equating will be used to place the new test forms on this newly-developed operational scale. To obtain Rasch item parameter estimates for the Spring 2014 ES/MS assessments, the RPCM was applied to the operational and embedded field-test items. Winsteps (Linacre, 2011) was used for all grade-level calibrations. The calibration 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 fixed common item parameter calibration. 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.

Ability Estimates

After the item parameter estimates were obtained for the Spring 2014 operational items for each grade-level assessment, student proficiencies were estimated 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 grade-level 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

For the Spring 2014 administration, 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 Spring 2014 administration, 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 4–12.

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 4–12.

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 27–30.

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 31.

Inter-Rater Agreement

An additional form of reliability, called inter-rater agreement, will also be assessed for future 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 will simultaneously observe a student taking the CoAlt assessment. Both raters will evaluate student performance and enter their scores into the online score entry system. The two independent ratings will then be compared to determine the consistency of the ratings. The metrics that will be tracked and reported are the correlation between the two independent ratings, perfect agreement and adjacent agreement. Correlations are used to evaluate the relationship or association between pairs of scores. 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.

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 pictures 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.

For future administrations, evidence will be also collected from external observers who will visit schools to observe teachers administering the assessment. These trained observers will collect information such as how teachers manage materials, the testing environment, and the test administration for students. A second set of scores for the students will also be gathered during this observation. These data will be collected as part of a planned inter-rater agreement study. More information about the 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 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 FOR SPRING 2014

This section contains an overview of the statistical summaries for the following administrations:

- Spring 2014 Operational Items
- Spring 2014 Embedded Field-Test Items

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: SPRING 2014 OPERATIONAL ITEMS

The following section provides high-level details about the Spring 2014 administration of the CoAlt: Science and Social Studies assessments.

Administration Summary

Approximately 2,600 students took the CoAlt: Science and Social Studies assessments. Tables 4–12 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 0.94 to 0.97. The SEM values for the total group ranged from 3.68 to 4.13.

Calibration Results

Item Statistics

Tables 13–16 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 17–20 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 21–22. The scale score distributions for each grade are shown in Tables 23–26. Tables 27–30 are provided and include the raw score, scale score, and CSEM values.

Decision Consistency and Accuracy

Table 31 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 32 provides the summary of teachers' responses to the test validation questions for each grade-level assessment.

CHAPTER 2: SPRING 2014 EMBEDDED FIELD-TEST ITEMS

The following section provides details around the field-test items that were embedded within the spring 2014 administration of the CoAlt: Science and Social Studies assessments.

Field-Test Items

Field-test items were included on each operational test form. Twenty-six 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 33 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. Field-test items were not flagged during the item performance review, and as a result, a data review meeting was not convened. Table 33 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 4–33

Table 4. 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	687	100	150.83	36.64	0	239	50.69	15.48	0	71	0.94	3.83
		Female	242	35.23	148.46	36.25	0	200	49.61	15.47	0	68	0.94	3.84
		Male	445	64.77	152.12	36.82	0	239	51.28	15.47	0	71	0.94	3.83
		American Indian	8	1.16	-	-	-	-	-	-	-	-	-	-
		Asian	14	2.04	-	-	-	-	-	-	-	-	-	-
		Black or African American	61	8.88	146.41	37.66	0	219	47.93	16.71	0	70	0.94	4.02
		Hispanic or Latino	261	37.99	152.10	35.08	0	239	51.63	14.35	0	71	0.93	3.74
		White	306	44.54	149.90	38.69	0	219	50.30	16.50	0	70	0.95	3.82
		Native Hawaiian or other Pacific Islander	3	0.44	-	-	-	-	-	-	-	-	-	-
		Two or More Races	28	4.08	153.82	22.66	98	219	50.82	11.63	16	70	0.87	4.23
		Missing	6	0.87	-	-	-	-	-	-	-	-	-	-
	7	Total	614	100	150.90	35.97	0	250	53.82	15.04	0	72	0.94	3.68
		Female	234	38.11	149.61	34.27	0	250	53.44	14.63	0	72	0.94	3.72
		Male	380	61.89	151.69	37.00	0	250	54.05	15.30	0	72	0.94	3.65
		American Indian	12	1.95	-	-	-	-	-	-	-	-	-	-
		Asian	11	1.79	-	-	-	-	-	-	-	-	-	-
		Black or African American	42	6.84	152.67	35.55	0	250	54.52	13.73	0	72	0.93	3.70
		Hispanic or Latino	195	31.76	154.53	29.89	0	209	55.80	12.95	0	70	0.93	3.46
		White	324	52.77	148.67	40.09	0	250	52.56	16.49	0	72	0.95	3.75
		Native Hawaiian or other Pacific Islander	2	0.33	-	-	-	-	-	-	-	-	-	-
		Two or More Races	24	3.91	148.67	37.11	0	191	53.79	15.52	0	68	0.94	3.95
		Missing	4	0.65	-	-	-	-	-	-	-	-	-	-

Table 5. 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	697	100	150.61	38.16	0	247	54.08	16.51	0	72	0.95	3.68
		Female	247	35.44	153.28	36.61	0	247	55.33	15.25	0	72	0.95	3.55
		Male	450	64.56	149.15	39.94	0	247	53.38	17.13	0	72	0.95	3.75
		American Indian	6	0.86	-	-	-	-	-	-	-	-	-	-
		Asian	23	3.30	139.26	41.06	0	184	48.22	19.70	0	68	0.96	3.91
		Black or African American	44	6.31	157.32	36.85	0	217	57.21	15.91	0	71	0.95	3.47
		Hispanic or Latino	253	36.30	149.89	35.80	0	247	53.93	16.06	0	72	0.95	3.67
		White	339	48.64	151.78	38.72	0	247	54.36	16.43	0	72	0.95	3.69
		Native Hawaiian or other Pacific Islander	2	0.29	-	-	-	-	-	-	-	-	-	-
		Two or More Races	19	2.73	132.90	61.91	0	200	48.42	23.01	0	70	0.98	3.34
		Missing	11	1.58	-	-	-	-	-	-	-	-	-	-
	8	Total	599	100	150.84	35.35	0	250	83.85	23.19	0	108	0.97	4.13
		Female	232	38.73	152.58	32.76	0	233	85.38	21.73	0	107	0.97	3.99
		Male	367	61.27	149.74	36.90	0	250	82.88	24.05	0	108	0.97	4.20
		American Indian	6	1.00	-	-	-	-	-	-	-	-	-	-
		Asian	13	2.17	-	-	-	-	-	-	-	-	-	-
		Black or African American	44	7.35	150.34	34.50	0	196	83.89	22.77	0	104	0.96	4.41
		Hispanic or Latino	224	37.40	151.09	34.40	0	204	84.87	22.12	0	105	0.97	3.92
		White	293	48.91	150.81	35.66	0	233	83.22	23.87	0	107	0.97	4.22
		Native Hawaiian or other Pacific Islander	2	0.33	-	-	-	-	-	-	-	-	-	-
		Two or More Races	13	2.17	-	-	-	-	-	-	-	-	-	-
		Missing	4	0.67	-	-	-	-	-	-	-	-	-	-

Table 6. 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	351	51.09	155.09	31.59	0	239	52.58	13.73	0	71	0.93	3.74
		Reduced Lunch Eligible	53	7.71	137.23	47.41	0	184	44.79	20.11	0	65	0.96	3.89
		Not Eligible	275	40.03	147.61	39.43	0	219	49.32	16.27	0	70	0.94	3.93
		Missing	8	1.16	-	-	-	-	-	-	-	-	-	-
	7	Free Lunch Eligible	275	44.79	155.40	29.73	0	250	55.87	13.12	0	72	0.93	3.53
		Reduced Lunch Eligible	53	8.63	158.00	26.64	55	250	56.74	10.62	6	72	0.89	3.54
		Not Eligible	277	45.11	145.51	42.25	0	250	51.30	17.14	0	72	0.95	3.83
		Missing	9	1.47	-	-	-	-	-	-	-	-	-	-
SC	5	Free Lunch Eligible	328	47.06	154.83	31.63	0	247	56.41	13.92	0	72	0.94	3.55
		Reduced Lunch Eligible	62	8.90	149.84	44.51	0	247	52.74	19.05	0	72	0.96	3.71
		Not Eligible	299	42.90	146.18	42.99	0	247	51.82	18.30	0	72	0.96	3.79
		Missing	8	1.15	-	-	-	-	-	-	-	-	-	-
	8	Free Lunch Eligible	278	46.41	153.68	32.84	0	233	86.03	21.37	0	107	0.97	3.93
		Reduced Lunch Eligible	52	8.68	146.92	42.79	0	190	81.87	28.08	0	103	0.98	3.98
		Not Eligible	262	43.74	148.37	36.48	0	250	81.78	24.02	0	108	0.97	4.36
		Missing	7	1.17	-	-	-	-	-	-	-	-	-	-

Table 7. 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	558	81.22	151.55	35.75	0	239	50.96	15.25	0	71	0.94	3.84
			NEP	96	13.97	149.29	34.99	0	193	50.09	15.14	0	67	0.94	3.75
			LEP	12	1.75	-	-	-	-	-	-	-	-	-	-
			FEP	7	1.02	-	-	-	-	-	-	-	-	-	-
			PHLOTE	4	0.58	-	-	-	-	-	-	-	-	-	-
			FELL	2	0.29	-	-	-	-	-	-	-	-	-	-
			Missing	8	1.16	-	-	-	-	-	-	-	-	-	-
		ELL Program-Bilingual	No	678	98.69	150.89	36.23	0	239	50.72	15.40	0	71	0.94	3.83
			Yes	0	0.00	-	-	-	-	-	-	-	-	-	-
			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	8	1.16	-	-	-	-	-	-	-	-	-	-
		ELL Program-ESL	No	564	82.10	151.28	36.16	0	239	50.87	15.35	0	71	0.94	3.85
			Yes	107	15.57	149.54	36.86	0	193	50.38	15.50	0	67	0.94	3.72
			Re-designated Monitored Y1	0	0.00	-	-	-	-	-	-	-	-	-	-
			Re-designated Monitored Y2	1	0.15	-	-	-	-	-	-	-	-	-	-
			Exited Y3	5	0.73	-	-	-	-	-	-	-	-	-	-
			Parent Choice	2	0.29	-	-	-	-	-	-	-	-	-	-
			Missing	8	1.16	-	-	-	-	-	-	-	-	-	-

Table 7. 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	498	81.11	150.17	38.38	0	250	53.46	15.76	0	72	0.95	3.71
			NEP	84	13.68	155.66	22.60	72	209	56.10	10.81	10	70	0.90	3.40
			LEP	9	1.47	-	-	-	-	-	-	-	-	-	-
			FEP	12	1.95	-	-	-	-	-	-	-	-	-	-
			PHLOTE	1	0.16	-	-	-	-	-	-	-	-	-	-
			FELL	1	0.16	-	-	-	-	-	-	-	-	-	-
			Missing	9	1.47	-	-	-	-	-	-	-	-	-	-
		ELL Program-Bilingual	No	603	98.21	150.96	36.19	0	250	53.85	15.10	0	72	0.94	3.67
			Yes	1	0.16	-	-	-	-	-	-	-	-	-	-
			Re-designated Monitored Y1	0	0.00	-	-	-	-	-	-	-	-	-	-
			Re-designated Monitored Y2	0	0.00	-	-	-	-	-	-	-	-	-	-
			Exited Y3	1	0.16	-	-	-	-	-	-	-	-	-	-
			Parent Choice	0	0.00	-	-	-	-	-	-	-	-	-	-
			Missing	9	1.47	-	-	-	-	-	-	-	-	-	-
		ELL Program-ESL	No	502	81.76	150.25	38.255	0	250	53.50	15.72	0	72	0.94	3.71
			Yes	83	13.52	155.55	20.33	84	209	56.24	9.82	14	70	0.88	3.43
			Re-designated Monitored Y1	1	0.16	-	-	-	-	-	-	-	-	-	-
			Re-designated Monitored Y2	3	0.49	-	-	-	-	-	-	-	-	-	-
			Exited Y3	5	0.81	-	-	-	-	-	-	-	-	-	-
			Parent Choice	11	1.79	-	-	-	-	-	-	-	-	-	-
			Missing	9	1.47	-	-	-	-	-	-	-	-	-	-

Table 8. 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	561	80.49	150.19	36.62	0	247	53.86	17.15	0	72	0.95	3.69
			NEP	103	14.78	151.33	33.98	0	247	54.51	13.78	0	72	0.93	3.63
			LEP	13	1.87	-	-	-	-	-	-	-	-	-	-
			FEP	7	1.00	-	-	-	-	-	-	-	-	-	-
			PHLOTE	5	0.72	-	-	-	-	-	-	-	-	-	-
			FELL	0	0.00	-	-	-	-	-	-	-	-	-	-
			Missing	8	1.15	-	-	-	-	-	-	-	-	-	-
		ELL Program-Bilingual	No	687	98.57	150.68	38.35	0	247	54.12	16.55	0	72	0.95	3.67
			Yes	1	0.14	-	-	-	-	-	-	-	-	-	-
			Re-designated Monitored Y1	0	0.00	-	-	-	-	-	-	-	-	-	-
			Re-designated Monitored Y2	0	0.00	-	-	-	-	-	-	-	-	-	-
			Exited Y3	1	0.14	-	-	-	-	-	-	-	-	-	-
			Parent Choice	0	0.00	-	-	-	-	-	-	-	-	-	-
			Missing	8	1.15	-	-	-	-	-	-	-	-	-	-
		ELL Program-ESL	No	568	81.49	150.24	39.44	0	247	53.88	17.11	0	72	0.95	3.69
			Yes	108	15.49	153.60	33.76	0	247	55.61	13.74	0	72	0.94	3.48
			Re-designated Monitored Y1	0	0.00	-	-	-	-	-	-	-	-	-	-
			Re-designated Monitored Y2	1	0.14	-	-	-	-	-	-	-	-	-	-
			Exited Y3	5	0.72	-	-	-	-	-	-	-	-	-	-
			Parent Choice	7	1.00	-	-	-	-	-	-	-	-	-	-
			Missing	8	1.15	-	-	-	-	-	-	-	-	-	-

Table 8. 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	491	81.97	150.31	36.03	0	250	83.29	23.81	0	108	0.97	4.16
			NEP	81	13.52	153.16	30.99	0	204	86.26	20.01	0	105	0.96	4.02
			LEP	13	2.17	-	-	-	-	-	-	-	-	-	-
			FEP	6	1.00	-	-	-	-	-	-	-	-	-	-
			PHLOTE	0	0.00	-	-	-	-	-	-	-	-	-	-
			FELL	1	0.17	-	-	-	-	-	-	-	-	-	-
			Missing	7	1.17	-	-	-	-	-	-	-	-	-	-
		ELL Program-Bilingual	No	589	98.33	150.71	35.57	0	250	83.75	23.32	0	108	0.97	4.14
			Yes	1	0.17	-	-	-	-	-	-	-	-	-	-
			Re-designated Monitored Y1	0	0.00	-	-	-	-	-	-	-	-	-	-
			Re-designated Monitored Y2	0	0.00	-	-	-	-	-	-	-	-	-	-
			Exited Y3	2	0.33	-	-	-	-	-	-	-	-	-	-
			Parent Choice	0	0.00	-	-	-	-	-	-	-	-	-	-
			Missing	7	1.17	-	-	-	-	-	-	-	-	-	-
		ELL Program-ESL	No	493	82.30	149.97	36.58	0	250	83.11	24.05	0	108	0.97	4.15
			Yes	85	14.19	153.66	30.65	0	204	86.54	19.68	0	105	0.96	4.01
			Re-designated Monitored Y1	0	0.00	-	-	-	-	-	-	-	-	-	-
			Re-designated Monitored Y2	2	0.33	-	-	-	-	-	-	-	-	-	-
			Exited Y3	2	0.33	-	-	-	-	-	-	-	-	-	-
			Parent Choice	10	1.67	-	-	-	-	-	-	-	-	-	-
			Missing	7	1.17	-	-	-	-	-	-	-	-	-	-

Table 9. 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	110	16.01	146.21	33.37	0	219	47.28	15.91	0	70	0.93	4.20
		Deaf-Blindness	0	0.00	-	-	-	-	-	-	-	-	-	-
		Developmental Delay	0	0.00	-	-	-	-	-	-	-	-	-	-
		Hearing Impairment	2	0.29	-	-	-	-	-	-	-	-	-	-
		Intellectual Disability	183	26.64	163.90	24.57	0	219	57.27	9.81	0	70	0.89	3.30
		Multiple Disabilities	274	39.88	141.51	41.68	0	239	46.36	17.03	0	71	0.94	4.09
		Orthopedic Impairment	0	0.00	-	-	-	-	-	-	-	-	-	-
		Other Health Impairment	5	0.73	-	-	-	-	-	-	-	-	-	-
		Physical Disability	70	10.19	147.47	42.75	0	193	49.94	17.13	0	67	0.95	3.67
		Emotional Disability	3	0.44	-	-	-	-	-	-	-	-	-	-
		Specific Learning Disability	22	3.20	172.09	11.68	147	193	60.86	4.37	49	67	0.53	3.00
		Speech Impairment	7	1.02	-	-	-	-	-	-	-	-	-	-
		Traumatic Brain Injury	8	1.16	-	-	-	-	-	-	-	-	-	-
		Visual Impairment	1	0.15	-	-	-	-	-	-	-	-	-	-
		None	2	0.29	-	-	-	-	-	-	-	-	-	-

Table 9. 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	86	14.01	149.62	26.86	72	250	52.37	11.97	10	72	0.89	3.93
		Deaf-Blindness	0	0.00	-	-	-	-	-	-	-	-	-	-
		Developmental Delay	0	0.00	-	-	-	-	-	-	-	-	-	-
		Hearing Impairment	0	0.00	-	-	-	-	-	-	-	-	-	-
		Intellectual Disability	195	31.76	164.08	16.41	116	228	60.19	6.01	33	71	0.68	3.39
		Multiple Disabilities	250	40.72	138.85	44.23	0	250	48.32	18.78	0	72	0.96	3.81
		Orthopedic Impairment	1	0.16	-	-	-	-	-	-	-	-	-	-
		Other Health Impairment	1	0.16	-	-	-	-	-	-	-	-	-	-
		Physical Disability	46	7.49	145.72	42.72	0	191	52.76	16.27	0	68	0.94	3.86
		Emotional Disability	2	0.33	-	-	-	-	-	-	-	-	-	-
		Specific Learning Disability	23	3.75	178.35	26.11	148	250	63.83	4.89	54	72	0.68	2.78
		Speech Impairment	3	0.49	-	-	-	-	-	-	-	-	-	-
		Traumatic Brain Injury	3	0.49	-	-	-	-	-	-	-	-	-	-
		Visual Impairment	1	0.16	-	-	-	-	-	-	-	-	-	-
		None	3	0.49	-	-	-	-	-	-	-	-	-	-

Table 10. 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	136	19.51	148.30	36.46	0	247	52.70	16.18	0	72	0.94	3.87
		Deaf-Blindness	0	0.00	-	-	-	-	-	-	-	-	-	-
		Developmental Delay	0	0.00	-	-	-	-	-	-	-	-	-	-
		Hearing Impairment	1	0.14	-	-	-	-	-	-	-	-	-	-
		Intellectual Disability	176	25.25	166.26	22.78	0	247	61.48	8.74	0	72	0.87	3.14
		Multiple Disabilities	266	38.16	138.00	45.03	0	217	48.68	19.45	0	71	0.96	3.88
		Orthopedic Impairment	1	0.14	-	-	-	-	-	-	-	-	-	-
		Other Health Impairment	12	1.72	-	-	-	-	-	-	-	-	-	-
		Physical Disability	55	7.89	150.76	23.54	65	217	53.62	11.97	5	71	0.89	3.98
		Emotional Disability	3	0.43	-	-	-	-	-	-	-	-	-	-
		Specific Learning Disability	21	3.01	174.10	15.88	137	200	64.57	5.40	47	70	0.73	2.80
		Speech Impairment	9	1.29	-	-	-	-	-	-	-	-	-	-
		Traumatic Brain Injury	10	1.43	-	-	-	-	-	-	-	-	-	-
		Visual Impairment	1	0.14	-	-	-	-	-	-	-	-	-	-
		None	6	0.86	-	-	-	-	-	-	-	-	-	-

Table 10. 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	14.86	146.24	33.34	0	214	79.25	24.16	0	106	0.96	4.64
		Deaf-Blindness	0	0.00	-	-	-	-	-	-	-	-	-	-
		Developmental Delay	0	0.00	-	-	-	-	-	-	-	-	-	-
		Hearing Impairment	2	0.33	-	-	-	-	-	-	-	-	-	-
		Intellectual Disability	184	30.72	161.94	24.99	0	233	91.32	14.87	0	107	0.94	3.64
		Multiple Disabilities	241	40.23	140.45	39.87	0	204	77.88	26.36	0	105	0.97	4.43
		Orthopedic Impairment	1	0.17	-	-	-	-	-	-	-	-	-	-
		Other Health Impairment	4	0.67	-	-	-	-	-	-	-	-	-	-
		Physical Disability	48	8.01	159.10	32.28	0	214	89.40	20.90	0	106	0.97	3.61
		Emotional Disability	3	0.50	-	-	-	-	-	-	-	-	-	-
		Specific Learning Disability	17	2.84	174.71	18.72	153	233	97.77	4.79	89	107	0.58	3.10
		Speech Impairment	2	0.33	-	-	-	-	-	-	-	-	-	-
		Traumatic Brain Injury	7	1.17	-	-	-	-	-	-	-	-	-	-
		Visual Impairment	0	0.00	-	-	-	-	-	-	-	-	-	-
		None	1	0.17	-	-	-	-	-	-	-	-	-	-

Table 11. 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	664	96.65	151.75	35.88	0	239	51.14	15.15	0	71	0.94	3.81
			Yes	23	3.35	124.39	47.88	0	180	37.87	19.31	0	64	0.94	4.55
		Braille	No	686	99.85	150.86	36.66	0	239	50.71	15.48	0	71	0.94	3.83
			Yes	1	0.15	-	-	-	-	-	-	-	-	-	-
		Eye Gaze	No	661	96.22	152.97	33.97	0	239	51.67	14.53	0	71	0.93	3.81
			Yes	26	3.78	96.39	56.21	0	161	25.77	18.13	0	57	0.95	4.26
		Modified Picture Symbols	No	667	97.09	151.32	36.88	0	239	51.01	15.43	0	71	0.94	3.79
			Yes	20	2.91	134.35	22.20	70	173	40.10	13.36	7	62	0.86	5.03
		Objects	No	665	96.80	151.83	35.50	0	239	51.13	15.10	0	71	0.94	3.81
			Yes	22	3.20	120.55	54.89	0	177	37.55	20.78	0	63	0.95	4.48
		Translation into Native Language	No	679	98.84	150.89	36.74	0	239	50.73	15.48	0	71	0.94	3.83
			Yes	8	1.16	-	-	-	-	-	-	-	-	-	-
		Other	No	621	90.39	152.21	35.22	0	239	51.32	14.98	0	71	0.94	3.82
			Yes	66	9.61	137.86	46.29	0	219	44.77	18.68	0	70	0.96	3.98
	7	Assistive Technology	No	578	94.14	152.24	34.85	0	250	54.40	14.55	0	72	0.94	3.64
			Yes	36	5.86	129.31	46.15	0	180	44.47	19.48	0	66	0.95	4.22
		Braille	No	610	99.35	150.86	36.06	0	250	53.80	15.08	0	72	0.94	3.68
			Yes	4	0.65	-	-	-	-	-	-	-	-	-	-
		Eye Gaze	No	594	96.74	152.93	33.03	0	250	54.72	13.81	0	72	0.93	3.67
			Yes	20	3.26	90.55	61.12	0	176	27.10	23.92	0	65	0.98	3.59
		Modified Picture Symbols	No	598	97.39	151.16	36.24	0	250	53.95	15.07	0	72	0.94	3.67
			Yes	16	2.61	141.25	22.72	90	176	48.88	13.38	17	65	0.92	3.82
		Objects	No	583	94.95	152.20	35.78	0	250	54.50	14.60	0	72	0.94	3.63
			Yes	31	5.05	126.45	30.98	49	163	40.94	17.58	5	61	0.94	4.41
		Translation into Native Language	No	609	99.19	150.86	36.10	0	250	53.79	15.09	0	72	0.94	3.68
			Yes	5	0.81	-	-	-	-	-	-	-	-	-	-
		Other	No	561	91.37	151.65	36.62	0	250	54.22	15.05	0	72	0.94	3.62
			Yes	53	8.63	142.89	27.19	49	191	49.60	14.37	5	68	0.92	4.19

Table 12. 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	659	94.55	152.44	36.67	0	247	54.95	15.79	0	72	0.95	3.62
			Yes	38	5.45	118.95	48.80	0	184	38.82	20.95	0	68	0.95	4.57
		Braille	No	695	99.71	150.69	38.18	0	247	54.13	16.50	0	72	0.95	3.67
			Yes	2	0.29	-	-	-	-	-	-	-	-	-	-
		Eye Gaze	No	673	96.56	152.40	36.42	0	247	55.01	15.63	0	72	0.95	3.63
			Yes	24	3.44	100.58	50.87	0	184	27.83	18.94	0	68	0.94	4.60
		Modified Picture Symbols	No	675	96.84	151.50	37.35	0	247	54.43	16.29	0	72	0.95	3.66
			Yes	22	3.16	123.55	51.87	0	174	43.18	19.57	0	66	0.96	4.17
		Objects	No	673	96.56	151.96	36.04	0	247	54.68	15.79	0	72	0.95	3.67
			Yes	24	3.44	113.00	67.68	0	247	37.17	25.58	0	72	0.98	3.92
		Translation into Native Language	No	695	99.71	150.71	38.08	0	247	54.13	16.42	0	72	0.95	3.68
			Yes	2	0.29	-	-	-	-	-	-	-	-	-	-
		Other	No	644	92.40	151.28	37.93	0	247	54.45	16.30	0	72	0.95	3.65
			Yes	53	7.60	142.49	40.31	0	247	49.49	18.44	0	72	0.95	4.00
	8	Assistive Technology	No	576	96.16	151.64	35.16	0	250	84.39	23.00	0	108	0.97	4.08
			Yes	23	3.84	130.74	34.93	0	166	70.26	24.37	0	96	0.96	5.13
		Braille	No	598	99.83	150.90	35.36	0	250	83.90	23.18	0	108	0.97	4.12
			Yes	1	0.17	-	-	-	-	-	-	-	-	-	-
		Eye Gaze	No	581	96.99	152.75	33.31	0	250	85.28	21.57	0	108	0.97	4.06
			Yes	18	3.01	89.17	44.28	0	142	37.72	26.83	0	81	0.96	5.69
		Modified Picture Symbols	No	582	97.16	152.09	34.00	0	250	84.67	22.39	0	108	0.97	4.09
			Yes	17	2.84	108.00	52.17	0	160	55.82	32.26	0	93	0.97	5.22
		Objects	No	577	96.33	151.91	34.33	0	250	84.54	22.49	0	108	0.97	4.11
			Yes	22	3.67	122.73	49.08	0	174	65.68	32.83	0	99	0.98	4.53
		Translation into Native Language	No	596	99.50	150.87	35.40	0	250	83.87	23.21	0	108	0.97	4.12
			Yes	3	0.50	-	-	-	-	-	-	-	-	-	-
		Other	No	557	92.99	151.84	33.92	0	250	84.53	22.46	0	108	0.97	4.09
			Yes	42	7.01	137.57	49.31	0	233	74.79	30.20	0	107	0.98	4.57

Table 13. Grade 4 Social Studies Classical Statistics

ITEM	TYPE	% 0	% 1	% 2	% 3	% 4	% 5	% 6	MEAN SCORE	ITEM-TOTAL CORR
1	SR	6.7	17.5	29.7	22.4	23.7			2.390	0.620
2	SR	6.4	11.8	16.2	19.2	46.4			2.875	0.755
3	SR	7.3	10.6	13.1	20.8	48.2			2.920	0.762
4	SR	5.8	9.9	14.6	12.4	57.4			3.055	0.744
5	SR	5.5	9.9	21.3	26.5	36.8			2.792	0.625
6	SR	6.0	9.3	19.4	18.8	46.6			2.907	0.684
7	SR	7.1	11.8	18.2	30.7	32.2			2.690	0.706
8	SR	7.0	13.2	20.7	31.0	28.1			2.600	0.706
9	SPT	6.8	1.0	1.0	24.6	30.6	25.8	10.2	3.891	0.680
10	SR	6.3	7.9	12.8	14.0	59.1			3.118	0.761
11	SR	7.1	8.7	15.6	19.4	49.2			2.948	0.683
12	SR	6.4	14.4	20.7	12.2	46.3			2.776	0.722
13	SPT	7.0	1.5	1.3	19.5	34.6	24.0	12.1	3.937	0.732
14	SR	7.1	9.6	13.1	16.4	53.7			3.000	0.780
15	SR	6.6	11.6	22.4	28.1	31.3			2.659	0.721
16	SR	6.7	8.6	18.8	32.5	33.5			2.774	0.615
17	SR	7.4	4.1	6.8	8.3	73.4			3.361	0.791

Table 14. Grade 7 Social Studies Classical Statistics

ITEM	TYPE	% 0	% 1	% 2	% 3	% 4	% 5	% 6	MEAN SCORE	ITEM-TOTAL CORR
1	SR	5.0	3.1	14.3	16.9	60.6			3.249	0.709
2	SR	5.2	12.5	29.0	23.6	29.6			2.599	0.628
3	SR	3.9	8.8	19.7	31.6	36.0			2.870	0.633
4	SR	4.7	5.9	14.2	30.9	44.3			3.042	0.677
5	SR	4.7	8.3	15.0	14.2	57.8			3.121	0.732
6	SR	4.2	6.4	13.8	19.2	56.4			3.171	0.719
7	SR	6.2	5.5	12.9	30.0	45.4			3.029	0.710
8	SR	5.5	7.8	18.1	18.1	50.5			3.002	0.751
9	SPT	8.1	0.8	1.3	29.6	29.2	17.8	13.2	3.769	0.765
10	SR	5.2	6.8	16.9	13.5	57.5			3.112	0.710
11	SR	5.9	6.5	18.9	20.2	48.5			2.990	0.718
12	SR	6.0	7.2	18.4	24.3	44.1			2.933	0.705
13	SPT	6.7	0.2	1.6	7.0	13.5	20.0	51.0	4.845	0.845
14	SR	5.7	8.6	26.2	16.9	42.5			2.819	0.638
15	SR	4.7	5.4	12.5	20.7	56.7			3.192	0.765
16	SR	5.5	7.8	20.7	15.8	50.2			2.972	0.705
17	SR	5.0	6.7	13.5	22.6	52.1			3.101	0.759

Table 15. Grade 5 Science Classical Statistics

ITEM	TYPE	% 0	% 1	% 2	% 3	% 4	% 5	% 6	MEAN SCORE	ITEM-TOTAL CORR
1	SR	6.6	6.6	12.9	13.1	60.8			3.149	0.740
2	SR	6.7	12.9	34.0	21.7	24.7			2.446	0.627
3	SR	7.0	3.6	8.2	7.3	73.9			3.374	0.786
4	SPT	6.7	0.9	1.0	9.5	23.5	27.3	31.1	4.485	0.781
5	SR	6.5	5.9	9.3	10.9	67.4			3.270	0.798
6	SR	7.0	12.2	16.8	21.2	42.8			2.805	0.749
7	SR	6.3	10.5	18.1	13.5	51.6			2.937	0.762
8	SR	6.7	9.5	23.4	26.4	34.0			2.714	0.688
9	SR	6.5	9.3	12.1	13.2	59.0			3.089	0.816
10	SR	7.0	13.2	24.1	18.2	37.4			2.659	0.666
11	SR	6.5	9.6	19.8	24.0	40.2			2.818	0.734
12	SR	5.9	6.5	14.3	17.2	56.1			3.112	0.758
13	SR	6.9	4.9	8.6	5.9	73.7			3.347	0.810
14	SPT	7.0	0.6	0.4	7.5	14.9	29.0	40.6	4.720	0.815
15	SR	7.2	3.4	8.0	6.5	74.9			3.385	0.778
16	SR	7.3	10.2	13.8	12.3	56.4			3.003	0.786
17	SR	7.5	8.9	23.8	19.7	40.2			2.762	0.625

Table 16. Grade 8 Science Classical Statistics

ITEM	TYPE	% 0	% 1	% 2	% 3	% 4	% 5	% 6	MEAN SCORE	ITEM-TOTAL CORR
1	SR	3.5	4.0	6.2	8.2	78.1			3.534	0.768
2	SR	4.7	9.3	12.5	14.2	59.3			3.140	0.781
3	SR	5.0	4.2	7.0	7.0	76.8			3.464	0.828
4	SR	4.5	5.0	10.7	16.2	63.6			3.294	0.804
5	SPT	6.2	0.5	1.0	4.5	9.0	17.9	60.9	5.070	0.820
6	SR	5.5	10.5	30.1	26.4	27.5			2.599	0.619
7	SR	3.7	7.5	17.4	25.5	45.9			3.025	0.718
8	SR	4.5	6.2	9.5	10.5	69.3			3.339	0.843
9	SR	5.0	5.2	6.2	15.0	68.6			3.371	0.756
10	SR	6.3	10.4	17.7	23.5	42.1			2.846	0.749
11	SR	4.3	5.3	7.2	11.4	71.8			3.409	0.798
12	SR	5.8	8.8	18.5	30.2	36.6			2.828	0.727
13	SR	5.0	8.2	10.5	20.9	55.4			3.135	0.784
14	SR	4.7	7.7	29.4	26.4	31.9			2.731	0.554
15	SR	5.5	5.7	9.0	15.5	64.3			3.274	0.753
16	SR	5.2	6.5	26.0	41.7	20.5			2.659	0.580
17	SR	6.2	5.7	13.9	23.0	51.3			3.075	0.747
18	SR	4.8	6.0	6.5	7.8	74.8			3.417	0.835
19	SR	5.2	6.0	11.9	12.2	64.8			3.254	0.773
20	SR	5.8	8.8	23.9	24.2	37.2			2.781	0.727
21	SR	6.0	10.4	27.4	25.0	31.2			2.651	0.665
22	SPT	5.8	1.2	0.8	22.4	27.5	22.2	20.0	4.114	0.730
23	SR	4.5	6.0	9.8	15.2	64.4			3.290	0.786
24	SR	6.3	3.3	6.2	7.2	77.0			3.451	0.849
25	SR	6.0	6.8	12.2	15.9	59.1			3.152	0.825
26	SR	5.5	6.8	18.4	26.2	43.1			2.945	0.674

Table 17. 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			1.08	1.13
2	SR	-0.0896	0	-1.5816	0.2833	0.8469	0.4514			0.90	0.88
3	SR	-0.0175	0	-1.1917	0.3133	0.4685	0.4099			0.87	0.83
4	SR	-0.2883	0	-1.5743	0.3035	1.3227	-0.0519			0.97	0.90
5	SR	-0.1462	0	-1.8403	-0.1310	0.8772	1.0941			1.21	1.18
6	SR	-0.1537	0	-1.4792	-0.1084	1.1001	0.4874			1.07	1.00
7	SR	0.1440	0	-1.4442	0.0042	0.3238	1.1162			0.94	0.90
8	SR	0.2000	0	-1.6249	-0.0263	0.4177	1.2336			0.94	0.91
9	SPT	0.3656	0	0.0496	-0.4036	-3.0552	0.3117	1.0192	2.0783	1.29	1.24
10	SR	-0.2637	0	-1.1523	0.1513	1.0072	-0.0062			0.92	0.92
11	SR	-0.0570	0	-1.0456	-0.0434	0.7401	0.3490			1.10	1.14
12	SR	-0.0156	0	-1.7108	0.2341	1.5167	-0.0399			0.89	0.87
13	SPT	0.3472	0	-0.1932	-0.2591	-2.5586	-0.0418	1.2165	1.8363	1.13	1.26
14	SR	-0.1037	0	-1.1420	0.2453	0.7699	0.1268			0.84	0.76
15	SR	0.1044	0	-1.6023	-0.1763	0.6724	1.1061			0.89	0.87
16	SR	0.0254	0	-1.3298	-0.2804	0.3814	1.2288			1.23	1.20
17	SR	-0.3513	0	-0.1788	0.0459	0.8822	-0.7493			0.95	0.72

Table 18. 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			1.12	1.16
2	SR	0.2771	0	-1.9415	-0.3409	1.1889	1.0935			1.10	1.15
3	SR	-0.1260	0	-2.2184	-0.1154	0.8068	1.5270			1.18	1.25
4	SR	-0.1419	0	-1.3578	-0.2732	0.3866	1.2444			1.13	1.11
5	SR	-0.1839	0	-1.5577	0.1067	1.2653	0.1858			0.99	0.92
6	SR	-0.3251	0	-1.5864	-0.0575	1.0020	0.6418			1.06	1.05
7	SR	0.0541	0	-0.7723	-0.3793	0.1813	0.9703			1.05	1.11
8	SR	0.0097	0	-1.2601	-0.2543	1.1077	0.4066			0.89	0.89
9	SPT	0.7132	0	0.8171	-0.8777	-3.0613	0.4725	1.2452	1.4041	0.92	0.92
10	SR	-0.1161	0	-1.2352	-0.2160	1.3941	0.0571			1.08	1.01
11	SR	0.0657	0	-0.9411	-0.5618	0.9948	0.5081			1.00	0.94
12	SR	0.1286	0	-1.0107	-0.4662	0.7297	0.7472			1.02	1.06
13	SPT	0.0528	0	2.4578	-2.4873	-1.1772	0.0694	0.6538	0.4835	0.82	0.79
14	SR	0.1696	0	-1.2830	-0.6076	1.4417	0.4489			1.18	1.24
15	SR	-0.2313	0	-1.1255	-0.2407	0.7713	0.5948			0.90	0.82
16	SR	0.0437	0	-1.1900	-0.4165	1.3518	0.2547			1.01	1.03
17	SR	-0.1001	0	-1.2131	-0.1457	0.6764	0.6824			0.91	0.88

Table 19. Grade 5 Science Item Parameter Estimates

ITEM	TYPE	B	D1	D2	D3	D4	D5	D6	D7	INFIT	OUTFIT
1	SR	-0.1437	0	-1.0111	-0.0670	1.0898	-0.0117			1.10	1.06
2	SR	0.4215	0	-1.8743	-0.5880	1.3082	1.1541			1.20	1.26
3	SR	-0.2893	0	-0.2463	-0.2520	1.2401	-0.7418			0.99	1.09
4	SPT	0.1041	0	0.2786	-0.4312	-1.9759	-0.2034	0.9468	1.3852	1.12	1.13
5	SR	-0.2553	0	-0.9204	0.1694	0.9878	-0.2369			0.94	0.68
6	SR	0.1763	0	-1.5482	0.1721	0.7220	0.6541			0.91	0.88
7	SR	-0.0098	0	-1.5741	0.0642	1.3711	0.1388			0.87	0.81
8	SR	0.2139	0	-1.4884	-0.4438	0.8203	1.1118			1.09	1.08
9	SR	-0.0989	0	-1.3749	0.3629	1.0028	0.0092			0.75	0.57
10	SR	0.2718	0	-1.6661	-0.1360	1.2025	0.5996			1.17	1.17
11	SR	0.1026	0	-1.5396	-0.1958	0.8134	0.9220			0.98	0.98
12	SR	-0.2134	0	-1.2969	-0.1462	0.9929	0.4503			0.99	0.87
13	SR	-0.2707	0	-0.5728	0.0422	1.4977	-0.9670			0.89	0.67
14	SPT	-0.0044	0	0.8846	0.0625	-2.5743	0.0117	0.4382	1.1773	1.04	1.36
15	SR	-0.2845	0	-0.1585	-0.2653	1.3182	-0.8943			1.12	1.10
16	SR	0.0500	0	-1.2119	0.2444	1.0996	-0.1321			0.82	0.79
17	SR	0.2299	0	-1.1736	-0.5346	1.0940	0.6142			1.39	1.39

Table 20. Grade 8 Science Item Parameter Estimates

ITEM	TYPE	B	D1	D2	D3	D4	D5	D6	D7	INFIT	OUTFIT
1	SR	-0.7634	0	-1.5443	0.2671	1.3195	-0.0423			1.23	0.80
2	SR	-0.0474	0	-1.7711	0.3571	1.1451	0.2689			0.87	0.77
3	SR	-0.3603	0	-0.8141	0.0398	1.3200	-0.5457			0.89	0.71
4	SR	-0.2696	0	-1.2724	-0.1535	0.9375	0.4884			0.90	0.78
5	SPT	-0.1010	0	1.1506	-1.0481	-1.3039	0.0981	0.6117	0.4916	1.28	1.12
6	SR	0.4834	0	-1.8358	-0.6399	1.1522	1.3235			1.12	1.11
7	SR	-0.1730	0	-2.2133	-0.1262	1.0349	1.3046			0.98	0.95
8	SR	-0.2804	0	-1.4199	0.2073	1.2636	-0.0510			0.78	0.58
9	SR	-0.2531	0	-1.0656	0.3267	0.4486	0.2903			1.16	1.03
10	SR	0.3747	0	-1.4519	-0.0941	0.7219	0.8241			0.87	0.83
11	SR	-0.3959	0	-1.3503	0.3325	0.9296	0.0881			0.99	0.78
12	SR	0.3385	0	-1.4944	-0.3164	0.5783	1.2326			0.95	0.96
13	SR	-0.0101	0	-1.5539	0.3304	0.5312	0.6924			0.88	0.81
14	SR	0.2510	0	-1.8418	-0.7948	1.2617	1.3750			1.38	1.39
15	SR	-0.0872	0	-1.0302	0.1066	0.6748	0.2488			1.10	1.06
16	SR	0.4331	0	-1.6492	-1.0314	0.5309	2.1497			1.28	1.22
17	SR	0.1454	0	-0.8954	-0.4357	0.5936	0.7375			1.02	1.00
18	SR	-0.2992	0	-1.2220	0.5239	1.1157	-0.4176			0.89	0.65
19	SR	-0.1110	0	-1.1627	-0.0987	1.2320	0.0294			1.02	0.85
20	SR	0.3557	0	-1.4858	-0.5260	1.0255	0.9862			0.91	0.87
21	SR	0.4862	0	-1.6473	-0.5321	1.0522	1.1272			1.09	1.08
22	SPT	0.4542	0	-0.2569	-0.2430	-3.0992	0.5924	1.3966	1.6101	1.15	1.21
23	SR	-0.2471	0	-1.4268	0.1288	0.9212	0.3767			0.99	0.82
24	SR	-0.1956	0	-0.1981	-0.1417	1.0274	-0.6876			0.83	0.53
25	SR	0.0809	0	-1.0553	-0.0694	0.8903	0.2344			0.76	0.67
26	SR	0.1912	0	-1.3405	-0.4691	0.7531	1.0565			1.17	1.13

Table 21. 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	166	24	238	35	254	37	29	4	283	41
	7	46	61	68	103	17	273	44	200	33	38	6	238	39
SC	5	45	61	68	117	17	269	39	232	33	79	11	311	45
	8	67	95	103	81	14	274	46	214	36	30	5	244	41

Table 22. 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
Grade 5 Science	0–134	135–159	160–183	184–250
Grade 8 Science	0–127	128–163	164–189	190–250

Table 23. Grade 4 Social Studies Scale Score Frequency Distributions

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	21	3.06	21	3.06
1	1	0.15	22	3.20
30	4	0.58	26	3.78
51	1	0.15	27	3.93
59	1	0.15	28	4.08
65	3	0.44	31	4.51
70	3	0.44	34	4.95
82	1	0.15	35	5.09
85	1	0.15	36	5.24
88	1	0.15	37	5.39
96	2	0.29	39	5.68
98	3	0.44	42	6.11
105	1	0.15	43	6.26
107	1	0.15	44	6.40
109	3	0.44	47	6.84
112	3	0.44	50	7.28
113	2	0.29	52	7.57
115	4	0.58	56	8.15
116	1	0.15	57	8.30
118	6	0.87	63	9.17
119	4	0.58	67	9.75
121	4	0.58	71	10.33
122	3	0.44	74	10.77
124	3	0.44	77	11.21
125	2	0.29	79	11.50
128	4	0.58	83	12.08
129	6	0.87	89	12.95
131	11	1.60	100	14.56
132	7	1.02	107	15.57

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
133	6	0.87	113	16.45
135	6	0.87	119	17.32
136	7	1.02	126	18.34
137	9	1.31	135	19.65
139	5	0.73	140	20.38
140	16	2.33	156	22.71
142	10	1.46	166	24.16
143	7	1.02	173	25.18
144	9	1.31	182	26.49
146	14	2.04	196	28.53
147	12	1.75	208	30.28
149	15	2.18	223	32.46
151	21	3.06	244	35.52
152	21	3.06	265	38.57
154	26	3.78	291	42.36
156	26	3.78	317	46.14
157	33	4.80	350	50.95
159	28	4.08	378	55.02
161	26	3.78	404	58.81
163	28	4.08	432	62.88
166	36	5.24	468	68.12
168	36	5.24	504	73.36
171	38	5.53	542	78.89

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
173	35	5.09	577	83.99
177	28	4.08	605	88.06
180	29	4.22	634	92.29
184	24	3.49	658	95.78
188	10	1.46	668	97.23
193	5	0.73	673	97.96
200	7	1.02	680	98.98
208	2	0.29	682	99.27
219	3	0.44	685	99.71
239	2	0.29	687	100.00

Table 24. Grade 7 Social Studies Scale Score Frequency Distributions

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	15	2.44	15	2.44
19	1	0.16	16	2.61
41	2	0.33	18	2.93
49	1	0.16	19	3.09
55	2	0.33	21	3.42
64	1	0.16	22	3.58
72	3	0.49	25	4.07
75	2	0.33	27	4.40
84	2	0.33	29	4.72
90	4	0.65	33	5.37
92	1	0.16	34	5.54
94	1	0.16	35	5.70
98	3	0.49	38	6.19
99	1	0.16	39	6.35
103	2	0.33	41	6.68
104	1	0.16	42	6.84
109	1	0.16	43	7.00
110	2	0.33	45	7.33
111	3	0.49	48	7.82
114	4	0.65	52	8.47
116	5	0.81	57	9.28
117	4	0.65	61	9.93
118	2	0.33	63	10.26
120	4	0.65	67	10.91
121	1	0.16	68	11.07
123	5	0.81	73	11.89
124	6	0.98	79	12.87
126	6	0.98	85	13.84
127	3	0.49	88	14.33
128	6	0.98	94	15.31

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
130	1	0.16	95	15.47
131	3	0.49	98	15.96
133	5	0.81	103	16.78
134	7	1.14	110	17.92
136	12	1.95	122	19.87
137	5	0.81	127	20.68
139	9	1.47	136	22.15
141	15	2.44	151	24.59
142	16	2.61	167	27.20
144	8	1.30	175	28.50
146	19	3.09	194	31.60
148	17	2.77	211	34.36
149	20	3.26	231	37.62
151	18	2.93	249	40.55
153	26	4.23	275	44.79
156	24	3.91	299	48.70
158	43	7.00	342	55.70
160	34	5.54	376	61.24
163	41	6.68	417	67.92
166	38	6.19	455	74.10
169	29	4.72	484	78.83
172	23	3.75	507	82.57
176	26	4.23	533	86.81
180	22	3.58	555	90.39
185	21	3.42	576	93.81

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
191	16	2.61	592	96.42
198	8	1.30	600	97.72
209	7	1.14	607	98.86
228	1	0.16	608	99.02
250	6	0.98	614	100.00

Table 25. Grade 5 Science Scale Score Frequency Distributions

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	24	3.44	24	3.44
1	1	0.14	25	3.59
51	2	0.29	27	3.87
59	1	0.14	28	4.02
65	3	0.43	31	4.45
70	2	0.29	33	4.73
79	2	0.29	35	5.02
85	2	0.29	37	5.31
88	2	0.29	39	5.60
90	1	0.14	40	5.74
92	1	0.14	41	5.88
98	1	0.14	42	6.03
100	1	0.14	43	6.17
101	1	0.14	44	6.31
104	2	0.29	46	6.60
106	1	0.14	47	6.74
107	2	0.29	49	7.03
110	2	0.29	51	7.32
111	2	0.29	53	7.60
112	1	0.14	54	7.75
113	1	0.14	55	7.89
116	4	0.57	59	8.46
117	3	0.43	62	8.90
118	1	0.14	63	9.04
119	4	0.57	67	9.61
121	7	1.00	74	10.62
122	5	0.72	79	11.33
123	1	0.14	80	11.48
124	2	0.29	82	11.76

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
125	5	0.72	87	12.48
127	4	0.57	91	13.06
128	4	0.57	95	13.63
129	4	0.57	99	14.20
130	6	0.86	105	15.06
131	3	0.43	108	15.49
132	6	0.86	114	16.36
134	3	0.43	117	16.79
135	7	1.00	124	17.79
136	11	1.58	135	19.37
137	7	1.00	142	20.37
139	12	1.72	154	22.09
140	11	1.58	165	23.67
141	8	1.15	173	24.82
143	16	2.30	189	27.12
144	12	1.72	201	28.84
146	21	3.01	222	31.85
147	17	2.44	239	34.29
149	15	2.15	254	36.44
150	22	3.16	276	39.60
152	27	3.87	303	43.47
154	25	3.59	328	47.06
156	31	4.45	359	51.51
158	27	3.87	386	55.38
160	25	3.59	411	58.97

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
162	31	4.45	442	63.41
165	32	4.59	474	68.01
168	37	5.31	511	73.31
171	46	6.60	557	79.91
174	29	4.16	586	84.07
179	32	4.59	618	88.67
184	34	4.88	652	93.54
191	16	2.30	668	95.84
200	15	2.15	683	97.99
217	10	1.43	693	99.43
247	4	0.57	697	100.00

Table 26. Grade 8 Science Scale Score Frequency Distributions

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	13	2.17	13	2.17
1	1	0.17	14	2.34
19	1	0.17	15	2.50
28	1	0.17	16	2.67
56	1	0.17	17	2.84
59	2	0.33	19	3.17
62	2	0.33	21	3.51
74	2	0.33	23	3.84
76	1	0.17	24	4.01
78	1	0.17	25	4.17
81	1	0.17	26	4.34
83	2	0.33	28	4.67
86	2	0.33	30	5.01
87	1	0.17	31	5.18
90	2	0.33	33	5.51
92	1	0.17	34	5.68
93	3	0.50	37	6.18
94	1	0.17	38	6.34
98	2	0.33	40	6.68
99	1	0.17	41	6.84
100	1	0.17	42	7.01
101	2	0.33	44	7.35
103	1	0.17	45	7.51
104	3	0.50	48	8.01
105	1	0.17	49	8.18
106	1	0.17	50	8.35
107	3	0.50	53	8.85
109	1	0.17	54	9.02
111	2	0.33	56	9.35

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
112	2	0.33	58	9.68
113	1	0.17	59	9.85
116	4	0.67	63	10.52
117	1	0.17	64	10.68
118	1	0.17	65	10.85
120	6	1.00	71	11.85
121	1	0.17	72	12.02
123	2	0.33	74	12.35
124	2	0.33	76	12.69
125	2	0.33	78	13.02
126	2	0.33	80	13.36
127	1	0.17	81	13.52
128	6	1.00	87	14.52
129	2	0.33	89	14.86
130	3	0.50	92	15.36
131	6	1.00	98	16.36
132	2	0.33	100	16.69
133	7	1.17	107	17.86
134	7	1.17	114	19.03
135	1	0.17	115	19.20
136	8	1.34	123	20.53
137	5	0.83	128	21.37
138	6	1.00	134	22.37
140	2	0.33	136	22.70
141	6	1.00	142	23.71
142	9	1.50	151	25.21
143	10	1.67	161	26.88
144	13	2.17	174	29.05
146	13	2.17	187	31.22
147	8	1.34	195	32.55
148	14	2.34	209	34.89

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
150	13	2.17	222	37.06
151	11	1.84	233	38.90
153	10	1.67	243	40.57
154	23	3.84	266	44.41
156	13	2.17	279	46.58
158	28	4.67	307	51.25
160	27	4.51	334	55.76
162	21	3.51	355	59.27
164	31	5.18	386	64.44
166	33	5.51	419	69.95
169	28	4.67	447	74.62
171	28	4.67	475	79.30
174	33	5.51	508	84.81
177	25	4.17	533	88.98
181	19	3.17	552	92.15
185	17	2.84	569	94.99
190	9	1.50	578	96.49
196	9	1.50	587	98.00
204	5	0.83	592	98.83
214	4	0.67	596	99.50
233	2	0.33	598	99.83
250	1	0.17	599	100.00

Table 27. 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	30	21
3	42	17
4	51	15
5	59	14
6	65	13
7	70	12
8	75	11
9	79	10
10	82	10
11	85	9
12	88	9
13	91	9
14	94	8
15	96	8
16	98	8
17	100	7
18	102	7
19	104	7
20	105	7
21	107	7
22	109	7
23	110	7
24	112	7
25	113	7
26	115	7

27	116	6
28	118	6
29	119	6
30	121	6
31	122	6
32	124	6
33	125	6
34	126	6
35	128	6
36	129	6
37	131	6
38	132	6
39	133	6
40	135	6
41	136	6
42	137	6
43	139	6
44	140	6
45	142	6
46	143	6
47	144	6
48	146	7
49	147	7
50	149	7
51	151	7
52	152	7
53	154	7
54	156	7
55	157	7
56	159	7

57	161	8
58	163	8
59	166	8
60	168	8
61	171	9
62	173	9
63	177	10
64	180	10
65	184	11
66	188	12
67	193	13
68	200	14
69	208	16
70	219	20
71	239	28
72	250	52

Table 28. 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	19	21
3	32	18
4	41	15
5	49	14
6	55	13
7	60	12
8	64	11
9	68	11
10	72	10
11	75	10
12	78	9
13	81	9
14	84	9
15	86	8
16	88	8
17	90	8
18	92	8
19	94	7
20	96	7
21	98	7
22	99	7
23	101	7
24	103	7
25	104	7
26	106	7

27	107	7
28	109	6
29	110	6
30	111	6
31	113	6
32	114	6
33	116	6
34	117	6
35	118	6
36	120	6
37	121	6
38	123	6
39	124	6
40	126	6
41	127	6
42	128	7
43	130	7
44	131	7
45	133	7
46	134	7
47	136	7
48	137	7
49	139	7
50	141	7
51	142	7
52	144	7
53	146	7
54	148	7
55	149	7
56	151	8

57	153	8
58	156	8
59	158	8
60	160	8
61	163	9
62	166	9
63	169	10
64	172	10
65	176	11
66	180	12
67	185	13
68	191	14
69	198	16
70	209	20
71	228	28
72	250	53

Table 29. Grade 5 Science Scale Scores and Conditional Standard Error of Measurement (CSEM)

Raw Score	Scale Score	CSEM
0	0	48
1	1	26
2	39	19
3	51	16
4	59	14
5	65	12
6	70	11
7	75	10
8	79	10
9	82	9
10	85	9
11	88	8
12	90	8
13	92	7
14	94	7
15	96	7
16	98	7
17	100	7
18	101	6
19	103	6
20	104	6
21	106	6
22	107	6
23	108	6
24	110	6
25	111	6
26	112	6

27	113	6
28	115	6
29	116	6
30	117	6
31	118	6
32	119	6
33	121	6
34	122	6
35	123	6
36	124	6
37	125	6
38	127	6
39	128	6
40	129	6
41	130	6
42	131	6
43	132	6
44	134	6
45	135	6
46	136	6
47	137	6
48	139	6
49	140	6
50	141	6
51	143	6
52	144	6
53	146	6
54	147	6
55	149	6
56	150	7

57	152	7
58	154	7
59	156	7
60	158	7
61	160	8
62	162	8
63	165	8
64	168	9
65	171	9
66	174	10
67	179	11
68	184	12
69	191	14
70	200	18
71	217	25
72	247	47

Table 30. Grade 8 Science Scale Scores and Conditional Standard Error of Measurement (CSEM)

Raw Score	Scale Score	CSEM
0	0	50
1	1	28
2	7	20
3	19	16
4	28	14
5	34	13
6	40	12
7	45	11
8	49	11
9	53	10
10	56	9
11	59	9
12	62	9
13	65	8
14	68	8
15	70	8
16	72	8
17	74	7
18	76	7
19	78	7
20	80	7
21	81	7
22	83	6
23	84	6
24	86	6
25	87	6
26	88	6

27	90	6
28	91	6
29	92	6
30	93	6
31	94	6
32	95	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	105	5
43	106	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	119	5
58	120	5
59	120	5
60	121	5
61	122	5
62	123	5
63	124	5
64	125	5
65	126	5
66	127	5
67	128	5
68	128	5
69	129	5
70	130	5
71	131	5
72	132	5
73	133	5
74	134	5
75	135	5
76	136	5
77	137	5
78	138	5
79	140	6
80	141	6
81	142	6
82	143	6
83	144	6
84	146	6
85	147	6
86	148	6

87	150	6
88	151	6
89	153	7
90	154	7
91	156	7
92	158	7
93	160	7
94	162	8
95	164	8
96	166	8
97	169	8
98	171	9
99	174	9
100	177	10
101	181	10
102	185	11
103	190	12
104	196	14
105	204	16
106	214	19
107	233	27
108	250	50

Table 31. 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	0.60	0.33	0.41	0.40	0.68	0.10	0.22
	7	0.57	0.35	0.34	0.43	0.62	0.13	0.25
SC	5	0.56	0.32	0.35	0.44	0.62	0.15	0.23
	8	0.59	0.37	0.35	0.41	0.64	0.12	0.23

Table 32. Test Validity Questions Summary

Question	Grade	Very Familiar	Somewhat Familiar	Familiar	Somewhat Unfamiliar	Unfamiliar				
How familiar are you with this student?	4	88.50%	5.97%	3.49%	1.31%	0.73%				
	7	87.13%	6.03%	4.07%	1.14%	1.63%				
	5	90.24%	5.02%	3.73%	0.57%	0.43%				
	8	88.65 %	4.17%	4.84%	1.34%	1.00%				
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?	4	32.17%	25.91%	16.45%	11.06%	8.01%	3.06%	3.35%		
	7	21.50%	12.21%	15.47%	12.05%	27.69%	7.17%	3.91%		
	5	26.11%	24.82%	20.66%	10.62%	11.91%	2.73%	3.16%		
	8	20.20%	10.85%	11.85%	11.35%	33.06%	9.68%	3.01%		
Question	Grade	25%	50%	75%	100%	None				
Approximately how much instructional time for this content area is in the general education classroom?	4	14.85%	6.99%	13.39%	34.06%	30.71%				
	7	5.54%	4.23%	12.87%	28.18%	49.19%				
	5	12.77%	8.46%	15.06%	35.58%	28.12%				
	8	5.01%	6.18%	12.35%	30.38%	46.08%				
Question	Grade	Oral Language	Reading	Picture Communication	Tactile	Other	Do Not Know			
This student's primary receptive communication is:	4	88.79%	1.16%	5.09%	0.15%	3.06%	1.75%			
	7	91.53%	2.12%	2.61%	0.16%	1.47%	2.12%			
	5	89.38%	1.72%	4.73%	0.57%	2.01%	1.58%			
	8	89.32%	2.34%	3.51%	0.17%	2.50%	2.17%			
Question	Grade	Oral Language	Writing	Picture Communication	Tactile	Other	Do Not Know			
This student's primary expressive communication is:	4	81.95%	0.58%	6.11%	0.44%	9.32%	1.60%			
	7	82.74%	0.16%	7.33%	0.00%	7.49%	2.28%			

	5	82.50%	0.14%	8.03%	0.43%	7.46%	1.43%			
	8	84.31%	0.00%	5.68%	0.33%	7.18%	2.50%			
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.	4	29.99%	43.23%	10.19%	7.86%	5.97%	2.77%			
	7	34.04%	39.25%	14.01%	6.03%	2.93%	3.75%			
	5	37.73%	37.73%	12.05%	6.17%	3.87%	2.44%			
	8	40.07%	35.89%	11.02%	6.34%	2.84%	3.84%			
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	Missing
How much time did this student take on the assessment?	4	3.78%	42.21%	44.25%	9.02%	0.58%	0.00%	0.00%	0.15%	0.00%
	7	3.26%	37.30%	48.86%	6.84%	2.77%	0.33%	0.49%	0.00%	0.16%
	5	6.03%	47.63%	40.60%	3.87%	1.29%	0.29%	0.29%	0.00%	0.00%
	8	3.51%	44.57%	43.74%	5.68%	1.50%	0.33%	0.50%	0.17%	0.00%

Table 33. Items Field Tested and Item Performance Review Outcomes

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

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

Colorado Alternate Assessment (CoAlt)
Spring 2014 Standard Setting Report

COLORADO DEPARTMENT *of* EDUCATION

&

December 8, 2014

Table of Contents

OVERVIEW 4

PREPARATION FOR STANDARD SETTING 5

PANELIST SELECTION AND COMPOSITION 5

DEVELOPMENT OF PLDS 6

CREATION OF MATERIALS 6

Agendas 6

Slides and Script 6

CoAlt Test Book 7

Rubrics 7

Reasoned Judgment Score Profile Sheets 7

Item Mean Reports 7

Forms 7

TRAINING OF FACILITATORS AND DATA ANALYSTS 8

STANDARD SETTING MEETING ACTIVITIES 8

GENERAL SESSION 8

THE STANDARD SETTING PROCESS 9

Review and Discuss Performance Level Descriptors 9

Review Assessment Items 9

Reasoned Judgment Task 10

Development of Threshold Descriptors 10

Standard Setting Training and Practice Round 11

Readiness Survey 11

Round 1 12

Round 1 Feedback 12

Round 2 13

Round 2 Feedback 13

Round 3 13

Round 3 Feedback 13

Evaluation 14

ROUND 3 RECOMMENDED CUT SCORES 14

VERTICAL ARTICULATION 16

PARTICIPANTS 16

VERTICAL ARTICULATION PROCESS 16

Review of PLDs 17

Discuss Expectations 17

Review and Discuss Impact Data across Grades 17

Establish Shared Recommendation *17*
Evaluation *17*
VERTICAL ARTICULATION RECOMMENDED CUT SCORES 18
REASONABLENESS REVIEW 19
APPROVAL OF THE FINAL PERFORMANCE STANDARDS 20
REFERENCES 23
APPENDIX A: SCORING RUBRICS 24
APPENDIX B: PANEL COMPOSITION 25
APPENDIX C: PERFORMANCE LEVEL DESCRIPTORS 27
APPENDIX D: ASSESSMENT FRAMEWORKS 31
APPENDIX E: REASONED JUDGMENT TASK RATING SHEET 37
APPENDIX F: READINESS SURVEY 38
APPENDIX G: SAMPLE RATINGS FORMS 39
APPENDIX H: POLICY DESCRIPTORS 44
APPENDIX I: STANDARD SETTING EVALUATION 45
APPENDIX J: VERTICAL ARTICULATION EVALUATION 49
APPENDIX K: PROPOSED RECOMMENDED CUT SCORES 52
APPENDIX L: ADDITIONAL PANELIST MEETING EVALUATION 53

OVERVIEW

In July 2014, the Colorado Department of Education (CDE) held a two-day standard setting meeting where four standard setting committees were convened 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 assessments. The purpose of this document is to provide a detailed report of the standard setting process for the Spring 2014 administration of the new CoAlt: Science and Social Studies assessments at grades 4, 5, 7, and 8.

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. 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 (see Appendix A for the Supported Performance Task Scoring Rubric).

The new CoAlt: Science and Social Studies elementary and middle school assessments were first administered in Spring 2014 from April 14, 2014 to May 2, 2014. The performance standards, or cut scores, for those assessments were recommended in July 2014 to aid in the interpretability of the test scores. High school Science and Social Studies assessments are also available; however,

the first operational administration for high school will occur in November 2014. The subject and grade combinations for CoAlt are shown in Table 1.

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 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 July 2014 was convened to obtain cut score recommendations to assist the state in delineating thresholds for each of the four performance levels for the new Science and Social Studies assessments. 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 assessments. With this methodology, standard setting panelists review each test item and 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 used in this methodology to help panelists think about whether the student performance at and around the recommended cut scores make sense for the performance levels.

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 standard setting meeting included approximately nine panelists for each grade-level 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 K–12 educators, 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 grade-level curriculum. In addition to classroom teachers, special education administrators also participated in the meeting. Panelists from the CMAS Science and Social

Studies standard setting meeting were also recruited to participate in the CoAlt 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 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 grade level individually and then reviewed the PLDs across grade levels within a subject area. 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 Spring 2014 CoAlt: Science and Social Studies 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 assessment, the Spring 2014 CoAlt test book associated with each grade-level committee was provided to panelists to use as part of the standard setting process. All operational items that appeared on the Spring 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, as well as a document indicating each item number, item type, and the content the item assessed that corresponded to the Assessment Frameworks. The Assessment Frameworks for each grade 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.

Reasoned Judgment Score Profile Sheets

A sample of different patterns of student performance, or score profiles, from the Spring 2014 operational data was presented to the panelists for discussion during each grade-level breakout session. Panelists were asked to think about the profiles and indicate the performance level in which a student with a specific score profile should be categorized.

Item Mean Reports

Item means were provided to panelists as part of the feedback provided after Round 1 recommendations. An item mean is the average rubric score obtained by all the students who took an item.

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.

- Reasoned Judgment Task Ratings Sheet: After panelists reviewed students' score profiles, they recorded their performance level ratings for each profile on the reasoned judgment rating sheet. They then referred to the sheet during group discussions of their ratings or during later portions of the standard setting process. A sample Reasoned Judgment Task Ratings sheet is provided in Appendix E.

- 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. The Readiness Survey is provided in Appendix F.

- Ratings Recommendation Forms: The ratings forms were used to collect panelists' item ratings for each round. The ratings forms for Rounds 1–3 are provided in Appendix G.

- 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 elementary and middle school assessments was held on July 17–18, 2014. During the two-day meeting, panelists from each of the four standard setting committees received training on the assessment and the standard setting process, reviewed the grade-level PLDs, reviewed the Spring 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, engaged in table and committee-level discussions, and considered the impact of their cut scores on student performance when making their cut score recommendations. On the afternoon of the second day, a vertical articulation meeting was held. During this meeting, panelists were allowed to review the cut scores set by each grade-level committee and make adjustments if necessary. 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 subjects 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 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. 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 grade-level committee was facilitated independently but the same standardized process was used across all grades.

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. In addition to the PLDs, the facilitator also reviewed and discussed the policy descriptors for each performance level with the panelists. These policy descriptors provide a general description of the expectations of students in each performance category. The descriptions can be applied to all the CoAlt: Science and Social Studies assessments and can be found in Appendix H.

Review Assessment Items

To become more familiar with the test for which they would be setting performance standards, the panelists reviewed the CoAlt assessment items. After reviewing the test, 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.

Reasoned Judgment Task

After reviewing and discussing the CoAlt test items, panelists were introduced to the reasoned judgment process. In this process, panelists reviewed the assessment's test design, the scoring rubrics, and examples of score profiles showing how individual item scores are summed to create a total test score. A sample of seven different patterns of student performance, or score profiles, from the Spring 2014 operational data and the resulting total test scores were presented to the panelists for discussion. 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 the performance level in which a student with a specific score profile should be categorized.

Panelists recorded their performance level ratings for each student profile on a rating sheet and could refer to these ratings during later portions of the standard setting process. After the panelists individually rated their score profiles, the meeting facilitator asked panelists to discuss as a whole group what rating they gave certain score combinations and why. They were also asked to discuss those patterns of performance where there was high agreement in their performance level rating and those patterns of performance where they were not in strong agreement.

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 to 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, a set of seven items was presented to the panelists for the practice round. Panelists were asked to review each specific item, the policy descriptors, the PLDs, and the threshold student descriptors, and identify the knowledge and skills the item is asking the student to demonstrate. 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 following outlines the specific steps that were to be followed for the "Emerging Level" cut.

1. Review the items and the task listed on the rating sheet.
2. Identify the skills required for the item or task.
    - Think about how those skills relate to the PLDs.
3. Decide: Should threshold Emerging Level students be able to demonstrate the skills assessed by the item or task?
4. Decide: How should performance appear for the threshold students?
5. On the ratings sheet, indicate the item-level score you feel describes what a threshold student should be able to obtain.

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 which 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. 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 consider the knowledge and skills the item is asking the student to demonstrate, the policy descriptors, the PLDs, the scoring rubrics, and the threshold student descriptors. During Round 1, panelists received a round readiness form and a Round 1 rating form to complete. 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 rating 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 several examples of how students achieved total test scores at and around the recommended whole group cut scores. The profiles were intended to show the panelists the types of performance students are demonstrating at those raw scores and to help them 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 by discussing the items associated with and around the recommended cuts. 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 assessment items, the scoring rubrics, the policy descriptors, the PLDs, 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 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 level were displayed, and the impact data were based on the median test-level cut scores. Spring 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.

In addition to CoAlt impact data, CMAS impact data were also shown to panelists when considering the CoAlt impact data. Panelists were asked their expectations regarding the CMAS impact data in relation to the CoAlt impact data before being shown the CMAS 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 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 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.

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
Grade 4	**Round 1**	35	51	65
	Round 2	34	50	67
	Round 3	35	50	66
Grade 7	**Round 1**	37	51	68
	Round 2	37	52	68
	Round 3	37	55	68
Grade 5	**Round 1**	26	46	62
	Round 2	34	53	66
	Round 3	35	55	66
Grade 8	**Round 1**	40	70	96
	Round 2	41	69	94
	Round 3	60	82	100

Based on Round 3 recommendations, Tables 3 and 4 show the percentages of students who would fall into each performance level based on the Spring 2014 administration.

Table 3. Round 3 Impact Data for Social Studies

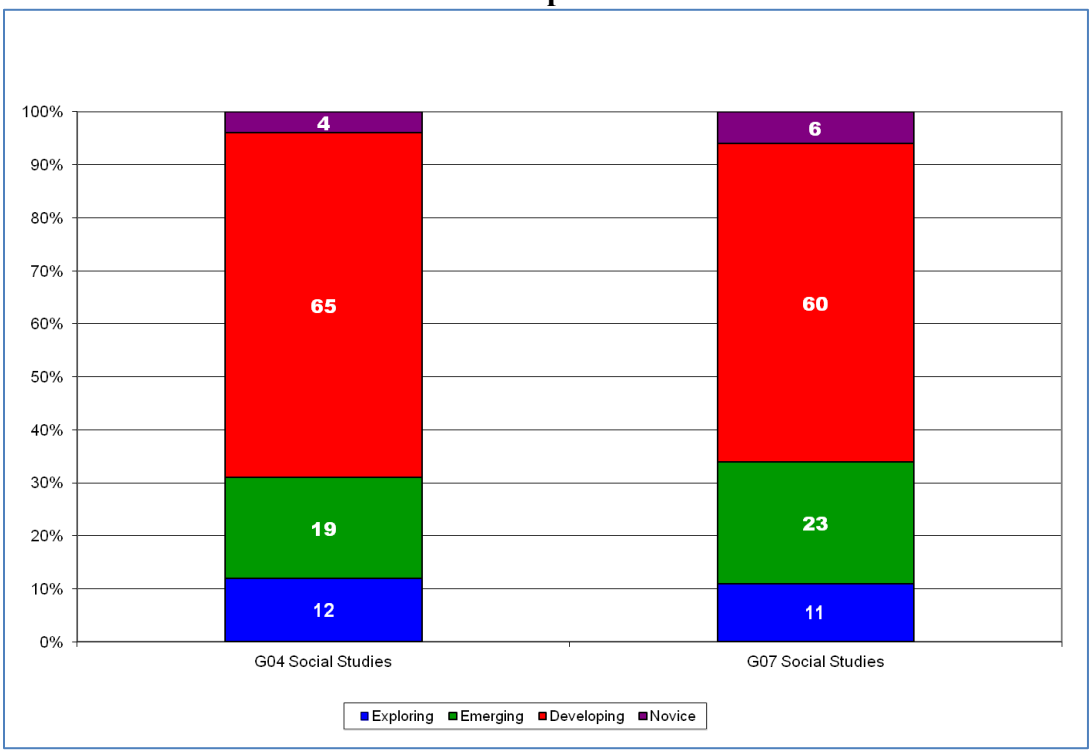
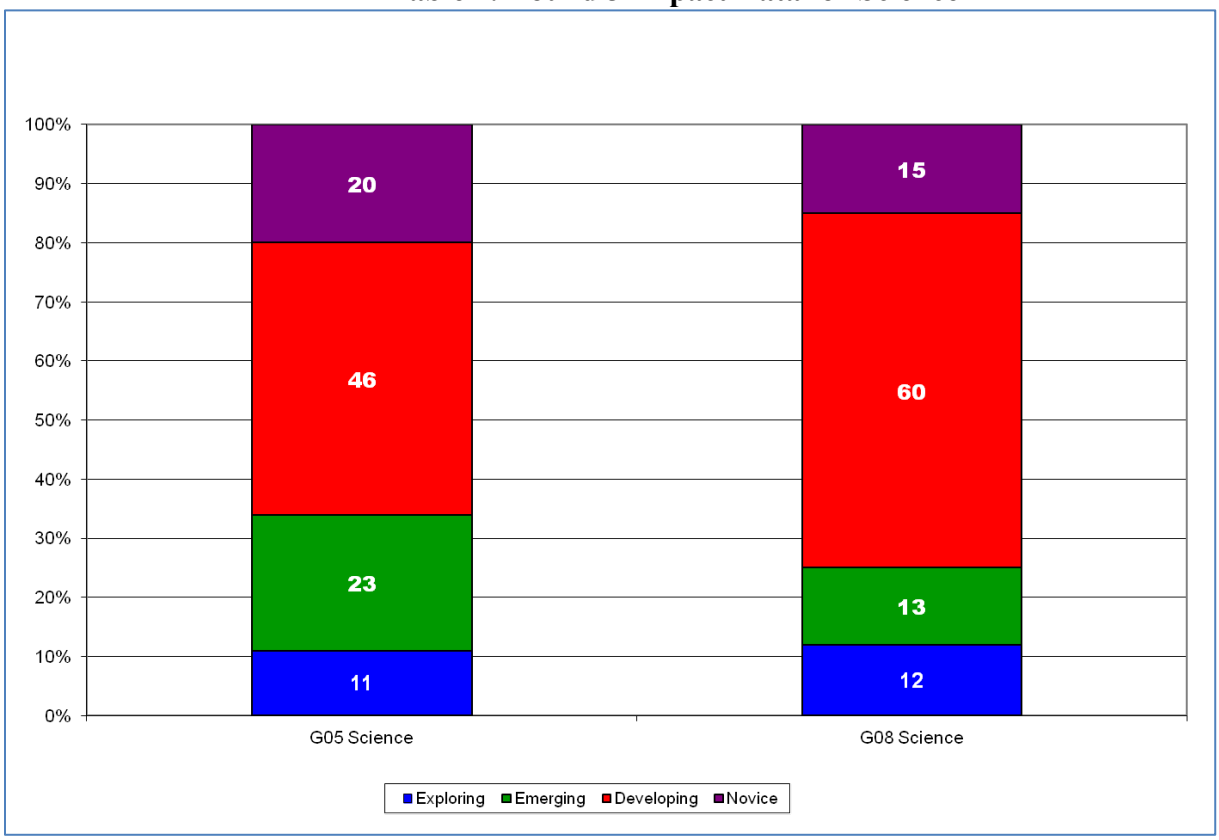


Table 4. Round 3 Impact Data for Science



VERTICAL ARTICULATION

Once the final performance standards were recommended for all grades and subjects, a subset of standard setting panelists convened to make cross-grade comparisons called vertical articulation. The purpose of vertical articulation was to review the impact data associated with the recommended cut scores across both grades within a subject to determine if the trend of the impact data is reasonable given the performance level descriptors, the test-taking population, and the skills/tasks presented on the various assessments.

Participants

Each subject-area vertical articulation committee was established by selecting four standard setting panelists from each grade-level meeting. From each grade, two educators with experience teaching students with significant cognitive disabilities were selected with the remaining panelists being a special education teacher, a content expert, and an administrator.

Vertical Articulation Process

The social studies and science committees convened in separate rooms but were facilitated with a standardized process. Parallel slides and scripts were prepared ahead of time to ensure that the same process was used across subjects. The following paragraphs outline the steps of the process.

Review of PLDs

After a brief introduction to the vertical articulation process, participants spent some time reviewing PLDs for both grades within the content area focusing especially on the grade in which they did not participate in the standard setting. The review of both PLDs helped provide a complete picture of the developmental continuum for the content area.

Discuss Expectations

After reviewing the PLDs, the expectations for impact across the grade levels were discussed as a group. Both CoAlt and CMAS expectations were discussed. The following questions were posed to the group:

- What are your expectations of the student performance data progression across the grades for CoAlt?
    - Do you expect similar percentages of students in performance levels across grades? Why or why not?
        - Is there a progression of skills in PLDs that suggest differential impact from elementary to middle school?
        - Do populations differ significantly as you move from grade to grade?
    - What other trends might you expect to see and why?
- What are your expectations of student performance data progression across the grades when comparing CoAlt and CMAS?
    - Do you expect similar percentages of students in performance levels across grades between CoAlt and CMAS? Why or why not?

Review and Discuss Impact Data across Grades

After discussing expectations, the impact data associated with the Round 3 recommended cuts from standard setting for each grades were provided in a side-by-side chart. Panelists were then encouraged to discuss how/if cut scores should change to be consistent with impact expectations.

Establish Shared Recommendation

After the discussion, the facilitator discussed the vertical articulation impact recommendation task. Throughout this discussion, it was stressed to panelists that the intent is not to undo all that was done in the standard setting workshops. Rather, the goal was to provide reasonable cut-score recommendations to policy makers that consider both the content-based recommendations and the expectations about how students should perform across performance levels. However, any desire to change the cuts needed to be justified based on the PLDs and the assessment items. Once the group reached a shared recommendation, results were displayed.

Evaluation

To end the meeting, participants completed a brief evaluation. This evaluation asked about participants' level of comfort with the vertical articulation procedure and their satisfaction with final cut score recommendations. The vertical articulation evaluation and results can be found in Appendix J. Upon completing the evaluations, panelists were thanked for their time and participation and dismissed.

Vertical Articulation Recommended Cut Scores

For social studies, the panelists recommended no adjustments to the cut scores, as reflected in Table 5. For science, the panelists recommended adjustments to the cut scores for each grade. During their discussion, panelists stated they believed the cut scores should be more rigorous and that the trends between grades 5 and 8 should be similar when considering the student population and the content. As a result, the Developing Level and Novice Level cut scores were adjusted for grade 5 and the Emerging Level, Developing Level, and Novice Level cut scores were adjusted for grade 8. The updated impact data for grades 5 and 8 are reflected in Table 6.

Table 5. Post-Vertical Articulation Impact for Social Studies

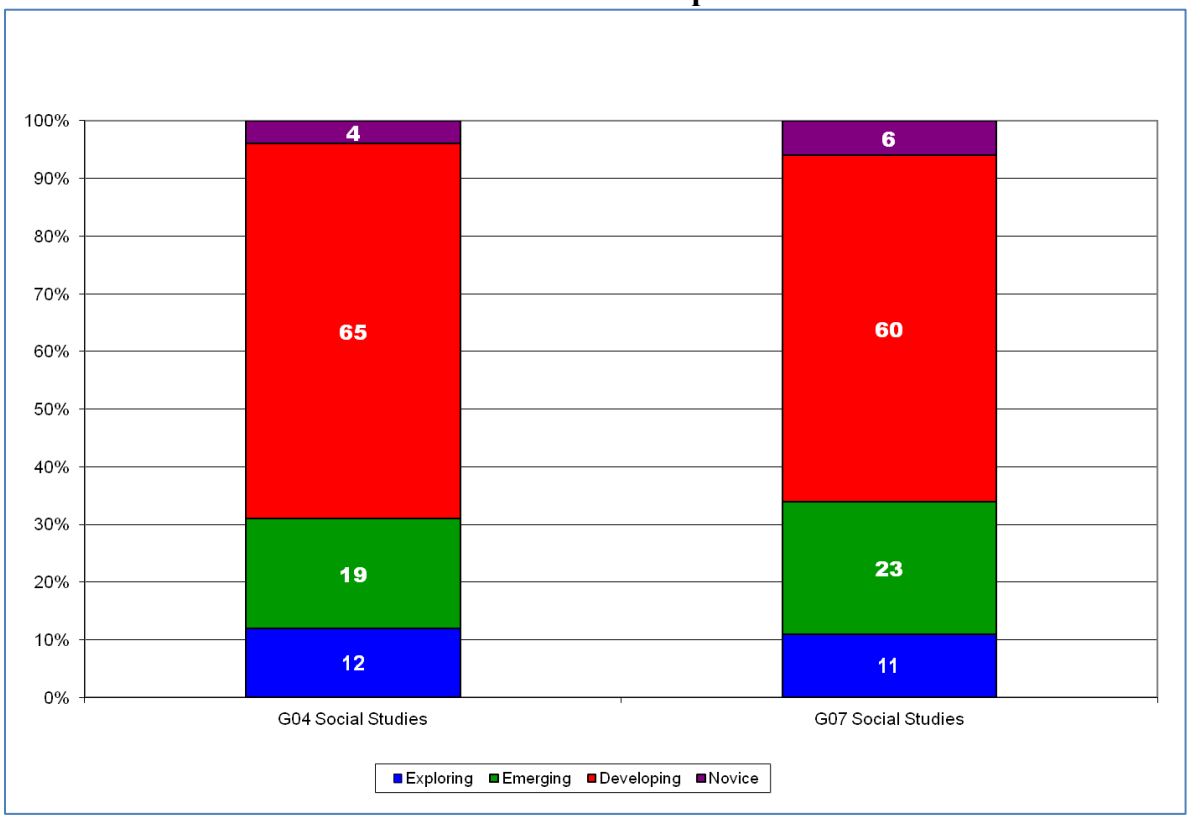
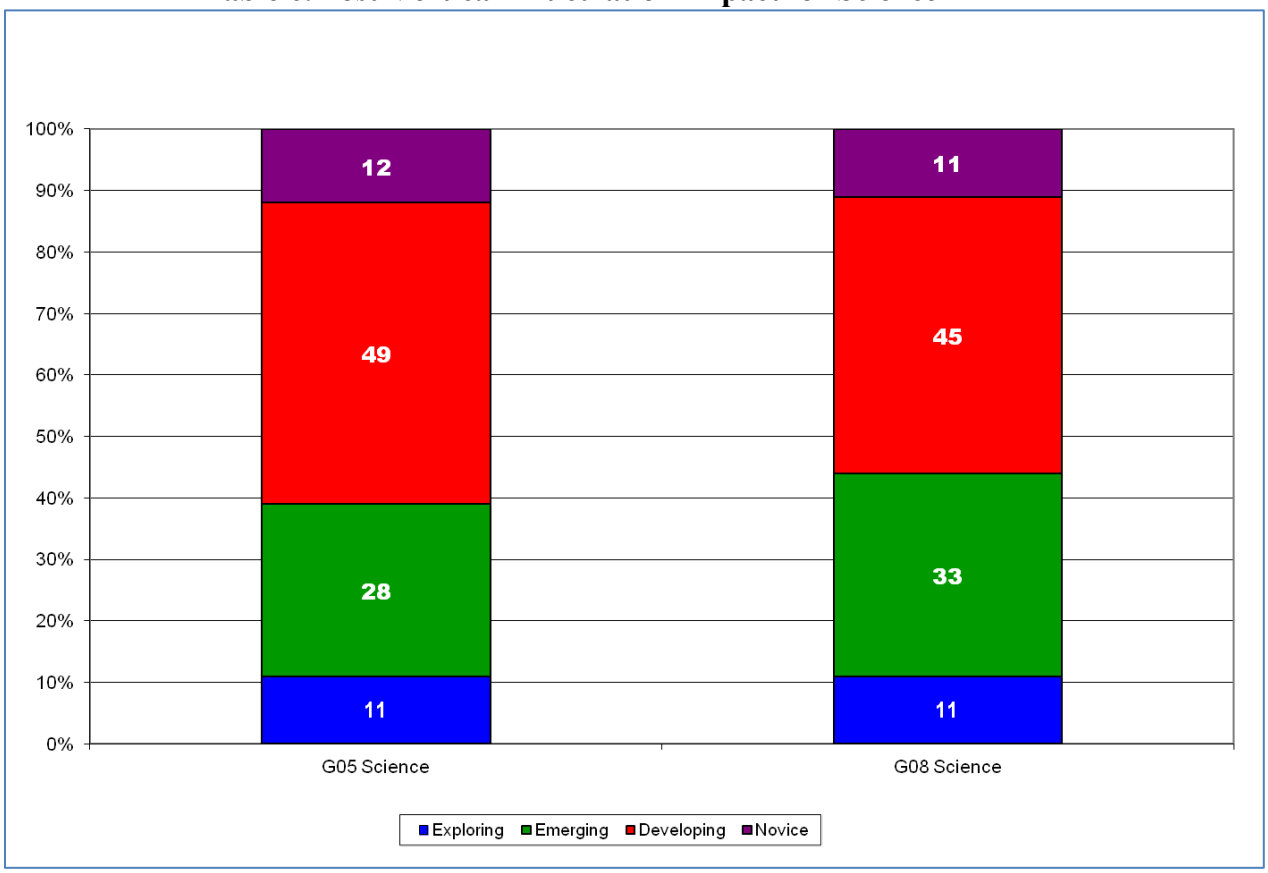


Table 6. Post-Vertical Articulation Impact for Science



At the completion of the vertical articulation, the recommendations were then reviewed by CDE.

Reasonableness Review

The reasonableness review process is intended to ensure that the performance standards contribute to a well-articulated and coherent assessment program. As part of this review, CDE reviewed the cut score recommendations from vertical articulation and considered additional information when evaluating the cut scores.

During the reasonableness review, CDE considered the following additional information:

- Alignment of grade-level test items with the corresponding PLDs
- Number of items students could get correct by chance alone
- Score profiles, which took into account the alignment of the test items with the PLDs, the content expectations shown in the PLDs, and the chance level
- Impact data associated with various score profiles
- Rigor of the content standards

The review and consideration given to this additional information was used to determine the Department-recommended cut scores. On August 11, 2014, CDE convened a half-day meeting with each subject-area standard setting committee to discuss the Department's adjustments to the cut scores and the rationales for the adjustments. The panelists who participated in the social

studies standard setting meeting met on the morning of August 11, and the panelists who participated in the science standard setting meeting met on the afternoon of the same day. A WebEx conference was also implemented for those educators who could not attend in person.

During the meetings, CDE presented additional information to the panelists and discussed how this additional data helped to shape their recommendations regarding the cut scores for each grade level. During the meetings, panelists provided their thoughts regarding the information being presented and whether they believed the rigor of the Department recommendations was appropriate. One set of recommended cut scores came out of the social studies meeting based on a shared consensus between the Department and the panelists. Several of the science panelists wanted more flexibility with the cut scores taking into account the student population. As a result, the science panelists recommended slightly lower cut scores. CDE provided the panelists' and the Department's recommendations to the State Board of Education for their review. The proposed recommended cut scores presented to the State Board of Education for social studies and science can be found in Appendix K.

To end the meetings, participants completed a brief evaluation. The evaluation and results can be found in Appendix L. Upon completing the evaluation, panelists were thanked for their time and participation and dismissed.

Approval of the Final Performance Standards

On August 13, 2014, the Colorado State Board of Education reviewed the cut score recommendations and approved the Department recommended cut scores for the CoAlt: Science and Social Studies assessments for grades 4, 5, 7, and 8. Table 7 presents the approved cut scores, and Table 8 presents the resulting scale score ranges for each performance level determined by the approved cut scores. Students with an "Inconclusive" designation were included in the Exploring Level for aggregation purposes.

Table 7. CoAlt: Science and Social Studies Final Cut Scores

	Emerging Level	Developing Level	Novice Level
Grade 4 Social Studies	46	58	66
Grade 7 Social Studies	46	61	68
Grade 5 Science	45	61	68
Grade 8 Science	67	95	103

Table 8. CoAlt: Science and Social Studies Scale Score Ranges

	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
Grade 5 Science	0–134	135–159	160–183	184–250
Grade 8 Science	0–127	128–163	164–189	190–250

Estimated impact data for the grades 4 and 7 social studies assessments were 41% and 38%, respectively, for the top two performance levels (Novice Level and Developing Level). For the grades 5 and 8 science assessments, estimated impact data were 44% and 41%, respectively, for the top two performance levels. The estimated impact data for social studies and science can be found in Tables 9 and 10.

Table 9. Estimated Final Impact for Social Studies

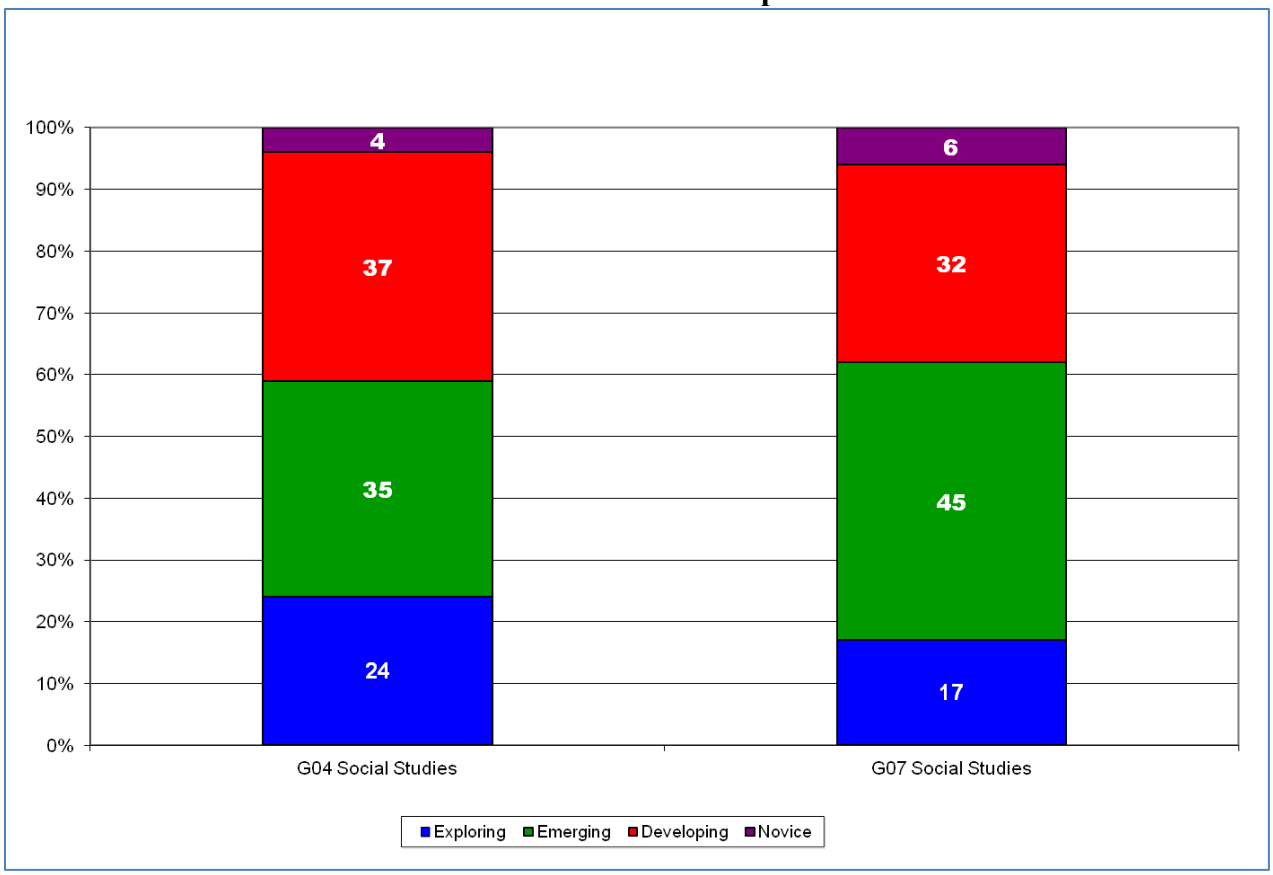
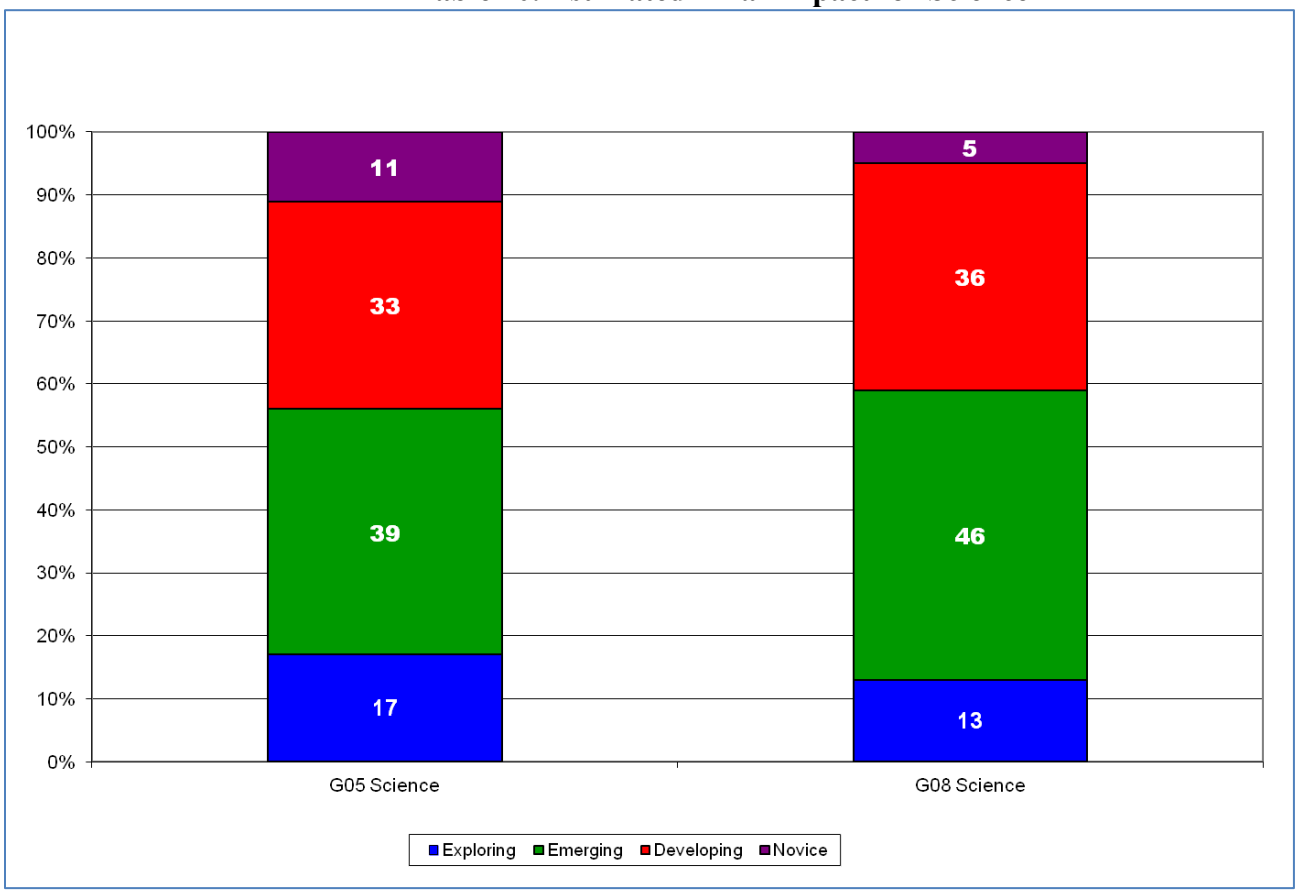


Table 10. Estimated Final Impact for Science



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	Special Ed Admin	Total
Grade 4 Social Studies	3	2	3	1	**9**
Grade 7 Social Studies	2	3	3	1	**9**
Grade 5 Science	3	1	3	1	**8**
Grade 8 Science	3	2	3	1	**9**
Total	**11**	**8**	**12**	**4**	**35**

Panelists Breakdown by School Setting

	Rural	Suburban	Urban	Omit	Total
Grade 4 Social Studies	2	5	2	0	**9**
Grade 7 Social Studies	3	6	0	0	**9**
Grade 5 Science	1	2	4	1	**8**
Grade 8 Science	2	5	2	0	**9**
Total	**8**	**18**	**8**	**1**	**35**

Panelists Breakdown by School Type

	Charter/Innovation School	Neither Charter nor Innovation School	District Level	Omit	Total
Grade 4 Social Studies	0	6	3	0	**9**
Grade 7 Social Studies	1	7	1	0	**9**
Grade 5 Science	1	5	1	1	**8**
Grade 8 Science	1	4	4	0	**9**
Total	**3**	**22**	**9**	**1**	**35**

Panelists Breakdown by Region

	Total
Denver Metro	**9**
North Central	**4**
Northeast	**3**
Northwest	**1**
Pikes Peak	**10**
Southeast	**2**
Southwest	**4**
West Central	**1**
Omit	**1**
Total	**35**

APPENDIX C: PERFORMANCE LEVEL DESCRIPTORS

Colorado Alternate Grade 4 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 Novice Level, with appropriate support, a student can typically:

- Identify historical eras, groups (e.g., miners, settlers and farmers), ideas, and themes in Colorado history
- Identify the cause and effect of growth in Colorado during various key events in U.S. history
- Integrate historical knowledge with geographical skills
- Recognize that particular dwellings, tools, and modes of transportation are specific to certain geographic areas and cultures in Colorado's history
- Identify regions and activities of Colorado based on specific physical features and label a map
- Identify choice and opportunity cost and compare the difference between the two
- Identify a specific perspective on an issue
- Identify the origins and structures of government

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

- Sequence Colorado historical events
- Identify the locations of specific activities or events in Colorado's history
- Identify specific factors that affected the growth of Colorado
- Match tools, modes of transportation, and products to natural resources or locations in Colorado
- Label a map using given map symbols
- Identify ways in which Colorado communities and markets were (and are) connected
- Identify the approximate value of goods
- Identify the functions of different levels of government
- Identify how people respond to positive and negative consequences

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

- Match historical Colorado cultures with related artifacts, modes of transportation, and resources
- Match physical, natural, and geographic features on a map to their appropriate symbols
- Identify types of goods, services and resources native to Colorado
- Recognize that items vary in their value
- Recognize that there are different levels of governance

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

- Identify artifacts (e.g., tools, housing, modes of transportation and clothing) related to Colorado history
- Identify features on a map of Colorado
- Recognize that items have value
- Recognize emergency situations and appropriate responses that affect members of the Colorado community
- Recognize that there are laws and rules

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

Colorado Alternate Grade 7 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 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.

Colorado Alternate Grade 5 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 Novice Level, with appropriate support, a student can typically:

- Demonstrate that the weight of a mixture is the same before and after separation
- Distinguish between healthy choices and unhealthy choices for the human body
- Compare and contrast characteristics between groups of plants and groups of animals
- Sort animals by observable characteristics
- Identify ways to conserve resources
- Identify landforms that are created by Earth's forces
- Identify forms of precipitation by physical characteristics

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

- Determine the weight of an individual component of a mixture after separation
- Identify the function of the internal organs of the human body
- Recognize a relationship between healthy choices and a healthy body
- Understand how plants and animals get the food they need to survive
- Compare the physical characteristics of plants to plants and animals to animals
- Distinguish between renewable and nonrenewable resources
- Identify forces that create common landforms
- Use weather-condition symbols to recognize different types of weather based on observable characteristics

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

- Identify physical properties of matter
- Select appropriate tools to separate simple mixtures based on physical properties
- Separate simple mixtures based on physical properties
- Identify the functions of the sensory organs, stomach, lungs, and heart.
- List ways to maintain a healthy body
- List observable characteristics of animals
- Match animals to animals and plants to plants based on similar physical characteristics
- List basic survival needs for plants and animals
- List Earth's resources
- Identify a source of energy as renewable or nonrenewable
- Label basic landforms of Earth
- Compare forms of precipitation

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

- Recognize physical properties of matter
- Identify observable parts of the human body
- Recognize basic survival needs for plants and animals
- Identify basic Earth resources
- Recognize basic landforms of Earth
- Identify common forms of precipitation (e.g., rain and snow)
- Recognize sources of daily/weekly weather information

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

Colorado Alternate Grade 8 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 Novice Level, with appropriate support, a student can typically:

- Match an object to itself before and after a physical or chemical change
- Compare and contrast different water or sound waves using wave characteristics
- Determine if different materials can absorb, reflect, or refract light
- Predict the effect of a human activity on a local ecosystem
- Identify why the appearances of the Sun and the moon change in the sky, including phases of the moon and eclipses

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

- Determine an object's directionality and compare the speeds of moving objects
- Determine sources for light and heat
- Determine if an object has undergone a physical or chemical change
- Identify sources of waves
- Identify human activities that have an effect on local ecosystems
- Identify traits that are passed down from parent to child
- Compare safe and unsafe practices during severe weather conditions
- Use models and simulations to explore the motions of Earth, the moon, and the Sun

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

- Recognize that the speed and direction of a force can change moving objects
- Compare different forms of energy
- Label chemical and physical changes
- Label different types of waves
- Recognize the effect of human activity on the local ecosystem
- Identify similarities and differences in parents and children
- Identify severe weather conditions and follow a simple action plan for severe weather
- Recognize facts and fiction in regards to space exploration

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

- Identify objects changing speed while moving
- Recognize that heat, light, and electricity are forms of energy
- Identify different types of waves
- Recognize stages of human aging
- Recognize different weather conditions
- Identify different climates
- Identify scientific tools related to weather and space exploration
- Acknowledge that celestial objects have patterns of movement

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) Social Studies Grade 4	Approximate % of Score Points	Approximate Score Points
1 History	**22%**	**16**
1. Organize and sequence events to understand the concepts of chronology and cause and effect in the history of Colorado		
I. Indicate one to three factors that affected the growth of Colorado (i.e. mining, farming, transportation, natural resources)		
2. The historical eras, individuals, groups, ideas and themes in Colorado history and their relationships to key events in the United States		
I. Identify significant artifacts related to Colorado history (e.g. cliff dwellings, covered wagons, mining tools, trains)		
2 Geography	**26.5%**	**19**
1. Use several types of geographic tools to answer questions about the geography of Colorado		
I. Identify features on a map of Colorado (i.e. mountains, river, plains, lakes)		
II. Create or illustrate features on a map of Colorado		
2. Connections within and across human and physical systems are developed		
I. Recognize that people live together in the same location (settlement)		
II. Identify how Colorado communities are connected (e.g. roads, railroads, airways, waterways)		
3 Economics	**26.5%**	**19**
1. People respond to positive and negative incentives		
I. Identify types of goods and services native to Colorado (e.g., tourism, steel, mining, agriculture, etc.)		
2. The relationship between choice and opportunity cost		
I. Demonstrate an understanding of the value of items (e.g. $1 can buy gum not a car)		
4 Civics	**25%**	**18**
1. Analyze and debate multiple perspectives on an issue		
I. Identify emergency situations that will affect the members of the community (e.g. fire, flood, power outage, tornado, blizzard)		
2. The origins, structure, and functions of the Colorado government		
I. Demonstrate an understanding that there are different levels of governance (i.e. federal, state, local, school, home)		
TOTAL	100%	72

Colorado Academic Standards with Extended Evidence Outcomes Alternate Assessment (CoAlt) Social Studies Grade 7	Approximate % of Score Points	Approximate Score Points
1 History	**26%**	**19**
1. Seek and evaluate multiple historical sources with different points of view to investigate a historical question and to formulate and defend a thesis with evidence		
I. Compare information from multiple sources related a significant historical event (e.g. a diary and a movie, different points of view)		
II. Develop a question related to a significant historical event (e.g. Who helped the colonists during the American Revolution?)		
2. The historical eras, individuals, groups, ideas and themes within regions of the Eastern Hemisphere and their relationships with one another		
I. Identify significant artifacts related to ancient civilizations of the Eastern Hemisphere (e.g. chariot, toga, Parthenon, pyramids, etc.)		
2 Geography	**26%**	**19**
1. Use geographic tools to ~~gather~~ data and make geographic inferences and predictions		
I. Use the cardinal points of the compass rose and directionality words, to describe a point on a map (e.g. the house is south of the park, turn left at the library)		
2. Regions have different issues and perspectives		
I. Determine what environmental resources influence settlement in different areas in one region in the Eastern Hemisphere (i.e. water, food, shelter)		
3 Economics	**17%**	**12**
1. Supply and demand influence price and profit in a market economy		
I. Identify examples of supply and demand within a familiar setting (e.g., needing more candy than pencils in the school store)		
II. Identify reasons an item would go on sale (e.g., end of season, overstocked, new version, etc.)		
2. The distribution of resources influences economic production and individual choices (PFL)		
I. Demonstrate an understanding that prices are increased by sales tax		
II. Identify three community needs or services that are paid for by taxes (e.g. roads, parks, police officers, libraries, etc.)		
4 Civics	**31%**	**22**
1. Compare how various nations define the rights, responsibilities, and roles of citizens		
I. Recognize all countries have laws		
II. Communicate the positive and negative consequences of adhering to laws/rules		
2. Different forms of government and international organizations and their influence in the world community		
I. Identify ways in which nations resolve their differences (e.g. war, treaties, diplomacy, embargo, etc.)		
TOTAL	100%	72

Colorado Academic Standards with Extended Evidence Outcomes Alternate Assessment (CoAlt) Science Grade 5	Approximate % of Score Points	Approximate Score Points
1 Physical Science	**16%**	**12**
1. Mixtures of matter can be separated regardless of how they were created; all weight and mass of the mixture are the same as the sum of weight and mass of its parts		
I. Separate simple mixtures based on physical properties		
II. Demonstrate that the weight of a mixture of solid objects before and after it is separated into parts is the same		
2 Life Science	**42%**	**30**
1. All organisms have structures and systems with separate functions		
I. Compare and contrast physical characteristics in plants and animals (plant/plant, animal/animal)		
II. Sort animals by observable characteristics based on a given group (birds, reptiles, insects and mammals)		
III. Identify how living organisms attain basic needs for survival		
2. Human body systems have basic structures, functions, and needs		
I. Identify the function of the main internal organs of the body		
II. Describe ways to maintain a healthy body		
3 Earth Systems Science	**42%**	**30**
1. Earth and sun provide a diversity of renewable and nonrenewable resources		
I. Identify Earth's resources (water, wind, and some fossil fuels such as coal, gas, solar)		
II. Identify ways to conserve resources (turn off lights, turn off water when brushing teeth)		
III. Distinguish between ~~renewable~~ and nonrenewable resources		
2. Earth's surface changes constantly through a variety of processes and forces		
I. Identify features of the Earth's surface (river, lakes, beaches, mountains, desert)		
II. Match Earth's materials to land forms (sand to beaches, rocks to mountains)		
III. Identify forces that can change the Earth's surface (erosion, deposition, climate, and human activity)		
3. Weather conditions change because of the uneven heating of Earth's surface by the Sun's energy. Weather changes are measured by differences in temperature, air pressure, wind, and water in the atmosphere and type of precipitation		
I. Make and record daily qualitative observations about the weather (temperature, wind, precipitation)		
II. Identify sources for daily/weekly weather information		
III. Compare forms of precipitation (rain, snow, hail)		
TOTAL	**100%**	**72**

Colorado Academic Standards with Extended Evidence Outcomes Alternate Assessment (CoAlt) Science Grade 8	Approximate % of Score Points	Approximate Score Points
1 Physical Science	**27%**	**29**
1. Identify and calculate the direction and magnitude of forces that act on an object, and explain the results in the object's change of motion		
I. Gather and record data about speed of moving objects		
II. Gather and record data about the direction force causes an object to move		
2. There are different forms of energy, and those forms of energy can be changed from one form to another – but total energy is conserved		
I. Select examples of different forms of energy (heat, sound, light, mechanical, and electrical)		
3. Distinguish between physical and chemical changes, noting that mass is conserved during any change		
I. Identify an object/substance as having undergone a chemical or physical change		
II. Identify an object before and after a chemical change or physical change		
4. Recognize that waves such as electromagnetic, sound, seismic, and water have common characteristics and unique properties		
I. Identify sources of waves (e.g. bell, sun, vibration, seismic, water)		
II. Compare and contrast different water or sound waves using amplitude, frequency, wavelength, and speed		
III. Identify materials that absorb, reflect, and refract light		
2 Life Science	**27%**	**29**
1. Human activities can deliberately or inadvertently alter ecosystems and their resiliency		
I. Predict the effect of a human activity on a local ecosystem		
2. Organisms reproduce and transmit genetic information (genes) to offspring, which influences individuals' traits in the next generation		
I. Label the stages of human aging/ maturation (birth, infancy, early childhood, adolescence, adulthood, death)		
II. Identify two human traits that are passed from one generation to the next		
3 Earth Systems Science	**46%**	**50**
1. Weather is a result of complex interactions of Earth's atmosphere, land and water, that are driven by energy from the sun, and can be predicted and described through complex models		
I. Identify severe weather conditions		
II. Follow a simple action plan for severe weather		
III. Compare safe versus unsafe practices during severe weather conditions (blizzards, flood, tornado, lightning)		
2. Earth has a variety of climates defined by average temperature, precipitation, humidity, air pressure, and wind that have changed over time in a particular location		
I. Compare different climates on Earth using characteristics such as temperature, hot/cold, precipitation, rain/snow etc.		
II. Identify tools to measure temperature, wind, and precipitation		
3. The solar system is comprised of various objects that orbit the Sun and are classified based on their characteristics		

Colorado Academic Standards with Extended Evidence Outcomes Alternate Assessment (CoAlt) Science Grade 8	Approximate % of Score Points	Approximate Score Points
I. Demonstrate that Earth's rotation causes the Sun to appear differently throughout the day (ex. sunrise, high noon, sunset)		
II. Recognizes that celestial objects have patterns of movement (moon and stars around sun)		
III. Explain why planets' temperatures are dependent on their proximity to the Sun		
IV. Distinguish between fact and fiction regarding space exploration (e.g. Star Wars vs. factual space exploration)		
4. The relative positions and motions of Earth, Moon, and Sun can be used to explain observable effects such as seasons, eclipses, and Moon phases		
I. Provide reasoning for the Moon's appearance during the Moon's common phases (crescent, half and, full moon)		
II. Conduct an investigation of how the Moon changes appearance during the month (crescent, new, half and, full moon). Not applicable for assessment		
III. Identify the relationship between orientation of the Earth and seasons		
IV. Distinguish features of solar and lunar eclipses		
V. Identify tools to view objects in the sky and space		
TOTAL	**100%**	**108**

APPENDIX E: REASONED JUDGMENT TASK RATING SHEET

Colorado Alternate Assessment (CoAlt)
Standard Setting Meeting

Reasoned Judgment Task Ratings

Assessment: _____

Panelist ID: _____

Instructions: Consider the sample of score profiles below. For each combination of scores, determine if student performance should be considered as Novice, Developing, Emerging, or Exploring. Write your rating in the appropriate box below.

Score Profile	Your Performance Level Rating
Score Profile 1	
Score Profile 2	
Score Profile 3	
Score Profile 4	
Score Profile 5	
Score Profile 6	
Score Profile 7	

APPENDIX F: READINESS SURVEY

CoAlt Standard-Setting Round 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 G: SAMPLE RATINGS FORMS

Colorado Alternate Assessment (CoAlt)

Assessment: Grade 4 Social Studies

Panelist ID: _____

Table Number: _____

Round 1 Ratings

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

Item/Task	Item Type	Score Range	Round 1		
			Emerging	Developing	Novice
Item 1	SR	0–4			
Item 2	SR	0–4			
Item 4	SR	0–4			
Item 5	SR	0–4			
Item 7	SR	0–4			
Item 8	SR	0–4			
Item 9	SR	0–4			
Item 11	SR	0–4			
Task 12 Prompt 1	SPT	0–2			
Task 12 Prompt 2		0–2			
Task 12 Prompt 3		0–2			

Item 13	SR	0–4			
Item 15	SR	0–4			
Item 16	SR	0–4			
Task 18 Prompt 1	SPT	0–2			
Task 18 Prompt 2		0–2			
Task 18 Prompt 3		0–2			
Item 19	SR	0–4			
Item 20	SR	0–4			
Item 22	SR	0–4			
Item 23	SR	0–4			

Colorado Alternate Assessment (CoAlt)

Assessment: Grade 4 Social Studies

Panelist ID: _____

Table Number: _____

Round 2 Ratings

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

Item/Task	Item Type	Score Range	Round 2		
			Emerging	Developing	Novice
Item 1	SR	0–4			
Item 2	SR	0–4			
Item 4	SR	0–4			
Item 5	SR	0–4			
Item 7	SR	0–4			
Item 8	SR	0–4			
Item 9	SR	0–4			
Item 11	SR	0–4			
Task 12 Prompt 1	SPT	0–2			
Task 12 Prompt 2		0–2			
Task 12 Prompt 3		0–2			
Item 13	SR	0–4			

Item 15	SR	0–4			
Item 16	SR	0–4			
Task 18 Prompt 1	SPT	0–2			
Task 18 Prompt 2		0–2			
Task 18 Prompt 3		0–2			
Item 19	SR	0–4			
Item 20	SR	0–4			
Item 22	SR	0–4			
Item 23	SR	0–4			

Colorado Alternate Assessment (CoAlt)

Assessment: Grade 4 Social Studies

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 H: POLICY DESCRIPTORS

2014 CoAlt Performance Level Descriptors
Social Studies Grades 4, 7, and High School

PL Label	Performance Level Descriptors
Inconclusive	The Inconclusive student's responses are not evident or are inconsistent when presented with a variety of social studies materials and concepts.
Exploring	The Exploring student demonstrates an initial understanding of concepts and skills represented by the Extended Evidence Outcomes of the Colorado Academic Standards for Social Studies. The student will need extensive academic supports to engage successfully in further studies in this content area.
Emerging	The Emerging student demonstrates a limited understanding of concepts and skills represented by the Extended Evidence Outcomes of the Colorado Academic Standards for Social Studies. The student will likely need moderate academic supports to engage successfully in further studies in this content area.
Developing	The Developing student demonstrates a foundational understanding of concepts and skills represented by the Extended Evidence Outcomes of the Colorado Academic Standards for Social Studies. The student is academically prepared to engage in further studies in this content area with appropriate supports.
Novice	The Novice student demonstrates a solid understanding of concepts and skills represented by the Extended Evidence Outcomes of the Colorado Academic Standards for Social Studies. The student is academically well prepared to engage in further studies in this content area with appropriate supports.

2014 CoAlt Performance Level Descriptors
Science Grades 5, 8, and High School

PL Label	Performance Level Descriptors
Inconclusive	The Inconclusive student's responses are not evident or are inconsistent when presented with a variety of scientific materials and concepts.
Exploring	The Exploring student demonstrates an initial understanding of concepts and skills represented by the Extended Evidence Outcomes of the Colorado Academic Standards for Science. The student will need extensive academic supports to engage successfully in further studies in this content area.
Emerging	The Emerging student demonstrates a limited understanding of concepts and skills represented by the Extended Evidence Outcomes of the Colorado Academic Standards for Science. The student will likely need moderate academic supports to engage successfully in further studies in this content area.
Developing	The Developing student demonstrates a foundational understanding of concepts and skills represented by the Extended Evidence Outcomes of the Colorado Academic Standards for Science. The student is academically prepared to engage in further studies in this content area with appropriate supports.
Novice	The Novice student demonstrates a solid understanding of concepts and skills represented by the Extended Evidence Outcomes of the Colorado Academic Standards for Science. The student is academically well prepared to engage in further studies in this content area with appropriate supports.

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?

- ☐ Grade 4 Social Studies
- ☐ Grade 7 Social Studies
- ☐ Grade 5 Science
- ☐ Grade 8 Science

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?"		☐	☐	☐	☐
	Grade 4	0%	0%	11%	89%
	Grade 5	0%	0%	13%	88%
	Grade 7	0%	0%	22%	78%
	Grade 8	0%	22%	11%	67%
If you cannot support, please explain why not:					
2. To what degree do you support the recommended cut score for "Developing Level?"		☐	☐	☐	☐
	Grade 4	0%	0%	33%	67%
	Grade 5	0%	0%	13%	88%
	Grade 7	0%	22%	67%	11%
	Grade 8	11%	33%	11%	44%
If you cannot support, please explain why not:					
3. To what degree do you support the recommended cut score for "Novice Level?"		☐	☐	☐	☐
	Grade 4	0%	11%	11%	78%
	Grade 5	0%	13%	0%	88%
	Grade 7	0%	0%	0%	100%
	Grade 8	0%	0%	33%	67%
If you cannot support, please explain why not:					

	Way too low	A bit low	Appropriate	A bit high	Way too high	(Omit)
4. The recommended cut score for "Emerging Level" is:	☐	☐	☐	☐	☐	
Grade 4	0%	33%	56%	11%	0%	
Grade 5	0%	0%	100%	0%	0%	
Grade 7	0%	11%	78%	0%	0%	11%
Grade 8	0%	11%	67%	0%	11%	11%
5. The recommended cut score for "Developing Level" is:	☐	☐	☐	☐	☐	
Grade 4	0%	33%	67%	0%	0%	
Grade 5	0%	0%	100%	0%	0%	
Grade 7	0%	67%	11%	11%	0%	11%
Grade 8	11%	22%	44%	0%	11%	11%
6. The recommended cut score for "Novice Level" is:	☐	☐	☐	☐	☐	
Grade 4	0%	0%	78%	11%	11%	
Grade 5	0%	0%	88%	13%	0%	
Grade 7	0%	0%	89%	0%	0%	11%
Grade 8	0%	11%	56%	11%	11%	11%

		Strongly Disagree	Disagree	Agree	Strongly Agree	(Omit)
7. The Modified Extended Angoff Method was explained clearly by the group facilitator.		☐	☐	☐	☐	
	Grade 4	0%	0%	22%	78%	
	Grade 5	0%	0%	63%	38%	
	Grade 7	0%	11%	0%	89%	
	Grade 8	0%	0%	22%	78%	
8. I had a solid understanding of what the test was intended to measure.		☐	☐	☐	☐	
	Grade 4	0%	0%	11%	89%	
	Grade 5	0%	0%	50%	50%	
	Grade 7	0%	0%	33%	67%	
	Grade 8	0%	0%	22%	78%	
9. I could clearly distinguish between performance levels.		☐	☐	☐	☐	
	Grade 4	0%	0%	22%	78%	
	Grade 5	0%	0%	50%	50%	
	Grade 7	0%	0%	78%	22%	
	Grade 8	0%	0%	67%	33%	
10. After the first round of recommendations, I felt comfortable with the standard setting procedure.		☐	☐	☐	☐	
	Grade 4	0%	0%	22%	78%	
	Grade 5	0%	0%	38%	63%	
	Grade 7	0%	0%	33%	67%	
	Grade 8	0%	0%	33%	67%	
11. I found the feedback on the comparison of all panelists' recommendations to be useful in standard setting.		☐	☐	☐	☐	
	Grade 4	0%	0%	33%	67%	
	Grade 5	0%	0%	25%	75%	
	Grade 7	0%	0%	33%	67%	
	Grade 8	0%	0%	11%	89%	
12. I found the item mean score information to be useful in standard setting.		☐	☐	☐	☐	
	Grade 4	0%	0%	33%	67%	
	Grade 5	0%	0%	25%	75%	
	Grade 7	0%	0%	44%	56%	
	Grade 8	0%	0%	22%	78%	
13. I found the score profile information to be useful in standard setting.		☐	☐	☐	☐	
	Grade 4	0%	0%	11%	89%	
	Grade 5	0%	0%	25%	75%	
	Grade 7	0%	0%	44%	56%	
	Grade 8	0%	0%	33%	67%	
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.		☐	☐	☐	☐	
	Grade 4	0%	0%	22%	78%	
	Grade 5	0%	0%	0%	88%	13%
	Grade 7	0%	0%	33%	67%	
	Grade 8	0%	0%	22%	78%	
15. Table and group discussions were open and honest.		☐	☐	☐	☐	
	Grade 4	0%	0%	11%	89%	
	Grade 5	0%	0%	13%	75%	13%
	Grade 7	0%	0%	11%	89%	
	Grade 8	0%	0%	0%	100%	

		Strongly Disagree	Disagree	Agree	Strongly Agree	(Omit)
16. I believe that my opinions were considered and valued by my group.		☐	☐	☐	☐	
	Grade 4	0%	0%	11%	89%	
	Grade 5	0%	0%	25%	75%	
	Grade 7	0%	0%	22%	78%	
	Grade 8	0%	0%	11%	89%	
17. The facilitator led the group through the standard setting process without imposing ideas about where cut scores should be.		☐	☐	☐	☐	
	Grade 4	0%	0%	0%	100%	
	Grade 5	0%	0%	38%	63%	
	Grade 7	0%	0%	11%	89%	
	Grade 8	0%	0%	33%	67%	
18. I am confident that the final cut score recommendations reflect the performance level descriptors associated with CoAlt.		☐	☐	☐	☐	
	Grade 4	0%	11%	33%	56%	
	Grade 5	0%	0%	13%	75%	13%
	Grade 7	0%	0%	78%	11%	11%
	Grade 8	0%	0%	44%	44%	11%
19. I am confident that the final cut score recommendations reflect high expectations consistent with the Extended Evidence Outcomes of the Colorado Academic Standards.		☐	☐	☐	☐	
	Grade 4	0%	0%	56%	44%	
	Grade 5	0%	0%	25%	75%	
	Grade 7	0%	0%	67%	22%	11%
	Grade 8	0%	0%	33%	67%	

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

APPENDIX J: VERTICAL ARTICULATION EVALUATION

Colorado Alternate Assessment (CoAlt)
Vertical Articulation 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 vertical articulation meeting did you participate?

☐ Social Studies ☐ Science

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 Elementary School "Emerging Level?"	☐	☐	☐	☐
Science	0%	13%	38%	50%
Social Studies	0%	0%	13%	88%
If you cannot support, please explain why not:				
2. To what degree do you support the recommended cut score for Elementary School "Developing Level?"	☐	☐	☐	☐
Science	0%	25%	38%	38%
Social Studies	0%	0%	38%	63%
If you cannot support, please explain why not:				
3. To what degree do you support the recommended cut score for Elementary School "Novice Level?"	☐	☐	☐	☐
Science	0%	0%	63%	38%
Social Studies	0%	13%	0%	88%
If you cannot support, please explain why not:				
4. To what degree do you support the recommended cut score for Middle School "Emerging Level?"	☐	☐	☐	☐
Science	0%	25%	38%	38%
Social Studies	0%	0%	13%	88%
If you cannot support, please explain why not:				
5. To what degree do you support the recommended cut score for Middle School "Developing Level?"	☐	☐	☐	☐
Science	0%	38%	38%	25%
Social Studies	0%	0%	38%	63%
If you cannot support, please explain why not:				
6. To what degree do you support the recommended cut score for Middle School "Novice Level?"	☐	☐	☐	☐
Science	0%	0%	63%	38%
Social Studies	0%	13%	0%	88%
If you cannot support, please explain why not:				

	Way too low	A bit low	Appropriate	A bit high	Way too high
7. The recommended cut score for Elementary School "Emerging Level" is:	☐	☐	☐	☐	☐
Science	0%	13%	75%	13%	0%
Social Studies	0%	13%	88%	0%	0%
8. The recommended cut score for Elementary School "Developing Level" is:	☐	☐	☐	☐	☐
Science	13%	13%	75%	0%	0%
Social Studies	0%	38%	63%	0%	0%
9. The recommended cut score for Elementary School "Novice Level" is:	☐	☐	☐	☐	☐
Science	0%	0%	100%	0%	0%
Social Studies	0%	0%	88%	0%	13%

	Way too low	A bit low	Appropriate	A bit high	Way too high
10. The recommended cut score for Middle School "Emerging Level" is:	☐	☐	☐	☐	☐
Science	0%	25%	63%	13%	0%
Social Studies	0%	13%	88%	0%	0%
11. The recommended cut score for Middle School "Developing Level" is:	☐	☐	☐	☐	☐
Science	13%	25%	50%	13%	0%
Social Studies	0%	38%	63%	0%	0%
12. The recommended cut score for Middle School "Novice Level" is:	☐	☐	☐	☐	☐
Science	0%	13%	75%	13%	0%
Social Studies	0%	0%	88%	0%	13%

	Strongly Disagree	Disagree	Agree	Strongly Agree	(Omit)
13. Table and group discussions were open and honest.	☐	☐	☐	☐	
Science	0%	38%	25%	38%	
Social Studies	0%	0%	25%	75%	
14. I believe that my opinions were considered and valued by my group.	☐	☐	☐	☐	
Science	0%	25%	38%	38%	
Social Studies	0%	0%	25%	75%	
15. The facilitator led the group through the vertical articulation process without imposing ideas about where cut scores should be.	☐	☐	☐	☐	
Science	0%	0%	25%	75%	
Social Studies	0%	0%	25%	75%	
16. I am confident that the final cut score recommendations reflect the performance level descriptors associated with CoAlt.	☐	☐	☐	☐	
Science	0%	0%	50%	25%	25%
Social Studies	0%	13%	13%	75%	
17. I am confident that the final cut score recommendations reflect high expectations consistent with the Extended Evidence Outcomes of the Colorado Academic Standards.	☐	☐	☐	☐	
Science	0%	25%	25%	25%	25%
Social Studies	0%	0%	25%	75%	

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

APPENDIX K: PROPOSED RECOMMENDED CUT SCORES

		Emerging Level	Developing Level	Novice Level
Grade 4 Social Studies	**Final Recommendation**	46	58	66
Grade 7 Social Studies	**Final Recommendation**	46	61	68
Grade 5 Science	**Group 1 Recommendation**	41	59	68
	Group 2 (Dept. Recommendation)	45	61	
Grade 8 Science	**Group 1 Recommendation**	61	91	101
	Group 2 (Dept. Recommendation)	67	95	103

APPENDIX L: ADDITIONAL PANELIST MEETING EVALUATION

Colorado Alternate Assessment (CoAlt)
Grades 4 and 7 Social Studies
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.

Indicate your response by checking the appropriate box.

	Do not support	Support with some reservation	Moderately support	Strongly support
1. I support the changes made to the Grade 4 Social Studies performance level descriptors.	0%	0%	10%	90%
2. To what degree do you support the recommended cut score for Grade 4 Social Studies "Emerging Level?"	0%	0%	10%	90%
If you cannot support, please explain why not:				
3. To what degree do you support the recommended cut score for Grade 4 Social Studies "Developing Level?"	0%	0%	10%	90%
If you cannot support, please explain why not:				
4. To what degree do you support the recommended cut score for Grade 4 Social Studies "Novice Level?"	0%	0%	10%	90%
If you cannot support, please explain why not:				
5. I support the changes made to the Grade 7 Social Studies performance level descriptors.	0%	0%	20%	80%
6. To what degree do you support the recommended cut score for Grade 7 Social Studies "Emerging Level?"	0%	0%	20%	80%
If you cannot support, please explain why not:				
7. To what degree do you support the recommended cut score for Grade 7 Social Studies "Developing Level?"	0%	0%	20%	80%
If you cannot support, please explain why not:				
8. To what degree do you support the recommended cut score for Grade 7 Social Studies "Novice Level?"	0%	0%	20%	80%
If you cannot support, please explain why not:				

	Way too low	A bit low	Appropriate	A bit high	Way too high
9. The recommended cut score for Grade 4 Social Studies "Emerging Level" is:	0%	0%	90%	10%	0%
10. The recommended cut score for Grade 4 Social Studies "Developing Level" is:	0%	0%	100%	0%	0%
11. The recommended cut score for Grade 4 Social Studies "Novice Level" is:	0%	0%	80%	20%	0%
12. The recommended cut score for Grade 7 Social Studies "Emerging Level" is:	0%	0%	100%	0%	0%
13. The recommended cut score for Grade 7 Social Studies "Developing Level" is:	0%	0%	100%	0%	0%
14. The recommended cut score for Grade 7 Social Studies "Novice Level" is:	0%	0%	90%	10%	0%

	Strongly Disagree	Disagree	Agree	Strongly Agree	(Omit)
15. Discussions were open and honest.	0%	0%	10%	80%	10%
16. I found the score profiles to be helpful.	0%	0%	10%	90%	
17. I believe that my opinions were considered and valued.	0%	0%	10%	90%	
18. I am confident that the final cut score recommendations reflect the performance level descriptors associated with CoAlt.	0%	0%	10%	90%	
19. I am confident that the final cut score recommendations reflect high expectations consistent with the Extended Evidence Outcomes of the Colorado Academic Standards.	0%	0%	20%	80%	

20. Please provide any suggestions you have for high school standard setting.

21. Please provide any suggestions you have for the type of guidance needed to help support score interpretation.

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

Colorado Alternate Assessment (CoAlt)
Grade 5 Science
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.

Indicate your response by checking the appropriate box.

	Do not support	Support with some reservation	Moderately support	Strongly support	(Omit)
1. I support the changes made to the Grade 5 Science performance level descriptors.	0%	0%	22%	78%	
2. Circle your recommended grade 5 Science "Emerging Level" cut score: 35 (0%) 41 (22%) 45 (67%) omit (11%)					
3. To what degree do you support the recommended cut score for Grade 5 Science "Emerging Level?"	0%	0%	33%	67%	
If you cannot support, please explain why not:					
4. Circle your recommended grade 5 Science "Developing Level" cut score: 59 (22%) 61 (67%) omit (11%)					
5. To what degree do you support the recommended cut score for Grade 5 Science "Developing Level?"	0%	0%	22%	67%	11%
If you cannot support, please explain why not:					
6. To what degree do you support the recommended cut score for Grade 5 Science "Novice Level?"	0%	0%	33%	67%	
If you cannot support, please explain why not:					

	Way too low	A bit low	Appropriate	A bit high	Way too high	(Omit)
7. The recommended cut score for Grade 5 Science "Emerging Level" is:	0%	0%	67%	22%	0%	11%
8. The recommended cut score for Grade 5 Science "Developing Level" is:	0%	0%	67%	22%	0%	11%
9. The recommended cut score for Grade 5 Science "Novice Level" is:	0%	0%	67%	22%	0%	11%

	Strongly Disagree	Disagree	Agree	Strongly Agree	(Omit)
10. Discussions were open and honest.	0%	11%	22%	67%	
11. I found the score profiles to be helpful.	0%	0%	56%	44%	
12. I believe that my opinions were considered and valued.	0%	0%	56%	33%	11%
13. I am confident that the final cut score recommendations reflect the performance level descriptors associated with CoAlt.	0%	0%	56%	44%	
14. I am confident that the final cut score recommendations reflect high expectations consistent with the Extended Evidence Outcomes of the Colorado Academic Standards.	0%	0%	56%	44%	

15. Please provide any suggestions you have for high school standard setting.

16. Please provide any suggestions you have for the type of guidance needed to help support score interpretation.

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

Colorado Alternate Assessment (CoAlt)
Grade 8 Science
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.

Indicate your response by checking the appropriate box.

	Do not support	Support with some reservation	Moderately support	Strongly support
1. I support the changes made to the Grade 8 Science performance level descriptors.	0%	0%	50%	50%
2. Circle your recommended grade 8 Science "Emerging Level" cut score: 61 (25%) 67 (63%) omit (13%)				
3. To what degree do you support the recommended cut score for Grade 8 Science "Emerging Level?"	0%	0%	50%	50%
If you cannot support, please explain why not:				
4. Circle your recommended grade 8 Science "Developing Level" cut score: 91 (25%) 95 (63%) omit (13%)				
5. To what degree do you support the recommended cut score for Grade 8 Science "Developing Level?"	0%	0%	50%	50%
If you cannot support, please explain why not:				
6. Circle your recommended grade 8 Science "Novice Level" cut score: 101 (25%) 103 (63%) omit (13%)				
7. To what degree do you support the recommended cut score for Grade 8 Science "Novice Level?"	0%	0%	50%	50%
If you cannot support, please explain why not:				

	Way too low	A bit low	Appropriate	A bit high	Way too high
8. The recommended cut score for Grade 8 Science "Emerging Level" is:	0%	0%	75%	25%	0%
9. The recommended cut score for Grade 8 Science "Developing Level" is:	0%	0%	75%	25%	0%
10. The recommended cut score for Grade 8 Science "Novice Level" is:	0%	0%	75%	25%	0%

	Strongly Disagree	Disagree	Agree	Strongly Agree
11. Discussions were open and honest.	0%	0%	38%	63%
12. I found the score profiles to be helpful.	0%	0%	50%	50%
13. I believe that my opinions were considered and valued.	0%	0%	63%	38%
14. I am confident that the final cut score recommendations reflect the performance level descriptors associated with CoAlt.	0%	0%	75%	25%
15. I am confident that the final cut score recommendations reflect high expectations consistent with the Extended Evidence Outcomes of the Colorado Academic Standards.	0%	0%	75%	25%

16. Please provide any suggestions you have for high school standard setting.

17. Please provide any suggestions you have for the type of guidance needed to help support score interpretation.

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

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

Student Performance Report

Colorado Alternate Assessment

Student: **FIRSTNAME LASTNAME**

SASID: *******6789** Birthdate: **06/06/2002**
School: **EXAMPLE ES (1234)**
District: **EXAMPLE ISD (1234)**

Spring 2014

Social Studies — Grade 7

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

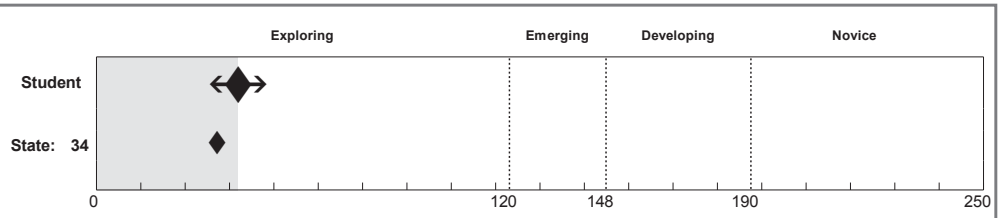
- Your student's performance is represented by a scale score. Scores are placed on a scale so that student performance can be compared across years.
- State averages are provided so that you can compare your student's performance to the performance of others.
- Scores are represented by diamonds. The 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 Child's Score

40

Exploring Command



The Colorado Academic Standards include expectations for student performance. Your child demonstrate an exploring command 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.	18	22	82% 35%
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.	0	16	0% 25%
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.	0	12	0% 18%
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.	0	22	0% 25%

Percent of Points Earned*: 0% 25% 50% 75% 100%

*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 child's mastery of the Extended Evidence Outcomes of the Colorado Academic Standards in Social Studies.

More information on the CoAlt assessment program: 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 test, please visit the Colorado Department of Education's website at **www.cde.state.co.us/standardsandinstruction**

District Performance Level Summary

Colorado Alternate Assessment

Spring 2014

District: **EXAMPLE DISTRICT (1234)**

Science

CONFIDENTIAL - DO NOT DISTRIBUTE

Grade 8

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

	Total Number Tested	Average Scale Score	Performance Levels										Developing and Novice		No Scores Reported
			Inconclusive		Exploring		Emerging		Developing		Novice				
			#	%	#	%	#	%	#	%	#	%	#	%	#
State	**644**	**151**	**13**	**2%**	**68**	**11%**	**274**	**46%**	**214**	**36%**	**30**	**5%**	**244**	**41%**	**45**
District	32	161	0	0%	0	0%	14	48%	15	52%	0	0%	15	52%	3
Gender															
Female	14	159	0	0%	0	0%	6	50%	6	50%	0	0%	6	50%	2
Male	18	162	0	0%	0	0%	8	47%	9	53%	0	0%	9	53%	1
Ethnicity/Race															
Hispanic or Latino	17	161	0	0%	0	0%	7	50%	7	50%	0	0%	7	50%	3
American Indian or Alaska Native	1	185	0	0%	0	0%	0	0%	1	100%	0	0%	1	100%	0
Asian	1	143	0	0%	0	0%	1	100%	0	0%	0	0%	0	0%	0
Black or African-American	1	156	0	0%	0	0%	1	100%	0	0%	0	0%	0	0%	0
White	12	160	0	0%	0	0%	5	42%	7	58%	0	0%	7	58%	0
Native Hawaiian or Other Pacific Islander	0	0	0	0%	0	0%	0	0%	0	0%	0	0%	0	0%	0
Two or more races	0	0	0	0%	0	0%	0	0%	0	0%	0	0%	0	0%	0
Not Indicated	0	0	0	0%	0	0%	0	0%	0	0%	0	0%	0	0%	0
Language Background															
English	25	162	0	0%	0	0%	10	43%	13	57%	0	0%	13	57%	2
Spanish	6	157	0	0%	0	0%	3	60%	2	40%	0	0%	2	40%	1
Other	1	143	0	0%	0	0%	1	100%	0	0%	0	0%	0	0%	0
Not Indicated	0	0	0	0%	0	0%	0	0%	0	0%	0	0%	0	0%	0
Language Proficiency															
Not Applicable	25	162	0	0%	0	0%	10	43%	13	57%	0	0%	13	57%	2
NEP	7	155	0	0%	0	0%	4	67%	2	33%	0	0%	2	33%	1
LEP	0	0	0	0%	0	0%	0	0%	0	0%	0	0%	0	0%	0
FEP	0	0	0	0%	0	0%	0	0%	0	0%	0	0%	0	0%	0
PHLOTE	0	0	0	0%	0	0%	0	0%	0	0%	0	0%	0	0%	0
FELL	0	0	0	0%	0	0%	0	0%	0	0%	0	0%	0	0%	0
Not Indicated	0	0	0	0%	0	0%	0	0%	0	0%	0	0%	0	0%	0

Content Standards Roster

Colorado Alternate Assessment

Spring 2014

School: **SAMPLE SCHOOL (5098)**

District: **SAMPLE DISTRICT (1080)**

Social Studies

CONFIDENTIAL - DO NOT DISTRIBUTE

Grade 7

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	191 - 250
Developing	163 - 190
Emerging	134 - 162
Exploring	1 - 133

	Overall Scale Score	Content Standards Performance			
		History	Geography	Economics	Civics
		Points Possible			
		22	16	12	22
		Percent of Points Earned			
State Average	151	73%	75%	74%	77%
District Average	172	86%	94%	83%	91%
School Average	172	86%	94%	83%	91%
STUDENT NAME	**Overall Performance Level**				
LASTNAME, FIRSTNAME M.	Developing	172	86%	94%	83%	91%
LASTNAME, FIRSTNAME M.	No Score					

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