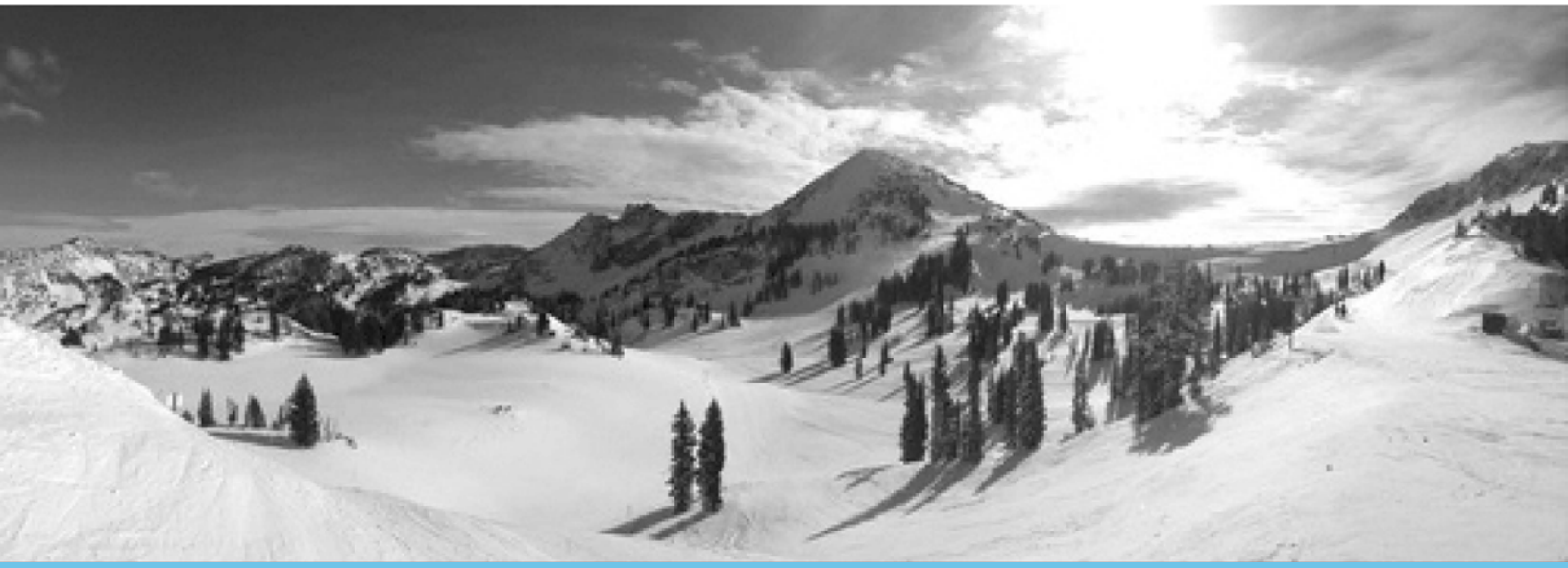
Colorado Alternate Assessment Program



Technical Report

Science and Social Studies

2018

Colorado Alternate Assessment (CoAlt): Science and Social Studies Technical Report 2017–2018

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 19
CHAPTER 7: REPORTING 21
Description of Scores *21*
Score Reports *22*
CHAPTER 8: CALIBRATION, EQUATING, AND SCALING 23
Calibration *23*
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 33
Sources of Validity Evidence *33*
Fairness *35*

PART II: STATISTICAL SUMMARIES **37**

CHAPTER 1: OPERATIONAL ITEMS 38
Administration Summary *38*
Calibration Results *38*
Performance Results *38*
Validity Evidence *39*
CHAPTER 2: EMBEDDED FIELD TEST ITEMS 40
Field Test Items *40*
Data Review *40*

REFERENCES **41**

CoAlt: Science and Social Studies Tables 7–51 ..... **43**

Appendices **103**

APPENDIX A: COALT: SCIENCE AND SOCIAL STUDIES ELIGIBILITY GUIDELINES 104
APPENDIX B: COALT: SCIENCE AND SOCIAL STUDIES TEST BLUEPRINTS 107

APPENDIX C: COALT: SCIENCE AND SOCIAL STUDIES SAMPLE SCORE REPORTS 110
APPENDIX D: COALT ALIGNMENT STUDY REPORT 115

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 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 2018 CoAlt: Science and Social Studies assessments, including content, assessment development, administration, scoring, and technical attributes.

The *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 2018 administrations, including results for both the operational items and the embedded field test items.

Overview of CoAlt

Purposes of the CoAlt Assessment Program

The Individuals with Disabilities Education Improvement Act of 2004 (IDEA) mandates that all students have access to the general curriculum and be included in each state's accountability system. The Reauthorization of the Elementary and Secondary Education Act (ESEA) in 2015 — also known as the Every Student Succeeds Act (ESSA)—continues to specify 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 general Colorado Measures of Academic Success (CMAS)—even with accommodations may take CoAlt. These students must be identified as having a cognitive disability; however, Intellectual Disability does not have to be the student's primary disability label for IDEA eligibility.

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:

- The student has been evaluated and determined to be eligible to receive special education services and has an IEP.
- The student has documented evidence of a cognitive disability.
- The student has a significant cognitive disability.
- 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 that include both the 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^{next}, 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 scaffolded attempts is reached. Teachers score the student's performance using a four-point scoring rubric that is built into the item.

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

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

CoAlt was administered in the following grades in Spring 2018:

- Social studies: grades 4 and 7 (The social studies assessments are administered on a sampling basis with schools participating once every three years. As a result, one-third of elementary and middle schools were assessed.)
- Science: grades 5, 8, and 11

The Standards

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

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

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

- Science
    - Physical Science: Students know and understand common properties, forms, and changes in matter and energy.
    - Life Science: Students know and understand the characteristics and structure of living things, the processes of life, and how living things interact with each other and their environment.
    - Earth Systems Science: Students know and understand the processes and interactions of Earth's systems and the structure and dynamics of Earth and other objects in space.
- Social Studies
    - History: Students develop moral understanding, define identity, and gain 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: Students gain an understanding of spatial perspectives and technologies for spatial analysis, awareness of interdependence of world regions and resources, and learn how places are connected at local, national, and global scales.

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

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

After the newly-developed items were field tested and the item performance data were obtained, the items went through data review where CDE assessment specialists evaluated item performance to recommend if an item should be accepted or rejected based on the student performance data. The items that were accepted were re-classified in the item bank as available for use in future operational assessments. Following the first operational administration of the ES/MS assessments in Spring 2014, performance standards were set and final cut scores were used for reporting purposes. Following the first operational administration for HS in Fall 2014, performance standards were recommended but no social studies cut scores were approved and science cut scores were approved for one year only and used for student-level reporting purposes.

Assessment Development Partners

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

Colorado Department of Education

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

Colorado Educator Community

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

Pearson

Pearson is responsible for the content development, logistics of 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, Senior Partner, The Colorado Education Initiative
- Dr. Jonathan Dings, Executive Director of Student Assessment and Program Evaluation, Boulder Valley School District
- Dr. Michael Kolen, Consultant
- Dr. Lisa Escarcega, Executive Director, Colorado Association of School Executives
- 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 item 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), was 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 is a tiered, inter-related process. The CoAlt 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 item-writing process included the following steps:

Specifications Development

Pearson created the test blueprint with input and approval from CDE. The CoAlt: Science and Social Studies test blueprints contain the number of test items by content standard and item type. The blueprints can be found in Appendix B. During this stage, Pearson also created an IDP which delineates the target number of items per standard, grade level expectation (GLE), and EEO. The IDP helped to forecast the number of items that were needed to create a robust operational item bank that would be refreshed over time.

Item Development

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

Item Reviews

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

Content and Editorial Review

Pearson's Assessment Development Services Department conducted a content review to evaluate standard and knowledge-and-skill match, quality of the items, adherence to the UD principles, 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.
- Evaluate changing roles and attitudes toward various groups.
- Review the role of language in setting and changing attitudes toward various groups.
- 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.
- Edit for inappropriate language usage or stereotyping with regard to sex, race, culture, ethnicity, class, or geographic region.

These reviews were conducted to ensure that all students would have an equal opportunity to demonstrate achievement regardless of their gender, ethnic background, religion, socio-economic

status, or geographic region. Items that were accepted based on the Pearson reviews were re-classified in ITTB as ready for CDE review.

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

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

Data Review

After the development of the items, selected items are placed on the operational test in embedded field-test positions. Following the operational administration, CDE and Pearson assessment specialists and psychometricians review student performance on the field-test items. Pearson provides the results of all statistical analyses. These analyses include classical statistics, item response theory, and differential item functioning (DIF) statistics so that CDE and Pearson can make informed judgments. The statistical information provided include:

- 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.
- DIF statistics by gender and ethnicity when group sample sizes are sufficient.

Student performance data are reviewed to determine if item performance is acceptable for the item to be used on future operational assessments. If a significant number of items are 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 that are accepted based on the evaluation of student performance are re-classified in the item bank as available for use on future operational assessments. Items that are rejected are 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.

Items that pass all stages of the development process (e.g., content and bias review, field test, and data review) are placed in the operational item bank to become eligible for use in future assessments.

Item Bank Statistics

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

The item summary statistics include the item sample size, item mean score, item-total correlation, response distribution that presents the percentage of students achieving each score point both overall and by ability level, and DIF classification for specific subgroups. 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
- efficient and deliberate use of varied content representative of the knowledge and skills in the content standards
- 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
- thorough review of individual items to establish the content within items are up to date and relevant
- selection of items with various stimuli types throughout the test form to enhance the test-taker experience by providing variation in the items presented
- 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 a proposed set of field test items
  - conduct content and psychometric reviews of each proposed set of items
  - construct a customer test map that provides content and psychometric information for each proposed item
  - manage the customer review process
  - provide the customer with copies of proposed items and the associated customer test map
  - revise the proposed item set based on customer comments
  - document edits/comments provided by the customer

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

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

Table 1. CoAlt: Science and Social Studies Operational Assessments

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

CHAPTER 4: TEST ADMINISTRATION PROCEDURES

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

Manuals

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

- *CMAS and CoAlt: Science and Social Studies Procedures Manual*
- *CoAlt: Science and Social Studies Examiner's Manual*
- *CMAS and CoAlt: Science and Social Studies Data Supplement*
- *PearsonAccess^{next} User Guide*

Training

CDE and Pearson conducted several in-person administration trainings for District Assessment Coordinators and district based special education staff across Colorado. CoAlt training materials were posted to the Support tab of PearsonAccess^{next} and on the CDE website 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^{next} 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 and CDE provides extensive training on how to implement accommodations. 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)
- Sign language
- Translation into student's native language
- Other
- None

Test Security

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

Selected Response (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

Supported Performance Tasks (SPTs) 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^{next}, 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^{next}. The points for the three prompts are added together to provide one score for the SPT item, with the maximum of 6 points possible. On the CoAlt: Science and Social Studies assessments, SR items never have more than three answer options, but the number of answer options available for the SPT items can vary by item and content area.

CHAPTER 6: STANDARD SETTING

To support the interpretation of student results, student performance on the CoAlt: Science and Social Studies assessments is described in terms of four performance levels: Advanced, At Target, Approaching Target, and Emerging (initial performance level labels were Novice, Developing, Emerging, Exploring). After the first operational administration of the ES/MS assessments in Spring 2014 and of the HS science assessment in Fall 2014, a standard setting meeting was held 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 in this methodology to help panelists think about the level of content knowledge that may be needed for a student to earn a specific rubric score, the patterns of performance (i.e., combinations of item scores) that lead to overall test scores, and whether various scoring patterns make sense for a given performance level. Different patterns of student performance, called score profiles, were presented to panelists with this approach. The score profile is a graphical representation of how a student could achieve a specific test score.

The standard setting meetings included 8–10 panelists for each committee. Panelists were grouped into tables of three within each meeting room. Panelists were selected for participation by CDE to represent the state in terms of gender and ethnicity as well as relevant demographic characteristics (e.g., school size, geographic location). The CoAlt panelists included educators who taught at the specific grade level, including special educators with experience working with students with significant cognitive disabilities, special educators with experience working with students with other types of disabilities, and content experts with knowledge of the subject-area curriculum. In addition to classroom teachers, special education administrators and higher education representatives also participated in the meetings. Panelists from the CMAS Science and Social Studies standard setting meetings were also recruited to participate. 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.

During the meetings, panelists from the standard setting committees received training on the assessment and the standard setting process, reviewed the grade-level PLDs, reviewed the operational items, reviewed the threshold student descriptors, and applied the Modified Extended Angoff method to establish cut score recommendations across three rounds of rating. During the process of establishing cut score recommendations, panelists also reviewed the content assessed by the CoAlt items and the concepts and skills found in the PLDs, engaged in table and

committee-level discussions, and considered the impact of their cut scores on student performance when making their cut score recommendations.

The proposed recommended cut scores from standard setting were presented to the State Board of Education for review and approval. The full standard setting report for the ES/MS assessments can be found in the *Spring 2014 CoAlt: Science and Social Studies Technical Report*. The full standard setting report for the HS assessment can be found in the *Spring 2015 CoAlt: Science and Social Studies Technical Report*.

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 *Interpretive Guide* at http://www.cde.state.co.us/assessment/cmas_coalt_interpretive_guide_2018.

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:

- Advanced
- At Target
- Approaching Target
- Emerging

These labels were updated in 2016 to match the levels used in the English Language Arts and Mathematics assessments. The cut scores and PLDs were not changed. 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 Emerging 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 C.

Student Performance Reports

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

Aggregate Reports

Two types of aggregate reports are produced for CoAlt:

- Content Standards Roster
- Performance Level Summary

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

CHAPTER 8: CALIBRATION, EQUATING, AND SCALING

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

Calibration

The Rasch Partial Credit Model

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

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

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

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

Equating and Scaling

Equating involves adjusting for differences in the difficulty of test forms, both within and across assessment administrations. Equating makes certain that students taking one form of a test are neither advantaged nor disadvantaged when compared to students taking a different form. Each time a new test form is constructed, equating is used to allow scores on the new form to be comparable to scores on the previous form by placing the scores on both forms on the same scale. It is the underlying Rasch scale obtained from calibration that facilitates equating of test forms. The Rasch scale can then be transformed to create scale scores to allow for the interpretation of test scores. The RPCM and Winsteps (Linacre, 2011) were used for all equating analyses.

Equating and Scaling

Two equating models were used with the Spring 2018 assessments. A pre-equating model was used for grades 4 and 7, and a post-equating model was used for grades 5, 8, and 11. The pre-equating process is one in which a newly-developed test form is linked through equating, before it is administered, to a set of items that appeared previously on one or more operational test forms. The pre-equated score tables that are created are then applied to the operational test administration results and no calibrations of the operational tests are necessary. By using this process, the difficulty level of the test form is known prior to its administration, and the anticipated raw scores that correspond to scale scores at performance standards can be identified. This model was chosen for social studies because the social studies assessments are administered on a sampling basis with schools participating once every three years. As a result, approximately one-third of the CoAlt population was assessed during the spring administration which is roughly 200 students per grade. It was determined that the pre-equating parameters based on a full student population would be more reliable than post-equating with the smaller sample (de Ayala, 2009).

For the post-equating process for the grades 5, 8, and 11 assessments, the fixed common items approach was used to equate the Spring 2018 grades 5, 8, and 11 assessments. The operational items used to equate the 2018 science assessments to the base scales are called anchor items. The anchor items are a set of common items that are already equated to the base scale and are placed on forms from adjacent administrations. This set of items represents the CoAlt blueprint in terms of content and item types and represents approximately 60% of a full form. To obtain equated Rasch parameter estimates for the Spring 2018 science assessments, anchor item parameter estimates were fixed to their previously equated item parameter estimates before calibrating the remaining non-anchor operational items. This method placed the non-anchor operational items on the same scale as the anchor items.

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

item from the anchor item set. It is important to also make sure that the remaining anchor set continues to be representative of the overall content and structure of the test.

Ability Estimates

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

Scale Scores

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

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

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

Steps in the Calibration and Scaling Process

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

Data Preparation

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

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

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

- Item sample size
- Response distribution
- Item mean score
- Item-total correlation

Calibration

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

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

CHAPTER 9: RELIABILITY

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

Cronbach's Alpha

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

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

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

where n is the number of items,

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

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

Cronbach's alpha ranges in value from 0.0 to 1.0, where higher values indicate a greater proportion of observed score variance is true score variance. Two factors affect estimates of internal consistency: test length and homogeneity of items. The longer the test, the more observed score variance is likely to be true score variance. The more similar the items, the more likely examinees will respond consistently across items within the test. For CoAlt, coefficient alpha estimates are provided for the overall test as well as for subgroups. The coefficient alpha estimates can be found in Tables 7–23.

Standard Error of Measurement

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

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

where s_x is the standard deviation of test scores and

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

There is an inverse relationship between the reliability coefficient (e.g., alpha) and SEM: the higher the reliability, the lower the SEM. SEM values can be found in Tables 7–23.

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 41–45.

Decision Consistency and Accuracy

The CoAlt: Science and Social Studies scales are divided into four performance levels: Advanced, At Target, Approaching Target, and Emerging. 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 46.

Several factors can affect classification consistency results. One factor is the number of performance levels. PC values using multi-level classification, the method used for CoAlt, are typically lower than PC values using binary classification because applying all the cut scores simultaneously allows for more opportunities for misclassifications due to the increased number of performance levels (Lee, Hanson, & Brennan, 2002; Wan, Brennan, & Lee, 2007). The distribution of observations in performance level categories can also affect consistency results. Prevalence deals with the affect that the distribution of observations can have on the magnitude of kappa values. Increased prevalence values found when evaluating the data indicate that the distribution of observations in categories is likely affecting kappa. It has been found that as the prevalence value increases the chance value also increases which results in a smaller kappa value (Bryt, Bishop, & Carlin, 1993).

Inter-Rater Agreement

An additional form of reliability, called inter-rater agreement, is also evaluated for CoAlt administrations. Inter-rater agreement examines the extent to which examinees would obtain the same score if scored by different scorers. For this method, two raters simultaneously observe a student taking the CoAlt assessment: a test examiner (i.e., a student's teacher) and a score monitor. Both raters evaluate student performance and enter their scores into the online score

entry system. The two independent ratings are then compared to determine the consistency of the ratings. The second set of scores provided by the score monitor is used only to establish the level of consistency in scoring. They are not used for student scoring and reporting.

Procedure

The sampling plan included five score monitors each conducting observations which would yield approximately 24 students with second scores for the CoAlt science assessments (grades 5, 8, and 11). Pearson selected five assessment specialists to serve as score monitors during the CoAlt assessment window. The assessment specialists were familiar with administering alternate assessments, including the CoAlt assessments, and familiar with the population of students who take alternate assessments.

Once Pearson selected the score monitors, they received training developed by CDE and Pearson staff via teleconference. Score monitors participated in training so that they would be consistent in their methods and scoring when conducting their observations. As part of the training, the meeting facilitator reviewed the purpose of score monitoring, the test materials, the scoring process, and the test administration procedures. In addition, the facilitator also reviewed the score monitor observation materials and the document to be used to obtain scores, descriptions of the testing environment, and test procedures used by the student's test administrator.

Pearson and CDE worked together to recruit schools to participate in score monitoring. Pearson contacted schools so that the sample of observed students would be representative of the geographic regions of the state and diverse in terms of gender and ethnicity. In addition, Pearson determined the number of students participating in the CoAlt assessment at each school and attempted to schedule observations in order to mitigate any impact on the classroom.

Ultimately, the five score monitors conducted 48 observations across nine school districts in Colorado. The school districts were located in four of the eight geographic regions of the state. During the Spring 2018 administration, 16 observations were conducted for grade 5 science, 17 observations were conducted for grade 8 science, and 15 observations were conducted for grade 11 science.

Results

In general, the score monitors indicated that the test administrators seemed comfortable with the students, were well prepared for administering the test, and provided accommodations that were appropriate for the student. They also noted that the testing rooms had adequate space and were free of visible materials that could provide assistance to test questions. However, some challenges were noted by the score monitors. For example, there were instances where a test administrator did not prepare the task manipulatives ahead of testing or did not always follow the scaffolding and the standardized script provided in the test books. CDE noted the issues and uses the score monitor feedback as part of test administrator training sessions.

The metrics used to evaluate inter-rater agreement were the correlation between two independent ratings, perfect agreement, and adjacent agreement. Correlations are used to evaluate the

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

Table 4. Descriptive Statistics for CoAlt: Science

	Population			Sample		
	N	Males	Females	N	Males	Females
Grade 5 Science	590	64%	36%	14	57%	43%
Grade 8 Science	617	63%	37%	17	59%	41%
HS Science	511	61%	39%	15	67%	33%

Correlation coefficients were calculated for each assessment. There were instances where students were tested across multiple days, and as a result, the score monitors were unable to observe the student taking all of the test items. This led to instances where second scores were missing. When this occurred, only those items that had scores from both the score monitor and the test administrator were included in the analysis.

The correlation of the item-level scores between the first and second scores ranged from 0.90 to 0.97 across grades 5, 8, and HS (see Table 5). The correlation of the test-level scores was also calculated. The correlation of the test-level scores for grades 5 and 8 science was 0.99, and the correlation for HS science was 70. The correlation coefficients indicate that the first and second scores are highly correlated for grades 5 and 8 science and less highly correlated for HS science. When noting their observations of the HS Test Examiners, score monitors described more instances where the Test Examiner did not always adhere to the scaffolding provided in the test books. This would result in more differences between the first and second scores for HS which could explain the lower correlations for HS science compared to grades 5 and 8 science.

Table 5. Correlations between First and Second Scores

	Item-Level Correlation	Test-Level Correlation
Grade 5 Science	0.94	0.99
Grade 8 Science	0.97	0.99
HS Science	0.90	0.70

Perfect and adjacent agreement rates were calculated and are summarized in Table 6. Perfect agreement rates of item-level scores were 91% for grade 5 science, 94% for grade 8 science, and 89% for HS science. Adjacent agreement rates of item-level scores ranged from 5% to 6% across the assessments.

Table 6. Percent Agreement between First and Second Scores

	Perfect Agreement	Adjacent Agreement
Grade 5 Science	91%	6%
Grade 8 Science	94%	5%
HS Science	89%	6%

The perfect and adjacent agreement rates indicate relatively high levels of agreement between scores. When perfect and adjacent agreement rates are combined, 95% to 99% of the ratings had the same or adjacent item scores. As with the correlations, HS science has smaller agreement rates than grades 5 and 8 science.

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., 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.

A formal alignment study was also conducted for CoAlt in 2016. The alignment study was conducted to demonstrate that CoAlt represents the full range of the EEOs and measures student knowledge in the same manner and at the same level of complexity as specified in the EEOs. The Human Resources Research Organization (HumRRO) was contracted by Pearson on behalf of CDE to conduct the independent alignment study. For the study, two panels (one per content area) of Colorado educators were convened to review the alignment between the CoAlt science and social studies items and the EEOs for science and social studies. Every effort was made to produce panels consisting of teachers reflecting the population of students who take the assessments. To conduct the content alignment review, HumRRO applied the Webb (2005) alignment method. This procedure is based on four indicators (or statistics) using the data gathered from two major tasks panelists are asked to complete: (a) providing depth of knowledge (DOK) ratings for the each of the EEOs for science and social studies, and (b) evaluating the science and social studies items by matching them to grade level EEOs, providing an item DOK rating, and selecting a rating of the overall alignment between item and standard.

The cumulative results of the study provide validity evidence to support that the content of the CoAlt science and social studies test items match the intended content as specified in the EEOs. Panelists from both content committees tended to agree that items were measuring the intended grade level expectations, and to rate items as highly aligned to the EEOs. Additional analyses by HumRRO found that panelists indicated that the CoAlt items reflect the intended content of the test blueprints, and that the large majority of items are highly aligned to the particular EEOs to which they were matched. Plans are also ongoing to evaluate the content coverage of the assessments as a way to possibly strengthen the alignment between the components of the assessment system. The full CoAlt alignment study report can be found in Appendix D.

Evidence Based on Response Processes

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

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

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

Evidence Based on Internal Structure

The internal structure of an assessment pertains to the degree to which the items on an assessment measure one underlying construct. When assessments are designed to measure one underlying construct, the internal components of the assessments should exhibit a high degree of homogeneity that can be measured in terms of the internal consistency estimates of reliability. As a result, the internal consistency for the CoAlt: Science and Social Studies assessments is

evaluated using reliability coefficients. These internal consistency estimates are described in Chapter 9 and provided for the overall test and various student subgroups in Part II of this report.

Evidence Based on Relations to Other Variables

Evidence was collected showing the correlation between student assessment scores and variables related to the student. Student scale scores were correlated with teachers' responses for several test validity questions to determine the strength of relationship between the variables. The test validity questions are variables related to the student (e.g., How familiar are you with this student?, How many hours per week does this student spend in instruction on this content area?, and Approximately how much instructional time for this content area is in the general education classroom?). The correlations between student scores and the familiarity of the test examiner with the student are small and indicate no meaningful to a weak relationship between the variables. The correlations between student scores and the instructional hours and instructional time variables are low to moderate positive correlations and indicate stronger relationships between student scores and the instructional hours and instructional time variables. The correlation coefficients between the student scores and the variables related to the student can be found in Tables 48–50.

Evidence for Validity and Consequences of Testing

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

Fairness

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

Universal Design

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

Differential Item Functioning

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

Accessibility and Accommodations

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

Released Items

Released items provide the opportunity for teachers and students to become familiar with the test design and scoring of the assessments before experiencing the items on an operational test. Teachers and students were provided the opportunity to experience sample items prior to the first operational administration of CoAlt and before each subsequent test administration.

PART II: STATISTICAL SUMMARIES

This section contains an overview of the statistical summaries for the Spring 2018 administration. Administration summaries, calibration results, performance results, reliability evidence, and validity evidence are included for the operational items. Test form summaries and item performance review outcomes are provided for the embedded field test items.

CHAPTER 1: OPERATIONAL ITEMS

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

Administration Summary

Approximately 2,100 students took the CoAlt: Science and Social Studies assessments. Tables 7–23 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 0.92 to 0.97. The SEM values for the total group ranged from 3.79 to 4.57.

Calibration Results

Item Statistics

Tables 24–28 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 29–33 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 34–35. The scale score distributions for each assessment are shown in Tables 36–40. Tables 41–45 are provided and include the raw score, scale score, and CSEM values.

Decision Consistency and Accuracy

Table 46 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 47 provides the summary of teachers' responses to the test validation questions for each assessment.

Correlations Between Student Scores and Variables Related to the Student

Tables 48–50 provide correlation coefficients related to validity evidence based on relations to other variables. Student scale scores were correlated with teachers' responses for several test validity questions to determine the strength of relationship between the variables.

CHAPTER 2: EMBEDDED FIELD TEST ITEMS

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

Field Test Items

Field test items were included on each operational test form. Thirty-four 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 51 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 a significant number of 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. Across the 34 field test items, no items were flagged. Table 51 shows the number of items field tested on each test and 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.

Byrt. T, Bishop. J, Carlin. J.B. Bias, prevalence and kappa. Journal of Clinical Epidemiology. 1993; 46 (5), 423-429.

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.

de Ayala, R. J. (2009). The theory and practice of item response theory. New York: The Guilford Press.

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.

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

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.

Wan, L., Brennan, R. L., & Lee W. (2007). Estimating classification consistency for complex assessments (CASMA Research Report No. 22). Iowa City, IA: University of Iowa.

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

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

COALT: SCIENCE AND SOCIAL STUDIES TABLES 7–51

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

Content	Grade	Subgroup	N	%	Scale Score				Raw Score				Alpha	SEM
					Mean	SD	Min	Max	Mean	SD	Min	Max		
SS	4	Total	190	100	146.16	32.68	0	214	48.78	15.52	0	70	0.93	4.20
		Female	72	37.89	147.67	33.04	0	195	49.47	15.96	0	68	0.94	4.12
		Male	118	62.11	145.24	32.56	0	214	48.36	15.29	0	70	0.92	4.25
		American Indian	1	0.53	-	-	-	-	-	-	-	-	-	-
		Asian	8	4.21	-	-	-	-	-	-	-	-	-	-
		Black or African American	11	5.79	-	-	-	-	-	-	-	-	-	-
		Hispanic or Latino	76	40.00	143.38	37.33	0	195	47.99	16.26	0	68	0.93	4.21
		White	87	45.79	148.41	28.94	26	195	49.79	14.67	2	68	0.92	4.25
		Native Hawaiian or other Pacific Islander	1	0.53	-	-	-	-	-	-	-	-	-	-
		Two or More Races	6	3.16	-	-	-	-	-	-	-	-	-	-
		Missing	0	0.00	-	-	-	-	-	-	-	-	-	-
	7	Total	193	100	156.00	31.83	0	250	55.15	13.32	0	72	0.92	3.79
		Female	72	37.31	160.54	20.45	116	227	57.60	8.24	32	71	0.80	3.66
		Male	121	62.69	153.31	36.79	0	250	53.69	15.43	0	72	0.94	3.85
		American Indian	2	1.04	-	-	-	-	-	-	-	-	-	-
		Asian	4	2.07	-	-	-	-	-	-	-	-	-	-
		Black or African American	17	8.81	143.94	49.50	0	198	51.59	19.73	0	69	0.97	3.59
		Hispanic or Latino	76	39.38	152.78	28.44	0	227	54.38	12.33	0	71	0.91	3.79
		White	81	41.97	162.37	29.79	87	250	56.69	12.42	15	72	0.91	3.80
		Native Hawaiian or other Pacific Islander	1	0.52	-	-	-	-	-	-	-	-	-	-
		Two or More Races	12	6.22	-	-	-	-	-	-	-	-	-	-
		Missing	0	0.00	-	-	-	-	-	-	-	-	-	-

Table 8. 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	588	100	147.50	33.86	0	223	50.25	15.87	0	71	0.94	4.01
		Female	212	36.05	149.10	32.71	0	206	51.09	15.42	0	70	0.93	3.95
		Male	376	63.95	146.60	34.50	0	223	49.78	16.12	0	71	0.94	4.04
		American Indian	6	1.02	-	-	-	-	-	-	-	-	-	-
		Asian	12	2.04	-	-	-	-	-	-	-	-	-	-
		Black or African American	39	6.63	148.13	25.45	41	184	50.39	14.43	2	67	0.92	4.02
		Hispanic or Latino	261	44.39	146.34	34.59	0	196	49.89	16.09	0	69	0.94	3.99
		White	241	40.99	147.08	36.15	0	206	50.02	16.62	0	70	0.94	4.05
		Native Hawaiian or other Pacific Islander	2	0.34	-	-	-	-	-	-	-	-	-	-
		Two or More Races	26	4.42	156.08	23.57	110	223	53.12	12.16	21	71	0.90	3.86
		Missing	1	0.17	-	-	-	-	-	-	-	-	-	-
	8	Total	616	100	148.46	32.35	0	250	82.02	21.57	0	108	0.96	4.57
		Female	228	37.01	147.91	32.08	0	250	81.74	21.85	0	108	0.96	4.54
		Male	388	62.99	148.79	32.54	0	231	82.19	21.44	0	107	0.95	4.59
		American Indian	6	0.97	-	-	-	-	-	-	-	-	-	-
		Asian	13	2.11	-	-	-	-	-	-	-	-	-	-
		Black or African American	58	9.42	151.98	22.21	85	213	83.38	17.12	22	106	0.92	4.78
		Hispanic or Latino	238	38.64	147.03	34.93	0	231	81.30	23.09	0	107	0.96	4.43
		White	271	43.99	149.17	32.40	0	250	82.42	21.33	0	108	0.95	4.64
		Native Hawaiian or other Pacific Islander	1	0.16	-	-	-	-	-	-	-	-	-	-
		Two or More Races	28	4.55	143.25	36.14	0	195	78.93	24.34	0	104	0.96	4.71
		Missing	1	0.16	-	-	-	-	-	-	-	-	-	-

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

Content	Grade	Subgroup	N	%	Scale Score				Raw Score				Alpha	SEM
					Mean	SD	Min	Max	Mean	SD	Min	Max		
SC	HS	Total	512	100	152.17	35.87	0	250	83.94	24.50	0	110	0.97	4.29
		Female	201	39.26	153.40	34.91	0	208	85.14	23.76	0	108	0.97	4.36
		Male	311	60.74	151.37	36.52	0	250	83.16	24.97	0	110	0.97	4.25
		American Indian	7	1.37	-	-	-	-	-	-	-	-	-	-
		Asian	16	3.13	155.69	22.31	97	199	85.56	19.55	25	107	0.95	4.48
		Black or African American	36	7.03	148.94	29.65	0	177	82.03	20.05	0	102	0.95	4.65
		Hispanic or Latino	193	37.70	154.55	34.62	0	250	86.04	22.64	0	110	0.97	4.13
		White	242	47.27	150.02	39.44	0	225	82.08	27.12	0	109	0.97	4.35
		Native Hawaiian or other Pacific Islander	1	0.20	-	-	-	-	-	-	-	-	-	-
		Two or More Races	16	3.13	154.63	24.57	83	184	84.94	23.02	16	104	0.96	4.71
		Missing	1	0.20	-	-	-	-	-	-	-	-	-	-

Table 10. Descriptive Statistics by Economically Disadvantaged

Content	Grade	Subgroup	N	%	Scale Score				Raw Score				Alpha	SEM
					Mean	SD	Min	Max	Mean	SD	Min	Max		
SS	4	No	77	40.53	144.62	35.23	1	189	48.26	16.39	1	67	0.93	4.26
		Yes	113	59.47	147.20	30.93	0	214	49.13	14.96	0	70	0.92	4.16
		Missing	0	0.00	-	-	-	-	-	-	-	-	-	-
	7	No	84	43.52	151.41	35.61	0	227	53.43	14.97	0	71	0.93	3.84
		Yes	109	56.48	159.55	28.23	42	250	56.47	11.79	4	72	0.90	3.73
		Missing	0	0.00	-	-	-	-	-	-	-	-	-	-
SC	5	No	247	42.01	143.93	37.12	0	206	48.33	17.19	0	70	0.94	4.19
		Yes	341	57.99	150.06	31.15	0	223	51.63	14.75	0	71	0.93	3.88
		Missing	0	0.00	-	-	-	-	-	-	-	-	-	-
	8	No	257	41.72	145.68	32.53	0	250	79.96	22.15	0	108	0.96	4.72
		Yes	359	58.28	150.40	32.15	0	213	83.45	21.08	0	106	0.96	4.46
		Missing	0	0.00	-	-	-	-	-	-	-	-	-	-
	HS	No	237	46.29	147.98	38.97	0	208	80.72	27.07	0	108	0.97	4.43
		Yes	275	53.71	155.68	32.67	0	250	86.63	21.76	0	110	0.96	4.18
		Missing	0	0.00	-	-	-	-	-	-	-	-	-	-

Table 11. 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	141	74.21	146.68	39.40	0	189	48.63	15.75	0	70	0.93	4.23
			NEP	36	18.95	134.48	58.48	0	195	50.03	15.20	12	68	0.93	3.93
			LEP	8	4.21	-	-	-	-	-	-	-	-	-	-
			PHLOTE	2	1.05	-	-	-	-	-	-	-	-	-	-
			FELL	0	0.00	-	-	-	-	-	-	-	-	-	-
			FEP Monitored Y1	2	1.05	-	-	-	-	-	-	-	-	-	-
			FEP Monitored Y2	0	0.00	-	-	-	-	-	-	-	-	-	-
			FEP Exited Y1	1	0.53	-	-	-	-	-	-	-	-	-	-
			FEP Exited Y2	0	0.00	-	-	-	-	-	-	-	-	-	-
			Missing	0	0.00	-	-	-	-	-	-	-	-	-	-

Table 12. 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	151	78.24	156.66	32.57	0	250	55.05	13.84	0	72	0.92	3.81
			NEP	26	13.47	151.54	34.85	0	191	54.77	13.43	0	68	0.92	3.81
			LEP	4	2.07	-	-	-	-	-	-	-	-	-	-
			PHLOTE	3	1.55	-	-	-	-	-	-	-	-	-	-
			FELL	0	0.00	-	-	-	-	-	-	-	-	-	-
			FEP Monitored Y1	3	1.55	-	-	-	-	-	-	-	-	-	-
			FEP Monitored Y2	5	2.59	-	-	-	-	-	-	-	-	-	-
			FEP Exited Y1	0	0.00	-	-	-	-	-	-	-	-	-	-
			FEP Exited Y2	1	0.52	-	-	-	-	-	-	-	-	-	-
			Missing	0	0.00	-	-	-	-	-	-	-	-	-	-

Table 13. 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	449	76.36	147.89	34.69	0	223	50.36	16.37	0	71	0.94	4.01
			NEP	102	17.35	148.25	30.52	0	189	51.08	13.35	0	68	0.91	3.94
			LEP	12	2.04	-	-	-	-	-	-	-	-	-	-
			PHLOTE	10	1.70	-	-	-	-	-	-	-	-	-	-
			FELL	2	0.34	-	-	-	-	-	-	-	-	-	-
			FEP Monitored Y1	6	1.02	-	-	-	-	-	-	-	-	-	-
			FEP Monitored Y2	2	0.34	-	-	-	-	-	-	-	-	-	-
			FEP Exited Y1	3	0.51	-	-	-	-	-	-	-	-	-	-
			FEP Exited Y2	2	0.34	-	-	-	-	-	-	-	-	-	-
			Missing	0	0.00	-	-	-	-	-	-	-	-	-	-

Table 14. Science 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	494	80.19	150.19	31.07	0	250	83.03	20.92	0	108	0.95	4.53
			NEP	92	14.94	147.52	26.13	0	184	82.13	18.32	0	102	0.94	4.62
			LEP	5	0.81	-	-	-	-	-	-	-	-	-	-
			PHLOTE	12	1.95	-	-	-	-	-	-	-	-	-	-
			FELL	7	1.14	-	-	-	-	-	-	-	-	-	-
			FEP Monitored Y1	6	0.97	-	-	-	-	-	-	-	-	-	-
			FEP Monitored Y2	0	0.00	-	-	-	-	-	-	-	-	-	-
			FEP Exited Y1	0	0.00	-	-	-	-	-	-	-	-	-	-
			FEP Exited Y2	0	0.00	-	-	-	-	-	-	-	-	-	-
			Missing	0	0.00	-	-	-	-	-	-	-	-	-	-

Table 15. Science Descriptive Statistics by English Language Proficiency (continued)

Content	Grade	Variable	Subgroup	N	%	Scale Score				Raw Score				Alpha	SEM
						Mean	SD	Min	Max	Mean	SD	Min	Max		
SC	HS	Language Proficiency	Not Applicable	393	76.76	150.19	31.07	0	250	83.15	24.97	0	109	0.97	4.37
			NEP	79	15.43	147.52	26.13	0	184	83.18	24.29	0	106	0.97	4.16
			LEP	13	2.54	-	-	-	-	-	-	-	-	-	-
			PHLOTE	7	1.37	-	-	-	-	-	-	-	-	-	-
			FELL	8	1.56	-	-	-	-	-	-	-	-	-	-
			FEP Monitored Y1	3	0.59	-	-	-	-	-	-	-	-	-	-
			FEP Monitored Y2	6	1.17	-	-	-	-	-	-	-	-	-	-
			FEP Exited Y1	1	0.20	-	-	-	-	-	-	-	-	-	-
			FEP Exited Y2	9	1.76	-	-	-	-	-	-	-	-	-	-
			Missing	0	0.00	-	-	-	-	-	-	-	-	-	-

Table 16. 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	38	20.00	141.84	37.15	0	189	46.76	16.26	0	67	0.93	4.23
		Deaf-Blindness	0	0.00	-	-	-	-	-	-	-	-	-	-
		Developmental Delay	0	0.00	-	-	-	-	-	-	-	-	-	-
		Emotional Disturbance	0	0.00										
		Hearing Impairment	1	0.53	-	-	-	-	-	-	-	-	-	-
		Intellectual Disability	42	22.11	158.38	19.53	114	214	54.45	9.80	26	70	0.86	3.66
		Multiple Disabilities	91	47.89	143.19	32.24	0	180	47.30	16.35	0	65	0.93	4.33
		Not Collected	0	0.00	-	-	-	-	-	-	-	-	-	-
		Orthopedic Impairment	1	0.53	-	-	-	-	-	-	-	-	-	-
		Other Health Impairment	10	5.26	-	-	-	-	-	-	-	-	-	-
		Specific Learning Disability	1	0.53	-	-	-	-	-	-	-	-	-	-
		Speech or Language Impairment	1	0.53	-	-	-	-	-	-	-	-	-	-
		Traumatic Brain Injury	1	0.53	-	-	-	-	-	-	-	-	-	-
		Visual Impairment	4	2.11	-	-	-	-	-	-	-	-	-	-
		Missing	0	0.00	-	-	-	-	-	-	-	-	-	-

Table 17. 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	26	13.47	140.46	31.01	42	198	47.58	15.86	4	69	0.92	4.57
		Deaf-Blindness	0	0.00	-	-	-	-	-	-	-	-	-	-
		Developmental Delay	0	0.00	-	-	-	-	-	-	-	-	-	-
		Emotional Disturbance	0	0.00										
		Hearing Impairment	0	0.00	-	-	-	-	-	-	-	-	-	-
		Intellectual Disability	56	29.02	167.59	23.33	126	227	59.98	7.71	39	71	0.80	3.46
		Multiple Disabilities	91	47.15	152.55	29.91	0	198	54.19	13.51	0	69	0.92	3.73
		Not Collected	0	0.00	-	-	-	-	-	-	-	-	-	-
		Orthopedic Impairment	2	1.04	-	-	-	-	-	-	-	-	-	-
		Other Health Impairment	10	5.18	-	-	-	-	-	-	-	-	-	-
		Specific Learning Disability	2	1.04	-	-	-	-	-	-	-	-	-	-
		Speech or Language Impairment	3	1.55	-	-	-	-	-	-	-	-	-	-
		Traumatic Brain Injury	2	1.04	-	-	-	-	-	-	-	-	-	-
		Visual Impairment	1	0.52	-	-	-	-	-	-	-	-	-	-
		Missing	0	0.00										

Table 18. 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	102	17.35	143.29	36.28	0	206	48.17	15.90	0	70	0.93	4.14
		Deaf-Blindness	0	0.00	-	-	-	-	-	-	-	-	-	-
		Developmental Delay	0	0.00	-	-	-	-	-	-	-	-	-	-
		Emotional Disturbance	5	0.85	-	-	-	-	-	-	-	-	-	-
		Hearing Impairment	2	0.34	-	-	-	-	-	-	-	-	-	-
		Intellectual Disability	155	26.36	158.73	21.24	0	223	56.39	10.60	0	71	0.89	3.56
		Multiple Disabilities	284	48.30	140.39	37.61	0	206	46.43	17.39	0	70	0.94	4.21
		Not Collected	0	0.00	-	-	-	-	-	-	-	-	-	-
		Orthopedic Impairment	4	0.68	-	-	-	-	-	-	-	-	-	-
		Other Health Impairment	23	3.91	158.78	26.77	61	189	55.83	13.63	4	68	0.92	3.78
		Specific Learning Disability	7	1.19	-	-	-	-	-	-	-	-	-	-
		Speech or Language Impairment	4	0.68	-	-	-	-	-	-	-	-	-	-
		Traumatic Brain Injury	2	0.34	-	-	-	-	-	-	-	-	-	-
		Visual Impairment	0	0.00	-	-	-	-	-	-	-	-	-	-
		Missing	0	0.00	-	-	-	-	-	-	-	-	-	-

Table 19. 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	91	14.77	148.105	27.32	0	213	80.50	19.90	0	106	0.94	5.01
		Deaf-Blindness	0	0.00	-	-	-	-	-	-	-	-	-	-
		Developmental Delay	0	0.00	-	-	-	-	-	-	-	-	-	-
		Emotional Disturbance	0	0.00	-	-	-	-	-	-	-	-	-	-
		Hearing Impairment	1	0.16	-	-	-	-	-	-	-	-	-	-
		Intellectual Disability	196	31.82	158.13	22.26	0	250	88.93	13.50	0	108	0.91	4.17
		Multiple Disabilities	293	47.56	141.46	37.21	0	195	77.43	25.06	0	104	0.97	4.68
		Not Collected	0	0.00	-	-	-	-	-	-	-	-	-	-
		Orthopedic Impairment	1	0.16	-	-	-	-	-	-	-	-	-	-
		Other Health Impairment	17	2.76	142.12	43.60	0	184	78.29	28.14	0	102	0.98	4.38
		Specific Learning Disability	11	1.79	-	-	-	-	-	-	-	-	-	-
		Speech or Language Impairment	3	0.49	-	-	-	-	-	-	-	-	-	-
		Traumatic Brain Injury	3	0.49	-	-	-	-	-	-	-	-	-	-
		Visual Impairment	0	0.00	-	-	-	-	-	-	-	-	-	-
		Missing	0	0.00	-	-	-	-	-	-	-	-	-	-

Table 20. Science Descriptive Statistics by Primary Disability (continued)

Content	Grade	Primary Disability	N	%	Scale Score				Raw Score				Alpha	SEM
					Mean	SD	Min	Max	Mean	SD	Min	Max		
SC	HS	Autism	70	13.67	147.83	38.17	0	199	79.50	28.67	0	107	0.98	4.37
		Deaf-Blindness	0	0.00	-	-	-	-	-	-	-	-	-	-
		Developmental Delay	0	0.00	-	-	-	-	-	-	-	-	-	-
		Emotional Disturbance	2	0.39	-	-	-	-	-	-	-	-	-	-
		Hearing Impairment	1	0.20	-	-	-	-	-	-	-	-	-	-
		Intellectual Disability	179	34.96	163.32	24.59	0	250	91.94	16.40	0	110	0.95	3.82
		Multiple Disabilities	230	44.92	143.82	40.42	0	199	78.20	27.03	0	107	0.97	4.63
		Not Collected	0	0.00	-	-	-	-	-	-	-	-	-	-
		Orthopedic Impairment	1	0.20	-	-	-	-	-	-	-	-	-	-
		Other Health Impairment	15	2.93	-	-	-	-	-	-	-	-	-	-
		Specific Learning Disability	5	0.98	-	-	-	-	-	-	-	-	-	-
		Speech or Language Impairment	4	0.78	-	-	-	-	-	-	-	-	-	-
		Traumatic Brain Injury	5	0.98	-	-	-	-	-	-	-	-	-	-
		Visual Impairment	0	0.00	-	-	-	-	-	-	-	-	-	-
		Missing	0	0.00	-	-	-	-	-	-	-	-	-	-

Table 21. 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 and Missing	168	88.42	147.41	32.18	0	214	49.38	15.53	0	70	0.93	4.11
			Yes	22	11.58	136.64	35.61	0	177	44.18	14.98	0	64	0.90	4.75
		Braille	No and Missing	187	98.42	146.39	32.85	0	214	48.95	15.54	0	70	0.93	4.16
			Yes	3	1.58	-	-	-	-	-	-	-	-	-	-
		Eye Gaze	No and Missing	188	98.95	146.72	32.14	0	214	49.07	15.24	0	70	0.92	4.19
			Yes	2	1.05	-	-	-	-	-	-	-	-	-	-
		Modified Picture Symbols	No and Missing	178	93.68	146.75	33.28	0	214	49.25	15.57	0	70	0.93	4.08
			Yes	12	6.32	-	-	-	-	-	-	-	-	-	-
		Objects	No and Missing	181	95.26	146.50	33.18	0	214	49.05	15.63	0	70	0.93	4.14
			Yes	9	4.74	-	-	-	-	-	-	-	-	-	-
		Sign Language	No and Missing	189	99.47	146.93	30.97	0	214	49.04	15.14	0	70	0.92	4.21
			Yes	1	0.53	-	-	-	-	-	-	-	-	-	-
		Translation into Native Language	No and Missing	190	100	146.16	32.68	0	214	48.78	15.52	0	70	0.93	4.20
			Yes	0	0.00	-	-	-	-	-	-	-	-	-	-
		Other	No and Missing	165	86.84	147.50	30.36	0	195	49.49	14.63	0	68	0.92	4.24
			Yes	25	13.16	137.32	44.91	0	214	44.12	20.19	0	70	0.96	3.93
	7	Assistive Technology	No and Missing	183	94.82	157.45	31.79	0	250	55.91	13.00	0	72	0.92	3.69
			Yes	10	5.18	-	-	-	-	-	-	-	-	-	-
		Braille	No and Missing	192	99.48	156.04	31.90	0	250	55.15	13.36	0	72	0.92	3.78
			Yes	1	0.52	-	-	-	-	-	-	-	-	-	-
		Eye Gaze	No and Missing	187	96.89	157.97	28.51	0	250	55.98	11.95	0	72	0.90	3.72
			Yes	6	3.11	-	-	-	-	-	-	-	-	-	-
		Modified Picture Symbols	No and Missing	175	90.67	156.24	30.19	0	227	55.50	12.67	0	71	0.91	3.74
			Yes	18	9.33	153.72	45.91	42	250	51.72	18.65	4	72	0.95	4.10
		Objects	No and Missing	188	97.41	156.46	31.98	0	250	55.38	13.31	0	72	0.92	3.77
			Yes	5	2.59	-	-	-	-	-	-	-	-	-	-
		Sign Language	No and Missing	190	98.45	156.26	31.94	0	250	55.26	13.32	0	72	0.92	3.76
			Yes	3	1.55	-	-	-	-	-	-	-	-	-	-
		Translation into Native Language	No and Missing	192	99.48	156.04	31.90	0	250	55.15	13.36	0	72	0.92	3.78
			Yes	1	0.52	-	-	-	-	-	-	-	-	-	-
		Other	No and Missing	180	93.26	155.83	32.63	0	250	55.00	13.64	0	72	0.92	3.78
			Yes	13	6.74	-	-	-	-	-	-	-	-	-	-

Table 22. 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 and Missing	531	90.31	149.33	2.02	0	223	51.20	15.24	0	71	0.93	3.95
			Yes	57	9.69	130.47	4.56	0	206	41.42	18.85	0	70	0.94	4.59
		Braille	No and Missing	587	99.83	147.48	33.88	0	223	50.24	15.88	0	71	0.94	4.01
			Yes	1	0.17	-	-	-	-	-	-	-	-	-	-
		Eye Gaze	No and Missing	564	95.92	149.04	32.69	0	223	51.15	15.19	0	71	0.93	3.95
			Yes	24	4.08	111.46	41.06	0	161	29.04	17.08	0	59	0.92	4.94
		Modified Picture Symbols	No and Missing	558	94.90	149.40	31.15	0	223	51.18	14.86	0	71	0.93	4.01
			Yes	30	5.10	112.27	56.72	0	196	33.03	23.13	0	69	0.97	3.94
		Objects	No and Missing	562	95.58	149.08	31.69	0	223	50.95	15.23	0	71	0.93	4.01
			Yes	26	4.42	113.35	55.81	0	165	35.12	21.55	0	61	0.96	4.10
		Sign Language	No and Missing	577	98.13	147.71	34.00	0	223	50.40	15.85	0	71	0.94	4.00
			Yes	11	1.87	-	-	-	-	-	-	-	-	-	-
		Translation into Native Language	No and Missing	585	99.49	147.50	33.94	0	223	50.25	15.91	0	71	0.94	4.01
			Yes	3	0.51	-	-	-	-	-	-	-	-	-	-
		Other	No and Missing	520	88.44	148.57	31.84	0	223	50.61	15.25	0	71	0.93	4.02
			Yes	68	11.56	139.37	46.01	0	189	47.54	19.95	0	68	0.96	3.96
	8	Assistive Technology	No and Missing	580	94.16	149.68	32.18	0	250	83.02	21.21	0	108	0.96	4.44
			Yes	36	5.84	128.89	28.93	0	169	65.86	21.29	0	97	0.92	6.17
		Braille	No and Missing	613	99.51	148.33	32.36	0	250	81.94	21.59	0	108	0.96	4.58
			Yes	3	0.49	-	-	-	-	-	-	-	-	-	-
		Eye Gaze	No and Missing	600	97.40	149.43	31.45	0	250	82.81	20.71	0	108	0.95	4.53
			Yes	16	2.60	112.44	44.52	0	169	52.31	31.47	0	97	0.97	5.80
		Modified Picture Symbols	No and Missing	590	95.78	149.25	32.37	0	250	82.74	21.18	0	108	0.96	4.51
			Yes	26	4.22	130.69	26.77	59	169	65.81	24.46	10	97	0.95	5.71
		Objects	No and Missing	600	97.40	148.88	32.14	0	250	82.38	21.14	0	108	0.95	4.57
			Yes	16	2.60	133.00	37.10	59	174	68.69	32.24	10	99	0.98	4.74
		Sign Language	No and Missing	605	98.21	148.64	32.39	0	250	82.15	21.52	0	108	0.96	4.55
			Yes	11	1.79	-	-	-	-	-	-	-	-	-	-
		Translation into Native Language	No and Missing	611	99.19	148.45	32.46	0	250	82.00	21.64	0	108	0.96	4.56
			Yes	5	0.81	-	-	-	-	-	-	-	-	-	-
		Other	No and Missing	555	90.10	149.00	30.89	0	231	82.59	20.44	0	107	0.95	4.57
			Yes	61	9.90	143.57	43.42	0	250	76.82	29.68	0	108	0.98	4.61

Table 23. Science Descriptive Statistics by Accommodation (continued)

Content	Grade	Accommodation	Subgroup	N	%	Scale Score				Raw Score				Alpha	SEM
						Mean	SD	Min	Max	Mean	SD	Min	Max		
SC	HS	Assistive Technology	No and Missing	473	92.38	154.67	33.47	0	250	85.90	22.68	0	110	0.97	4.19
			Yes	39	7.62	121.85	48.67	0	184	60.10	32.39	0	104	0.97	5.44
		Braille	No and Missing	505	98.63	152.74	34.99	0	250	84.28	24.07	0	110	0.97	4.29
			Yes	7	1.37	-	-	-	-	-	-	-	-	-	-
		Eye Gaze	No and Missing	493	96.29	155.53	29.06	0	250	85.93	21.37	0	110	0.96	4.30
			Yes	19	3.71	64.95	25.00	0	167	32.26	40.30	0	97	0.99	3.68
		Modified Picture Symbols	No and Missing	497	97.07	153.15	35.00	0	250	84.74	23.77	0	110	0.97	4.26
			Yes	15	2.93	-	-	-	-	-	-	-	-	-	-
		Objects	No and Missing	493	96.29	153.62	33.84	0	250	84.91	23.41	0	110	0.97	4.26
			Yes	19	3.71	114.68	60.77	0	172	58.68	36.91	0	100	0.98	5.13
		Sign Language	No and Missing	499	97.46	152.82	35.13	0	250	84.36	24.08	0	110	0.97	4.29
			Yes	13	2.54	-	-	-	-	-	-	-	-	-	-
		Translation into Native Language	No and Missing	505	98.63	152.68	35.00	0	250	84.22	24.09	0	110	0.97	4.30
			Yes	7	1.37	-	-	-	-	-	-	-	-	-	-
		Other	No and Missing	453	88.48	154.10	33.29	0	250	85.38	22.68	0	110	0.96	4.28
			Yes	59	11.52	137.37	49.49	0	199	72.88	33.76	0	107	0.98	4.37

Table 24. Grade 4 Social Studies Classical Statistics

ITEM	TYPE	% 0	% 1	% 2	% 3	% 4	% 5	% 6	MEAN SCORE	ITEM-TOTAL CORR
1	SR	7.4	23.2	29.5	19.5	20.5			2.226	0.573
2	SR	7.9	17.4	18.4	18.4	37.9			2.611	0.657
3	SR	6.8	22.1	12.6	22.6	35.8			2.584	0.618
4	SR	7.4	13.2	17.4	10.5	51.6			2.858	0.711
5	SR	6.8	13.2	23.7	23.7	32.6			2.621	0.647
6	SR	7.4	13.2	20.5	20.5	38.4			2.695	0.630
7	SR	6.3	10.5	27.9	18.4	36.8			2.689	0.608
8	SR	7.4	17.4	8.4	20.0	46.8			2.816	0.728
9	SPT	6.3	2.1	1.6	5.8	20.0	41.1	23.2	4.468	0.725
10	SR	7.4	21.1	14.7	17.9	38.9			2.600	0.702
11	SR	8.4	8.9	14.7	19.5	48.4			2.905	0.712
12	SR	9.5	20.5	22.1	11.1	36.8			2.453	0.681
13	SPT	6.8	3.2	1.6	25.8	32.6	21.1	8.9	3.732	0.668
14	SR	7.4	15.8	11.6	15.8	49.5			2.842	0.775
15	SR	7.4	13.7	17.4	27.4	34.2			2.674	0.650
16	SR	6.8	7.9	3.7	9.5	72.1			3.321	0.784
17	SR	6.8	14.2	20.0	21.6	37.4			2.684	0.665

Table 25. Grade 7 Social Studies Classical Statistics

ITEM	TYPE	% 0	% 1	% 2	% 3	% 4	% 5	% 6	MEAN SCORE	ITEM-TOTAL CORR
1	SR	2.1	14.5	15.5	20.7	47.2			2.964	0.611
2	SR	4.1	18.1	22.3	19.7	35.8			2.648	0.621
3	SR	3.6	7.3	16.1	36.3	36.8			2.953	0.576
4	SPT	3.1	0.5	1.6	17.1	26.4	21.8	29.5	4.466	0.683
5	SR	2.6	10.9	7.8	14.5	64.2			3.269	0.689
6	SR	3.6	5.7	16.6	19.7	54.4			3.155	0.648
7	SR	4.1	12.4	19.7	26.9	36.8			2.798	0.681
8	SR	2.1	11.4	9.3	14.0	63.2			3.249	0.688
9	SR	3.6	6.2	12.4	20.2	57.5			3.218	0.608
10	SPT	3.1	0.5	1.0	4.7	10.4	25.9	54.4	5.140	0.772
11	SR	2.6	8.3	11.4	24.9	52.8			3.171	0.533
12	SR	4.1	12.4	7.8	27.5	48.2			3.031	0.716
13	SR	3.6	11.9	11.9	20.7	51.8			3.052	0.761
14	SR	3.6	13.5	19.2	15.5	48.2			2.912	0.601
15	SR	4.7	5.7	6.7	17.6	65.3			3.332	0.662
16	SR	4.1	10.4	22.8	28.0	34.7			2.788	0.686
17	SR	3.1	8.3	21.2	20.2	47.2			3.000	0.698

Table 26. Grade 5 Science Classical Statistics

ITEM	TYPE	% 0	% 1	% 2	% 3	% 4	% 5	% 6	MEAN SCORE	ITEM-TOTAL CORR
1	SR	5.4	10.2	11.9	16.0	56.5			3.078	0.651
2	SR	5.4	8.2	8.0	8.5	69.9			3.293	0.751
3	SR	7.0	16.0	32.3	23.8	20.9			2.357	0.586
4	SPT	5.1	2.4	2.6	16.7	32.5	25.3	15.5	4.070	0.741
5	SR	8.2	16.3	17.5	16.2	41.8			2.672	0.743
6	SR	7.0	5.8	9.9	9.0	68.4			3.260	0.722
7	SR	8.3	16.8	23.5	19.0	32.3			2.502	0.672
8	SR	7.7	18.7	21.8	17.5	34.4			2.522	0.672
9	SR	7.8	9.9	9.9	6.8	65.6			3.126	0.827
10	SR	7.3	10.9	11.9	16.7	53.2			2.976	0.733
11	SR	8.8	17.5	25.7	23.3	24.7			2.374	0.647
12	SR	6.3	17.5	16.2	15.0	45.1			2.750	0.753
13	SR	7.8	13.9	17.2	19.9	41.2			2.726	0.715
14	SPT	7.0	1.7	1.7	15.5	27.2	24.3	22.6	4.177	0.801
15	SR	6.3	7.7	21.1	12.4	52.6			2.973	0.592
16	SR	6.8	14.8	18.5	26.0	33.8			2.653	0.728
17	SR	7.8	11.2	21.9	16.8	42.2			2.743	0.603

Table 27. Grade 8 Science Classical Statistics

ITEM	TYPE	% 0	% 1	% 2	% 3	% 4	% 5	% 6	MEAN SCORE	ITEM-TOTAL CORR
1	SR	5.7	9.6	18.3	23.2	43.2			2.886	0.626
2	SR	5.8	9.6	13.0	15.4	56.2			3.065	0.745
3	SR	5.0	4.4	8.4	12.2	70.0			3.377	0.733
4	SR	5.5	11.4	15.7	25.0	42.4			2.873	0.682
5	SPT	4.1	0.8	1.1	24.5	22.4	16.4	30.7	4.323	0.672
6	SR	5.2	5.8	8.0	13.3	67.7			3.325	0.763
7	SR	6.0	3.4	4.7	7.6	78.2			3.487	0.780
8	SR	4.9	5.8	14.4	33.3	41.6			3.008	0.592
9	SR	5.7	14.1	21.4	20.3	38.5			2.718	0.661
10	SR	5.4	10.4	21.3	20.8	42.2			2.841	0.604
11	SR	5.4	8.8	8.8	10.1	67.0			3.247	0.776
12	SR	5.5	10.1	22.2	19.0	43.2			2.843	0.675
13	SR	5.5	8.0	8.1	11.0	67.4			3.268	0.778
14	SR	4.7	7.0	23.1	21.3	44.0			2.929	0.617
15	SR	4.7	5.0	6.5	12.2	71.6			3.409	0.776
16	SR	5.2	10.4	23.4	33.3	27.8			2.680	0.606
17	SR	4.1	9.3	15.6	23.9	47.2			3.010	0.670
18	SR	5.0	11.4	20.0	24.2	39.4			2.817	0.637
19	SR	4.7	8.0	7.1	11.5	68.7			3.315	0.750
20	SR	4.9	3.9	8.4	16.9	65.9			3.351	0.705
21	SR	5.5	12.0	22.2	26.5	33.8			2.709	0.608
22	SPT	4.4	1.1	0.5	28.7	29.5	24.4	11.4	3.964	0.586
23	SR	4.7	7.5	9.1	14.6	64.1			3.260	0.761
24	SR	5.5	7.8	14.9	23.1	48.7			3.016	0.667
25	SR	4.4	8.9	13.6	28.6	44.5			2.998	0.666
26	SR	4.9	8.0	9.4	7.5	70.3			3.304	0.775

Table 28. HS Science Classical Statistics

ITEM	TYPE	% 0	% 1	% 2	% 3	% 4	% 5	% 6	MEAN SCORE	ITEM-TOTAL CORR
1	SR	5.1	4.3	4.3	8.2	78.1			3.500	0.799
2	SR	5.7	10.9	27.7	21.9	33.8			2.672	0.640
3	SR	6.4	7.6	10.4	21.1	54.5			3.096	0.755
4	SPT	4.9	1.4	1.2	9.2	19.5	26.6	37.3	4.660	0.734
5	SR	5.1	12.3	27.0	21.7	34.0			2.672	0.692
6	SR	5.7	10.7	18.2	25.0	40.4			2.838	0.730
7	SR	6.1	8.8	15.8	23.0	46.3			2.947	0.737
8	SR	7.0	10.9	17.4	26.6	38.1			2.777	0.744
9	SR	5.5	10.0	11.1	25.8	47.7			3.002	0.724
10	SR	4.9	9.4	12.3	17.6	55.9			3.102	0.739
11	SPT	4.3	1.6	0.6	11.7	24.4	34.0	23.4	4.461	0.670
12	SR	5.5	9.6	14.1	20.1	50.8			3.012	0.742
13	SR	6.1	8.0	14.5	25.0	46.5			2.979	0.778
14	SR	6.4	6.4	9.8	13.3	64.1			3.221	0.833
15	SR	6.3	8.2	9.4	14.6	61.5			3.170	0.809
16	SR	5.9	12.3	18.9	38.7	24.2			2.631	0.723
17	SR	6.3	7.2	11.5	19.7	55.3			3.105	0.813
18	SR	6.3	8.8	9.0	22.3	53.7			3.084	0.809
19	SR	6.3	8.2	15.8	18.2	51.6			3.006	0.803
20	SR	5.7	12.1	12.5	21.5	48.2			2.945	0.735
21	SPT	4.5	0.4	1.4	4.9	8.0	17.0	63.9	5.180	0.813
22	SR	5.3	6.4	17.8	20.9	49.6			3.031	0.689
23	SR	5.7	8.8	7.6	17.4	60.5			3.184	0.797
24	SR	5.9	6.8	7.4	11.9	68.0			3.293	0.822
25	SR	4.9	7.6	7.0	6.3	74.2			3.373	0.834
26	SR	5.5	9.0	17.2	17.2	51.2			2.996	0.650

Table 29. 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			1.05	1.06
2	SR	-0.0896	0	-1.5816	0.2833	0.8469	0.4514			1.00	0.95
3	SR	-0.0745	0	-1.7552	0.3614	0.6682	0.7256			0.77	0.75
4	SR	-0.2883	0	-1.5743	0.3035	1.3227	-0.0519			1.05	1.07
5	SR	-0.1462	0	-1.8403	-0.131	0.8772	1.0941			1.05	1.01
6	SR	-0.1537	0	-1.4792	-0.1084	1.1001	0.4874			1.19	1.16
7	SR	0.179	0	-1.0204	-0.1712	1.3677	-0.1762			1.24	1.26
8	SR	-0.3071	0	-1.8956	0.7291	0.7778	0.3887			0.99	0.97
9	SPT	-0.3582	0	-0.9166	-1.0555	-1.3048	-0.0936	0.8590	2.5115	1.37	1.24
10	SR	-0.1218	0	-1.6751	0.8345	0.3964	0.4442			1.14	1.07
11	SR	-0.057	0	-1.0456	-0.0434	0.7401	0.349			1.08	1.27
12	SR	-0.0156	0	-1.7108	0.2341	1.5167	-0.0399			0.96	0.90
13	SPT	0.3472	0	-0.1932	-0.2591	-2.5586	-0.0418	1.2165	1.8363	1.16	1.37
14	SR	-0.1037	0	-1.142	0.2453	0.7699	0.1268			1.06	0.95
15	SR	0.0543	0	-1.5601	-0.1821	0.8104	0.9318			1.20	1.19
16	SR	-0.3513	0	-0.1788	0.0459	0.8822	-0.7493			0.80	0.64
17	SR	0.0549	0	-1.3119	-0.1083	0.4443	0.9758			1.14	1.05

Table 30. Grade 7 Social Studies Item Parameter Estimates

ITEM	TYPE	B	D1	D2	D3	D4	D5	D6	D7	INFIT	OUTFIT
1	SR	0.0297	0	-1.6361	0.1561	1.0223	0.4577			0.83	0.83
2	SR	0.4368	0	-1.6058	-0.278	0.9952	0.8886			1.06	1.08
3	SR	-0.126	0	-2.2184	-0.1154	0.8068	1.527			1.35	1.44
4	SPT	0.5472	0	0.1997	-0.2972	-2.3954	0.3878	0.9377	1.1675	0.91	0.90
5	SR	-0.1839	0	-1.5577	0.1067	1.2653	0.1858			1.24	1.11
6	SR	-0.3251	0	-1.5864	-0.0575	1.002	0.6418			1.13	1.16
7	SR	0.4928	0	-1.1893	-0.2117	0.7468	0.6542			0.82	0.78
8	SR	0.0097	0	-1.2601	-0.2543	1.1077	0.4066			0.92	0.96
9	SR	-0.1161	0	-1.2352	-0.216	1.3941	0.0571			1.00	1.04
10	SPT	-0.1555	0	0.7307	-1.1437	-1.0072	0.022	0.4551	0.9431	0.90	0.86
11	SR	0.0657	0	-0.9411	-0.5618	0.9948	0.5081			1.11	0.98
12	SR	0.1286	0	-1.0107	-0.4662	0.7297	0.7472			1.06	1.08
13	SR	0.0491	0	-0.9163	-0.3126	1.2027	0.0262			0.96	1.06
14	SR	0.1696	0	-1.283	-0.6076	1.4417	0.4489			1.08	0.97
15	SR	-0.2313	0	-1.1255	-0.2407	0.7713	0.5948			0.99	0.79
16	SR	0.3	0	-1.3652	-0.2564	0.729	0.8927			1.13	1.04
17	SR	-0.1001	0	-1.2131	-0.1457	0.6764	0.6824			1.16	0.99

Table 31. Grade 5 Science Item Parameter Estimates

ITEM	TYPE	B	D1	D2	D3	D4	D5	D6	D7	INFIT	OUTFIT
1	SR	-0.2143	0	-1.5546	0.5073	0.8214	0.2259			1.21	1.23
2	SR	-0.3967	0	-1.1400	0.7813	1.0200	-0.6612			1.11	0.88
3	SR	0.5586	0	-1.5202	-0.2995	0.8327	0.9869			1.09	1.09
4	SPT	0.3165	0	-0.2580	-1.0519	-1.7798	-0.1750	1.2326	2.0319	1.15	1.31
5	SR	0.2788	0	-1.2349	0.3587	0.6595	0.2167			0.85	0.77
6	SR	-0.2055	0	-0.5300	0.0376	1.1234	-0.6309			1.09	1.04
7	SR	0.4173	0	-1.4518	0.0101	0.9372	0.5044			0.99	0.97
8	SR	0.3605	0	-1.6544	0.2340	1.0062	0.4142			1.00	1.09
9	SR	-0.0277	0	-0.8373	0.5352	1.3066	-1.0046			0.68	0.51
10	SR	-0.0290	0	-1.0785	-0.1014	1.0991	0.0807			1.01	0.89
11	SR	0.6061	0	-1.5585	-0.3207	0.8678	1.0114			1.12	1.10
12	SR	0.1046	0	-1.8181	0.6378	1.0246	0.1557			0.80	0.75
13	SR	0.2412	0	-1.3002	0.2227	0.6611	0.4165			0.93	0.87
14	SPT	0.2842	0	-0.1153	-0.2753	-1.9832	0.0295	1.0300	1.3142	0.91	0.99
15	SR	-0.0677	0	-0.9432	-0.2691	1.2902	-0.0779			1.37	1.40
16	SR	0.2340	0	-1.6347	0.2165	0.5012	0.9170			0.87	0.89
17	SR	0.2870	0	-0.8303	-0.4614	1.3720	-0.0803			1.24	1.34

Table 32. Grade 8 Science Item Parameter Estimates

ITEM	TYPE	B	D1	D2	D3	D4	D5	D6	D7	INFIT	OUTFIT
1	SR	0.1529	0	-1.5952	-0.1173	0.8843	0.8282			1.20	1.25
2	SR	-0.0474	0	-1.7711	0.3571	1.1451	0.2689			0.97	0.84
3	SR	-0.1818	0	-0.8881	0.1008	0.8168	-0.0295			1.05	0.97
4	SR	0.2132	0	-1.7709	0.2698	0.7363	0.7648			0.99	0.99
5	SPT	0.1292	0	-0.5492	-0.6137	-2.5213	1.1065	1.6160	0.9617	1.13	1.18
6	SR	-0.3475	0	-1.6923	0.3380	1.0049	0.3494			1.01	0.81
7	SR	-0.0975	0	-0.1245	0.3202	0.626	-0.8216			0.94	0.83
8	SR	0.1216	0	-1.1964	-0.2787	0.2693	1.2057			1.26	1.36
9	SR	0.4250	0	-1.6801	0.1421	0.9772	0.5609			0.95	0.95
10	SR	0.3076	0	-1.4943	-0.1204	1.0207	0.5940			1.17	1.12
11	SR	-0.0618	0	-1.2927	0.3729	1.3843	-0.4645			0.89	0.70
12	SR	0.3179	0	-1.4048	-0.1970	1.1360	0.4658			0.97	1.00
13	SR	0.0458	0	-1.0708	0.6637	0.7989	-0.3918			0.83	0.66
14	SR	0.1420	0	-1.3917	-0.5194	1.1967	0.7143			1.15	1.14
15	SR	-0.2042	0	-0.9779	0.4906	0.6390	-0.1517			0.91	0.71
16	SR	0.4278	0	-1.6729	-0.2961	0.5676	1.4014			1.14	1.15
17	SR	0.0002	0	-1.9335	0.2577	0.8115	0.8643			1.01	0.97
18	SR	0.3207	0	-1.5338	-0.2799	0.7203	1.0935			1.12	1.11
19	SR	-0.0808	0	-0.9915	0.3719	0.5418	0.0778			1.02	0.87
20	SR	-0.1574	0	-0.7099	-0.0673	0.5461	0.2310			1.09	1.07
21	SR	0.4242	0	-1.6286	-0.0777	0.7529	0.9535			1.09	1.11
22	SPT	0.4586	0	-0.8730	0.4121	-3.7385	0.7442	1.2908	2.1645	1.38	1.39
23	SR	-0.2471	0	-1.4268	0.1288	0.9212	0.3767			1.04	0.80
24	SR	0.1654	0	-0.8388	-0.4949	0.8324	0.5013			1.11	1.07
25	SR	0.1560	0	-1.2389	-0.242	0.1589	1.3220			1.17	1.07
26	SR	-0.0515	0	-1.5727	0.7610	1.570	-0.7584			0.78	0.62

Table 33. HS Science Item Parameter Estimates

ITEM	TYPE	B	D1	D2	D3	D4	D5	D6	D7	INFIT	OUTFIT
1	SR	-0.7904	0	-1.8221	0.3446	1.6074	-0.1299			1.06	0.82
2	SR	0.3213	0	-1.6946	-0.7476	1.3998	1.0425			1.19	1.16
3	SR	0.0778	0	-1.2721	0.2797	0.4113	0.5810			1.02	0.91
4	SPT	-0.1258	0	-1.2440	-0.1547	-1.5832	0.2891	1.1881	1.5048	1.32	1.20
5	SR	0.2136	0	-1.9328	-0.5023	1.3974	1.0377			1.03	1.09
6	SR	0.1839	0	-1.9743	0.0984	0.8286	1.0473			0.98	0.96
7	SR	0.1451	0	-1.5979	0.0141	0.7573	0.8265			0.99	0.98
8	SR	0.2739	0	-1.9122	0.0042	0.8958	1.0121			0.93	0.88
9	SR	0.0018	0	-1.6694	-0.0917	0.9192	0.8418			1.08	1.07
10	SR	-0.1691	0	-2.1532	0.5199	1.0258	0.6076			1.04	1.14
11	SPT	-0.1656	0	-2.0478	0.4884	-2.4899	0.4181	1.2851	2.3461	1.51	1.55
12	SR	0.0134	0	-1.8899	0.3058	0.8790	0.7051			1.00	1.06
13	SR	0.1162	0	-1.5156	0.0063	0.6010	0.9083			0.87	0.82
14	SR	-0.2761	0	-1.1544	0.0227	0.9848	0.1469			1.01	0.77
15	SR	0.0976	0	-0.8830	-0.1765	0.8767	0.1827			0.83	0.80
16	SR	0.3105	0	-2.0915	-0.3357	0.6357	1.7916			0.99	0.95
17	SR	-0.1199	0	-1.3246	-0.2161	0.5995	0.9412			0.95	0.81
18	SR	0.0565	0	-1.0219	-0.1391	0.6189	0.5421			0.86	0.74
19	SR	0.1186	0	-1.4258	-0.049	1.0076	0.4672			0.78	0.71
20	SR	0.1144	0	-2.0098	0.6314	0.6339	0.7444			0.99	1.08
21	SPT	-0.1877	0	0.8594	-0.8307	-1.4616	0.4506	0.5493	0.4329	1.24	2.28
22	SR	-0.0764	0	-1.6770	-0.3111	1.1499	0.8382			1.16	1.09
23	SR	0.1579	0	-0.7580	-0.1354	0.9368	-0.0434			0.81	0.81
24	SR	0.0783	0	-0.4765	0.1000	1.1304	-0.7540			0.77	0.79
25	SR	-0.4007	0	-1.8729	0.9034	1.5548	-0.5852			0.83	0.58
26	SR	0.0155	0	-1.8372	0.0482	1.2447	0.5443			1.31	1.22

Table 34. Cut Scores and Students in Each Performance Level

Content	Grade	Cut Scores			Performance Levels									
		Approaching Target	At Target	Advanced	Emerging		Approaching Target		At Target		Advanced		At Target and Advanced Combined	
					N	%	N	%	N	%	N	%	N	%
SS	4	46	59	66	53	28	85	45	45	24	7	4	52	27
	7	44	60	68	26	14	83	43	67	35	17	9	84	44
SC	5	41	58	67	105	18	241	41	205	35	37	6	242	41
	8	63	94	103	67	11	358	58	175	28	16	3	191	31
	HS	72	95	106	98	19	188	37	205	40	21	4	226	44

Table 35. Scale Score Ranges for Each Performance Level

	Emerging Level	Approaching Target Level	At Target Level	Advanced 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
HS Science	0–139	140–163	164–192	193–250

Table 36. Grade 4 Social Studies Scale Score Frequency Distributions

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
000	2	1.05	2	1.05
001	1	0.53	3	1.58
026	1	0.53	4	2.11
038	1	0.53	5	2.63
055	1	0.53	6	3.16
061	1	0.53	7	3.68
066	1	0.53	8	4.21
070	1	0.53	9	4.74
081	1	0.53	10	5.26
084	1	0.53	11	5.79
087	1	0.53	12	6.32
090	1	0.53	13	6.84
092	1	0.53	14	7.37
097	1	0.53	15	7.89
107	1	0.53	16	8.42
114	1	0.53	17	8.95
117	3	1.58	20	10.53
118	3	1.58	23	12.11
120	1	0.53	24	12.63
121	2	1.05	26	13.68
124	2	1.05	28	14.74
127	2	1.05	30	15.79
129	1	0.53	31	16.32
130	2	1.05	33	17.37
131	1	0.53	34	17.89
133	1	0.53	35	18.42
134	3	1.58	38	20.00
135	4	2.11	42	22.11

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
137	1	0.53	43	22.63
138	3	1.58	46	24.21
139	2	1.05	48	25.26
141	5	2.63	53	27.89
143	2	1.05	55	28.95
144	8	4.21	63	33.16
145	3	1.58	66	34.74
146	4	2.11	70	36.84
148	6	3.16	76	40.00
149	7	3.68	83	43.68
151	7	3.68	90	47.37
152	6	3.16	96	50.53
154	7	3.68	103	54.21
156	6	3.16	109	57.37
158	7	3.68	116	61.05
159	12	6.32	128	67.37
161	10	5.26	138	72.63
163	9	4.74	147	77.37
166	7	3.68	154	81.05
168	7	3.68	161	84.74
171	8	4.21	169	88.95
174	4	2.11	173	91.05
177	6	3.16	179	94.21
180	4	2.11	183	96.32
188	3	1.58	186	97.89
189	1	0.53	187	98.42
195	2	1.05	189	99.47
214	1	0.53	190	100.00

Table 37. Grade 7 Social Studies Scale Score Frequency Distributions

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
000	2	1.04	2	1.04
042	2	1.04	4	2.07
061	1	0.52	5	2.59
087	1	0.52	6	3.11
093	1	0.52	7	3.63
107	1	0.52	8	4.15
110	1	0.52	9	4.66
116	3	1.55	12	6.22
119	1	0.52	13	6.74
121	1	0.52	14	7.25
123	1	0.52	15	7.77
125	1	0.52	16	8.29
126	3	1.55	19	9.84
128	1	0.52	20	10.36
129	1	0.52	21	10.88
131	2	1.04	23	11.92
132	3	1.55	26	13.47
134	5	2.59	31	16.06
135	4	2.07	35	18.13
136	3	1.55	38	19.69
138	1	0.52	39	20.21
139	3	1.55	42	21.76
141	1	0.52	43	22.28
143	2	1.04	45	23.32
144	2	1.04	47	24.35
146	7	3.63	54	27.98
147	5	2.59	59	30.57
149	6	3.11	65	33.68
151	6	3.11	71	36.79

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
153	14	7.25	85	44.04
155	7	3.63	92	47.67
157	9	4.66	101	52.33
159	8	4.15	109	56.48
163	10	5.18	119	61.66
164	3	1.55	122	63.21
166	11	5.70	133	68.91
169	7	3.63	140	72.54
172	8	4.15	148	76.68
176	8	4.15	156	80.83
180	10	5.18	166	86.01
184	10	5.18	176	91.19
191	4	2.07	180	93.26
198	6	3.11	186	96.37
208	2	1.04	188	97.41
227	4	2.07	192	99.48
250	1	0.52	193	100.00

Table 38. Grade 5 Science Scale Score Frequency Distributions

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
000	12	2.04	12	2.04
001	4	0.68	16	2.72
041	5	0.85	21	3.57
061	2	0.34	23	3.91
077	1	0.17	24	4.08
081	2	0.34	26	4.42
088	3	0.51	29	4.93
091	1	0.17	30	5.10
093	2	0.34	32	5.44
098	1	0.17	33	5.61
100	1	0.17	34	5.78
101	3	0.51	37	6.29
105	2	0.34	39	6.63
107	1	0.17	40	6.80
108	2	0.34	42	7.14
110	3	0.51	45	7.65
113	1	0.17	46	7.82
114	4	0.68	50	8.50
116	2	0.34	52	8.84
117	3	0.51	55	9.35
118	2	0.34	57	9.69
119	2	0.34	59	10.03
121	1	0.17	60	10.20
122	5	0.85	65	11.05
123	4	0.68	69	11.73
125	1	0.17	70	11.90
126	4	0.68	74	12.59
127	5	0.85	79	13.44
128	5	0.85	84	14.29

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
129	3	0.51	87	14.80
131	6	1.02	93	15.82
132	3	0.51	96	16.33
133	5	0.85	101	17.18
134	4	0.68	105	17.86
135	8	1.36	113	19.22
136	8	1.36	121	20.58
138	11	1.87	132	22.45
139	6	1.02	138	23.47
140	14	2.38	152	25.85
141	9	1.53	161	27.38
143	11	1.87	172	29.25
144	14	2.38	186	31.63
145	22	3.74	208	35.37
146	17	2.89	225	38.27
148	15	2.55	240	40.82
149	14	2.38	254	43.20
151	18	3.06	272	46.26
152	14	2.38	286	48.64
154	22	3.74	308	52.38
155	21	3.57	329	55.95
157	17	2.89	346	58.84
160	26	4.42	372	63.27
161	33	5.61	405	68.88
163	29	4.93	434	73.81
165	23	3.91	457	77.72
167	25	4.25	482	81.97
170	15	2.55	497	84.52
173	22	3.74	519	88.27
176	15	2.55	534	90.82
179	17	2.89	551	93.71

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
184	14	2.38	565	96.09
189	13	2.21	578	98.30
196	6	1.02	584	99.32
206	3	0.51	587	99.83
223	1	0.17	588	100.00

Table 39. Grade 8 Science Scale Score Frequency Distributions

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
000	14	2.27	14	2.27
010	1	0.16	15	2.44
031	1	0.16	16	2.60
038	1	0.16	17	2.76
052	1	0.16	18	2.92
059	1	0.16	19	3.08
068	1	0.16	20	3.25
070	1	0.16	21	3.41
074	1	0.16	22	3.57
078	2	0.32	24	3.90
080	2	0.32	26	4.22
085	2	0.32	28	4.55
091	2	0.32	30	4.87
092	1	0.16	31	5.03
095	2	0.32	33	5.36
097	1	0.16	34	5.52
099	1	0.16	35	5.68
102	1	0.16	36	5.84
105	2	0.32	38	6.17

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
109	2	0.32	40	6.49
110	1	0.16	41	6.66
111	2	0.32	43	6.98
113	1	0.16	44	7.14
114	1	0.16	45	7.31
115	2	0.32	47	7.63
117	2	0.32	49	7.95
118	3	0.49	52	8.44
120	1	0.16	53	8.60
122	3	0.49	56	9.09
123	2	0.32	58	9.42
124	2	0.32	60	9.74
125	2	0.32	62	10.06
126	3	0.49	65	10.55
127	2	0.32	67	10.88
128	9	1.46	76	12.34
129	4	0.65	80	12.99
130	4	0.65	84	13.64
131	5	0.81	89	14.45
132	8	1.30	97	15.75
133	7	1.14	104	16.88
134	6	0.97	110	17.86
136	9	1.46	119	19.32
137	17	2.76	136	22.08
138	3	0.49	139	22.56
139	5	0.81	144	23.38
140	5	0.81	149	24.19
141	11	1.79	160	25.97
142	8	1.30	168	27.27
144	8	1.30	176	28.57
145	13	2.11	189	30.68

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
146	8	1.30	197	31.98
147	15	2.44	212	34.42
148	17	2.76	229	37.18
149	21	3.41	250	40.58
151	15	2.44	265	43.02
152	14	2.27	279	45.29
153	26	4.22	305	49.51
155	27	4.38	332	53.90
156	20	3.25	352	57.14
158	25	4.06	377	61.20
159	28	4.55	405	65.75
161	20	3.25	425	68.99
164	25	4.06	450	73.05
165	26	4.22	476	77.27
167	23	3.73	499	81.01
169	26	4.22	525	85.23
171	15	2.44	540	87.66
174	22	3.57	562	91.23
177	15	2.44	577	93.67
180	13	2.11	590	95.78
184	10	1.62	600	97.40
190	4	0.65	604	98.05
195	5	0.81	609	98.86
202	3	0.49	612	99.35
213	2	0.32	614	99.68
231	1	0.16	615	99.84
250	1	0.16	616	100.00

Table 40. HS Science Scale Score Frequency Distributions

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
000	12	2.34	12	2.34
001	3	0.59	15	2.93
025	3	0.59	18	3.52
035	1	0.20	19	3.71
054	1	0.20	20	3.91
062	2	0.39	22	4.30
074	1	0.20	23	4.49
079	1	0.20	24	4.69
081	1	0.20	25	4.88
083	1	0.20	26	5.08
095	2	0.39	28	5.47
097	1	0.20	29	5.66
100	1	0.20	30	5.86
111	2	0.39	32	6.25
113	2	0.39	34	6.64
114	4	0.78	38	7.42
116	2	0.39	40	7.81
117	1	0.20	41	8.01
118	2	0.39	43	8.40
119	2	0.39	45	8.79
120	2	0.39	47	9.18
121	2	0.39	49	9.57
122	4	0.78	53	10.35
126	1	0.20	54	10.55
127	2	0.39	56	10.94
128	1	0.20	57	11.13
129	3	0.59	60	11.72
130	3	0.59	63	12.30
131	3	0.59	66	12.89

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
132	2	0.39	68	13.28
133	6	1.17	74	14.45
134	3	0.59	77	15.04
135	2	0.39	79	15.43
136	3	0.59	82	16.02
137	4	0.78	86	16.80
138	9	1.76	95	18.55
139	3	0.59	98	19.14
140	8	1.56	106	20.70
141	3	0.59	109	21.29
142	3	0.59	112	21.88
143	2	0.39	114	22.27
144	4	0.78	118	23.05
145	11	2.15	129	25.20
146	5	0.98	134	26.17
147	4	0.78	138	26.95
148	6	1.17	144	28.13
149	4	0.78	148	28.91
150	7	1.37	155	30.27
151	10	1.95	165	32.23
152	9	1.76	174	33.98
153	13	2.54	187	36.52
154	9	1.76	196	38.28
155	10	1.95	206	40.23
157	10	1.95	216	42.19
158	18	3.52	234	45.70
159	21	4.10	255	49.80
160	15	2.93	270	52.73
162	16	3.13	286	55.86
164	20	3.91	306	59.77
165	29	5.66	335	65.43

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
167	21	4.10	356	69.53
168	24	4.69	380	74.22
170	16	3.13	396	77.34
172	29	5.66	425	83.01
175	20	3.91	445	86.91
177	12	2.34	457	89.26
180	15	2.93	472	92.19
184	11	2.15	483	94.34
188	8	1.56	491	95.90
193	12	2.34	503	98.24
199	5	0.98	508	99.22
208	2	0.39	510	99.61
225	1	0.20	511	99.80
250	1	0.20	512	100.00

Table 41. Grade 4 Social Studies Scale Scores and Conditional Standard Error of Measurement (CSEM)

Raw Score	Scale Score	CSEM
0	0	52
1	1	29
2	26	21
3	38	17
4	47	15
5	55	14
6	61	13
7	66	12
8	70	11
9	74	10
10	78	10
11	81	10
12	84	9
13	87	9
14	90	9
15	92	8
16	95	8
17	97	8
18	99	8
19	101	8
20	103	7
21	105	7
22	107	7
23	109	7
24	110	7
25	112	7

26	114	7
27	115	7
28	117	7
29	118	7
30	120	6
31	121	6
32	123	6
33	124	6
34	126	6
35	127	6
36	129	6
37	130	6
38	131	6
39	133	6
40	134	6
41	135	6
42	137	6
43	138	6
44	139	6
45	141	6
46	143	6
47	144	6
48	145	6
49	146	6
50	148	6
51	149	7
52	151	7
53	152	7
54	154	7
55	156	7

56	158	7
57	159	7
58	161	8
59	163	8
60	166	8
61	168	8
62	171	9
63	174	9
64	177	10
65	180	10
66	188	11
67	189	12
68	195	14
69	203	16
70	214	20
71	234	28
72	250	52

Table 42. 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	20	21
3	33	18
4	42	15
5	50	14
6	56	13
7	61	12
8	65	11
9	69	11
10	73	10
11	76	10
12	79	9
13	82	9
14	85	8
15	87	8
16	89	8
17	91	8
18	93	8
19	95	7
20	97	7
21	99	7
22	101	7
23	102	7
24	104	7

25	106	7
26	107	7
27	109	7
28	110	7
29	112	7
30	113	7
31	115	7
32	116	7
33	118	7
34	119	6
35	121	6
36	122	6
37	123	6
38	125	6
39	126	6
40	128	6
41	129	6
42	131	6
43	132	7
44	134	7
45	135	7
46	136	7
47	138	7
48	139	7
49	141	7
50	143	7
51	144	7
52	146	7
53	147	7
54	149	7

55	151	7
56	153	7
57	155	8
58	157	8
59	159	8
60	163	8
61	164	9
62	166	9
63	169	9
64	172	10
65	176	10
66	180	11
67	184	12
68	191	14
69	198	16
70	208	20
71	227	28
72	250	53

Table 43. Grade 5 Science Scale Scores and Conditional Standard Error of Measurement (CSEM)

Raw Score	Scale Score	CSEM
0	0	48
1	1	27
2	41	19
3	52	16
4	61	14
5	67	12
6	73	11
7	77	11
8	81	10
9	85	9
10	88	9
11	91	8
12	93	8
13	95	8
14	98	7
15	100	7
16	101	7
17	103	7
18	105	7
19	107	6
20	108	6
21	110	6
22	111	6
23	113	6
24	114	6

25	116	6
26	117	6
27	118	6
28	119	6
29	121	6
30	122	6
31	123	6
32	125	6
33	126	6
34	127	6
35	128	6
36	129	6
37	131	6
38	132	6
39	133	6
40	134	6
41	135	6
42	136	6
43	138	6
44	139	6
45	140	6
46	141	6
47	143	6
48	144	6
49	145	6
50	146	6
51	148	6
52	149	6
53	151	6
54	152	6

55	154	6
56	155	7
57	157	7
58	160	7
59	161	7
60	163	7
61	165	8
62	167	8
63	170	8
64	173	9
65	176	9
66	179	10
67	184	11
68	189	12
69	196	14
70	206	18
71	223	26
72	250	47

Table 44. Grade 8 Science Scale Scores and Conditional Standard Error of Measurement (CSEM)

Raw Score	Scale Score	CSEM
0	0	50
1	1	28
2	10	20
3	22	16
4	31	14
5	38	13
6	43	12
7	48	11
8	52	10
9	56	10
10	59	9
11	62	9
12	65	9
13	68	8
14	70	8
15	72	8
16	74	7
17	76	7
18	78	7
19	80	7
20	82	7
21	83	7
22	85	7
23	86	6
24	88	6

25	89	6
26	91	6
27	92	6
28	93	6
29	95	6
30	96	6
31	97	6
32	98	6
33	99	6
34	100	6
35	102	6
36	103	5
37	104	5
38	105	5
39	106	5
40	107	5
41	108	5
42	109	5
43	110	5
44	111	5
45	112	5
46	113	5
47	114	5
48	115	5
49	115	5
50	116	5
51	117	5
52	118	5
53	119	5
54	120	5

55	121	5
56	122	5
57	122	5
58	123	5
59	124	5
60	125	5
61	126	5
62	127	5
63	128	5
64	128	5
65	129	5
66	130	5
67	131	5
68	132	5
69	133	5
70	134	5
71	135	5
72	136	5
73	137	5
74	137	5
75	138	5
76	139	5
77	140	5
78	141	5
79	142	5
80	144	5
81	145	5
82	146	6
83	147	6
84	148	6

85	149	6
86	151	6
87	152	6
88	153	6
89	155	6
90	156	6
91	158	7
92	159	7
93	161	7
94	164	7
95	165	7
96	167	8
97	169	8
98	171	8
99	174	9
100	177	9
101	180	10
102	184	11
103	190	12
104	195	13
105	202	15
106	213	19
107	231	27
108	250	50

Table 45. HS Science Scale Scores and Conditional Standard Error of Measurement (CSEM)

Raw Score	Scale Score	CSEM
0	0	44
1	1	24
2	25	17
3	35	14
4	43	13
5	49	11
6	54	11
7	58	10
8	62	9
9	66	9
10	69	8
11	72	8
12	74	8
13	77	7
14	79	7
15	81	7
16	83	7
17	85	7
18	87	6
19	88	6
20	90	6
21	92	6
22	93	6
23	95	6
24	96	6

25	97	6
26	99	5
27	100	5
28	101	5
29	102	5
30	103	5
31	104	5
32	105	5
33	107	5
34	108	5
35	109	5
36	110	5
37	111	5
38	112	5
39	113	5
40	114	5
41	114	5
42	115	5
43	116	5
44	117	5
45	118	5
46	119	5
47	120	5
48	121	5
49	122	4
50	122	4
51	123	4
52	124	4
53	125	4
54	126	4

55	126	4
56	127	4
57	128	4
58	129	4
59	130	4
60	130	4
61	131	4
62	132	4
63	133	4
64	133	4
65	134	4
66	135	4
67	136	4
68	137	4
69	137	4
70	138	4
71	139	4
72	140	4
73	140	4
74	141	4
75	142	4
76	143	4
77	144	5
78	145	5
79	145	5
80	146	5
81	147	5
82	148	5
83	149	5
84	150	5

85	151	5
86	152	5
87	153	5
88	154	5
89	155	5
90	157	5
91	158	5
92	159	6
93	160	6
94	162	6
95	164	6
96	165	6
97	167	6
98	168	7
99	170	7
100	172	7
101	175	8
102	177	8
103	180	9
104	184	9
105	188	10
106	193	12
107	199	13
108	208	17
109	225	24
110	250	43

Table 46. 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	0.63	0.35	0.43	0.37	0.68	0.29	0.03
	7	0.58	0.34	0.37	0.42	0.66	0.14	0.21
SC	5	0.62	0.34	0.42	0.38	0.69	0.12	0.19
	8	0.62	0.46	0.31	0.38	0.68	0.32	0.01
	HS	0.62	0.35	0.42	0.38	0.70	0.09	0.22

Table 47. Test Validity Questions Summary

Question	Subject	Grade	Very Familiar	Somewhat Familiar	Familiar	Somewhat Unfamiliar	Unfamiliar	Missing			
How familiar are you with this student?	SS	4	78.95%	4.21%	6.32%	8.42%	1.58%	0.53%			
		7	92.23%	5.18%	1.55%	0.52%	0.52%	0.00%			
	SC	5	35.20%	26.70%	19.22%	9.52%	8.50%	0.85%			
		8	87.34%	7.14%	4.87%	0.16%	0.32%	0.16%			
		HS	86.52%	7.42%	5.47%	0.39%	0.00%	0.20%			
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	Missing	
How many hours per week does this student spend in instruction on this content area?	SS	4	35.26%	26.32%	16.84%	8.95%	8.95%	0.53%	2.63%	0.53%	
		7	21.24%	13.99%	8.29%	19.69%	26.42%	9.84%	0.52%	0.00%	
	SC	5	22.62%	25.51%	19.22%	14.97%	8.84%	5.78%	3.06%	0.00%	
		8	14.94%	11.53%	9.42%	14.12%	34.90%	14.77%	0.16%	0.16%	
		HS	19.53%	14.45%	13.28%	21.29%	24.61%	6.45%	0.20%	0.20%	
Question		Grade	25%	50%	75%	100%	None	Missing			
Approximately how much instructional time for this content area is in the general education classroom?	SS	4	19.47%	9.47%	7.37%	28.42%	34.74%	0.53%			
		7	8.81%	3.63%	7.25%	37.31%	43.01%	0.00%			
	SC	5	17.01%	9.18%	12.07%	30.61%	31.12%	0.00%			
		8	6.33%	5.68%	12.82%	36.53%	38.31%	0.32%			
		HS	11.52%	4.49%	3.52%	17.19%	63.09%	0.20%			
Question		Grade	Oral Language	Reading	Picture Communication	Tactile	Other	Do Not Know	Missing		
This student's primary receptive communication is:	SS	4	91.58%	2.11%	3.16%	0.53%	1.05%	1.05%	0.53%		
		7	96.89%	1.55%	0.52%	0.00%	1.04%	0.00%	0.00%		
	SC	5	89.97%	1.87%	4.25%	0.51%	2.72%	0.68%	0.00%		
		8	91.56%	2.60%	3.41%	0.32%	1.95%	0.00%	0.16%		
		HS	92.19%	1.95%	4.10%	0.00%	1.56%	0.00%	0.20%		
Question		Grade	Oral Language	Writing	Picture Communication	Tactile	Other	Do Not Know	Missing		
This student's primary expressive communication is:	SS	4	75.79%	0.00%	14.21%	1.05%	6.84%	1.58%	0.53%		
		7	84.46%	1.55%	9.33%	0.00%	4.15%	0.52%	0.00%		
	SC	5	81.80%	0.17%	9.01%	0.68%	7.82%	0.51%	0.00%		

		8	85.06%	0.81%	7.63%	0.16%	6.17%	0.00%	0.16%		
		HS	83.20%	0.78%	10.16%	0.00%	5.47%	0.20%	0.20%		
Question		Grade	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree	Do Not Know	Missing		
I feel that the student's responses accurately reflect their understanding of the material.	SS	4	31.58%	40.53%	13.16%	7.89%	4.21%	2.11%	0.53%		
		7	33.16%	46.63%	11.92%	5.70%	1.04%	1.55%	0.00%		
	SC	5	25.85%	47.62%	14.46%	7.48%	3.06%	1.53%	0.00%		
		8	34.74%	41.40%	15.58%	5.52%	1.62%	0.97%	0.16%		
		HS	38.67%	40.43%	11.52%	6.05%	2.54%	0.59%	0.20%		
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?	SS	4	2.63%	34.21%	51.05%	7.37%	2.11%	0.53%	0.00%	0.53%	1.58%
		7	3.11%	47.67%	47.67%	1.04%	0.00%	0.52%	0.00%	0.00%	0.00%
	SC	5	3.57%	43.03%	43.88%	6.63%	1.87%	0.85%	0.00%	0.00%	0.17%
		8	2.27%	42.05%	47.24%	5.36%	0.97%	0.16%	0.65%	0.16%	1.14%
		HS	2.15%	35.16%	53.32%	6.25%	2.15%	0.20%	0.59%	0.00%	0.20%

Table 48. Correlation Between Student Scores and Familiarity with the Student

Subject	Assessment	N	Correlation
Social Studies	Grade 4	190	0.06
	Grade 7	193	-0.02
Science	Grade 5	588	-0.25
	Grade 8	616	-0.01
	HS	512	-0.02

Table 49. Correlation Between Student Scores and Hours Per Week in Instruction on the Content Area

Subject	Assessment	N	Correlation
Social Studies	Grade 4	190	0.18
	Grade 7	193	0.29
Science	Grade 5	588	0.19
	Grade 8	616	0.38
	HS	512	0.10

Table 50. Correlation Between Student Scores and How Much Instructional Time in the Content Area Is in the General Education Classroom

Subject	Assessment	N	Correlation
Social Studies	Grade 4	190	0.28
	Grade 7	193	0.23
Science	Grade 5	588	0.26
	Grade 8	616	0.35
	HS	512	0.18

Table 51. Items Field Tested and Item Performance Review Outcomes

	Social Studies		Science		
	Grade 4	Grade 7	Grade 5	Grade 8	HS
Number of field test forms	1	1	1	2	2
Number of items field tested	6	6	6	8	8
Item performance review outcome					
• Flagged Items	0	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)? ☐ Is a current Individualized Education Program (IEP) in place or being 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 to receive special education services, 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 Options:**
☐ 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.***	☐ **Yes**. Both elements affirm that the student's evaluated performance falls within range of the most **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 <u>alternate</u> standards-based instruction and appropriate alternate assessment.** ☐ The documented evidence supports the existence of a significant cognitive disability. However the IEP Team determines that with appropriate adaptations (supports and accommodations), the student will receive daily instruction based on the Colorado Academic Standards enrolled **grade-level expectations**. (The student then 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.** ☐ **Yes.** Although the documented evidence supporting the existence of a significant cognitive disability does not fall into the lower ranges, the IEP Team has considered the impact and severity of the disability along with other related factors in order to determine that the student qualifies to receive modified 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.**

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. 9/15

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)
CMAS: **Reading/ Writing (ELA)** **Math** **Social Studies** **Science**	☐ Grade-level **classroom/ district** assessments ☐ with accommodation ☐ without accommodation ☐ **State** Summative Assessment ☐ with accommodations allowed for use on state assessment ☐ without accommodation ☐ Unique Request- pending approval by CDE Assessment Unit	☐ Alternate **classroom/ district** assessments based on alternate standards ☐ Alternate **State** Summative Assessments (Gr. 3-9 and 11) **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.
Other	☐ ACCESS for ELLs (K-12) ☐ with allowable accommodations ☐ Grade 10 Preparatory Exam ☐ Grade 11 College Entrance Exam	☐ Alternate ACCESS for ELLs (Gr. 1-12) ☐ 10th Grade DLM Alternate Assessment ☐ 11th Grade DLM Alternate Assessment
Dual Assessment	*Dual assessment is NOT an option beginning with the 2014-15 school year. If a student meets the guidelines to receive instruction on alternate standards and take alternate assessment based upon those alternate standards, then ALL tested content areas or other state-mandated assessments required for the student's enrolled grade level, will be ALTERNATE assessments.	

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 achievement 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. 9/15

APPENDIX B: COALT: SCIENCE AND SOCIAL STUDIES TEST BLUEPRINTS

CoAlt Blueprint – Grade 4 Social Studies

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

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

CoAlt Blueprint – Grade 5 Science

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

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

CoAlt Blueprint – Grade 7 Social Studies

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

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

CoAlt Blueprint – Grade 8 Science

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

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

CoAlt Blueprint – HS Science

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

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

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

Student Performance Report

Colorado Alternate Assessment

Student: **FIRSTNAME M. LASTNAME**

SASID: **9999999999** Birthdate: **07/07/2003**
School: **SCHOOL NAME (9999)**
District: **DISTRICT NAME (9999)**

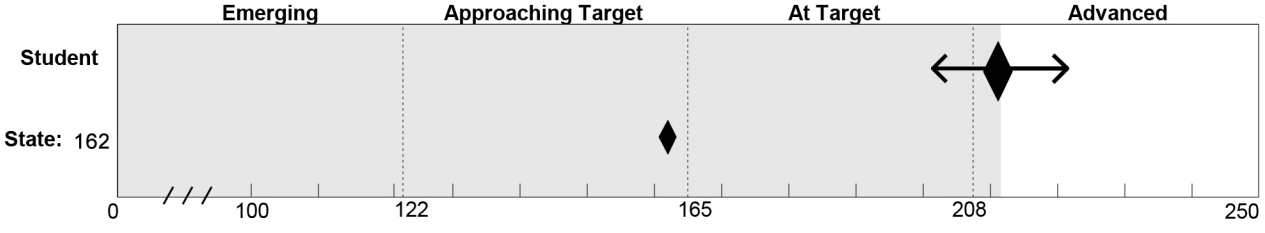
Spring 2016

Social Studies — Grade 7

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

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

Your Student's Score
212
Advanced



Percent of CO students by Performance Level: 13.1% | 15.3% | 33.3% | 7.1%

The Extended Evidence Outcomes of the Colorado Academic Standards include expectations for student performance. Your student demonstrated a novice understanding of high school 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.	15	22	67% / 49%
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.	12	16	75% / 31%
Economics			
Economics teaches how society manages its scarce resources, how people make decisions, how people interact in the domestic and international markets, and how forces and trends affect the economy as a whole. Personal financial literacy applies the economic way of thinking to help individuals understand how to manage their own scarce resources.	11	12	92% / 75%
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.	18	22	83% / 63%

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

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

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.

With appropriate support, Advanced students 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

With appropriate support, At Target students 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

With appropriate support, Approaching Target students 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

With appropraite support, Emerging students 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.

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

School Performance Level Summary

Colorado Alternate Assessment

Spring 2016

District: **DISTRICT NAME (9999)**

Social Studies

CONFIDENTIAL - DO NOT DISTRIBUTE

Grade 7

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

	Number of Valid Scores	Average Scale Score	Performance Levels								At Target and Advanced		No Scores Reported	Total Number of Students
			Emerging		Approaching Target		At Target		Advanced					
			#	%	#	%	#	%	#	%	#	%	#	#
State	**59,723**	**601**	**25,987**	**43.5%**	**23,242**	**38.9%**	**7,788**	**13.0%**	**2,706**	**4.5%**	**10,494**	**17.6%**	**4,605**	**64,328**
District	525	555	328	62.5%	170	32.4%	26	5.0%	1	0.2%	27	5.1%	2	527
Gender														
Female	128	567	75	58.6%	45	35.2%	7	5.5%	1	0.8%	8	6.3%	1	129
Male	128	550	81	63.3%	40	31.3%	7	5.5%	0	0.0%	7	5.5%	1	129
Ethnicity/Race														
Hispanic or Latino	214	555	133	62.1%	71	33.2%	9	4.2%	1	0.5%	10	4.7%	2	216
American Indian or Alaska Native	1	613	0	0.0%	1	100.0%	0	0.0%	0	0.0%	0	0.0%	0	1
Asian	2	653	0	0.0%	2	100.0%	0	0.0%	0	0.0%	0	0.0%	0	2
Black or African-American	5	604	1	20.0%	3	60.0%	1	20.0%	0	0.0%	1	20.0%	0	5
White	29	567	19	65.5%	6	20.7%	4	13.8%	0	0.0%	4	13.8%	0	29
Native Hawaiian or Other Pacific Islander	0	0	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0
Two or more races	5	529	3	60.0%	2	40.0%	0	0.0%	0	0.0%	0	0.0%	0	5
Not Indicated	0	0	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0
Economic Disadvantage														
Free/Reduces Lunch Eligible	95	563	53	55.8%	32	33.7%	9	9.5%	1	1.1%	10	10.5%	1	96
Not Eligible for Free/Reduced Lunch	160	555	103	64.4%	52	32.5%	5	3.1%	0	0.0%	5	3.1%	1	161
Language Proficiency														
Not English Proficient (NEP)	95	563	53	55.8%	32	33.7%	9	9.5%	1	1.1%	10	10.5%	1	96
Limited English Proficient (LEP)	14	449	14	100.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	14
NEP and LEP	103	541	78	75.7%	25	24.3%	0	0.0%	0	0.0%	0	0.0%	1	104
Not NEP and LEP	117	530	92	78.6%	25	21.4%	0	0.0%	0	0.0%	0	0.0%	1	118
Fluent English Proficient	43	622	11	25.6%	27	62.8%	5	11.6%	0	0.0%	5	11.6%	0	43
Primary Home Language other than English	0	0	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0
Former English Language Learner	1	650	0	0.0%	1	100.0%	0	0.0%	0	0.0%	0	0.0%	0	1
Not in ELL Program	139	582	64	46.0%	60	43.2%	14	10.1%	1	0.7%	15	10.8%	1	140
Not Indicated	0	0	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0

This report is NOT for public review. Distribution within your school/district must be in accordance with state and federal privacy laws, and local school board policy.

mmddccyy-Z9999999-99999-9999-9999999

Content Standards Roster

Colorado Alternate Assessment

Spring 2016

School: **SCHOOL NAME (9999)**
District: **DISTRICT NAME (9999)**

Science

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
Advanced	190 - 250
At Target	164 - 189
Approaching Target	128 - 163
Emerging	1 - 127

	Overall Scale Score	Content Standards Performance		
		Physical Science	Life Science	Earth Systems Science
		Points Possible		
		28	30	50
		Percent of Points Earned		
State Average	162	80%	85%	83%
District Average	143	69%	78%	77%
School Average	159	79%	80%	82%

	STUDENT NAME	Overall Performance Level				
1	ALASTNAME, FIRSTNAME A.	At Target	175	67%	100%	97%
2	BLAST, FIRST	Advanced	200	93%	97%	100%
3	BBLAST, FIRST	Emerging	127	75%	47%	64%
4	BDLAST, FIRST	Approaching Target	129	79%	63%	56%
5	CLASTNAME, FIRST E.	At Target	165	82%	93%	92%

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

mmddccyy-Z9999999-99999-9999-9999999

This report is NOT for public review. Distribution within your school/district must be in accordance with state and federal privacy laws, and local school board policy.

APPENDIX D: COALT ALIGNMENT STUDY REPORT

2016 No. 004

HumRRO
HUMAN RESOURCES RESEARCH ORGANIZATION



Independent Alignment Review of the Colorado Alternate Assessment (CoAlt) Science and Social Studies Tests

Final Report

Prepared for:	Colorado Department of Education 201 E. Colfax Ave. Denver, CO 80203	Prepared under:	Contract # C 13-13 Task Order 22
Authors:	Emily R, Dickinson Arthur A. Thacker Richard C. Deatz	Date:	January 15, 2016

Headquarters: 66 Canal Center Plaza, Suite 700, Alexandria, VA 22314 | Phone: 703.549.3611 | Fax: 703.549.9025 | **humrro.org**

Independent Alignment Review of the Colorado Alternate Assessment (CoAlt) Science and Social Studies Tests

Table of Contents

Executive Summary vii
Scope of Work vii
Methodology vii
Review of Content Alignment vii
Summary of Results viii
Key Findings and Conclusions viii
Alignment of CoAlt Science to Colorado Extended Evidence Outcomes ix
Alignment of CoAlt Social Studies to Colorado Extended Evidence Outcomes x
Recommendations x

Chapter 1: Introduction 1
Organization and Contents of the Report 1

Chapter 2: Alignment Study Design and Methodology 3
Alignment of Assessments and Standards 3
Webb Alignment Method 3
Alignment Workshop Process for Science and Social Studies 4
Review of Content Alignment 4
Panelists 4
Training 5
Materials 5
Procedures 6

Chapter 3: Results: Science Content Alignment 9
Reliability Results 9
Panelist-Test Developer Analyses 9
Webb Alignment Results 9
Categorical Concurrence 10
Depth-of-Knowledge Consistency 15
Range of Knowledge Correspondence 19
Balance-of-Knowledge Representation 23
Summary and Discussion on Webb Alignment Indicators 26

Chapter 4: Results: Social Studies Content Alignment 29
Reliability Results 29
Panelist-Test Developer Analyses 29
Webb Alignment Results 29
Categorical Concurrence 30

Depth-of-Knowledge Consistency 34
Range of Knowledge Correspondence 37
Balance-of-Knowledge Representation 39
Summary and Discussion on Webb Alignment Indicators 42

Chapter 5: Summary and Recommendations 45

References 47

Appendix A. Content Alignment Results: Science A-1
Categorical Concurrence A-1
Depth-of-Knowledge Consistency A-5
Range-of-Knowledge Correspondence A-10
Balance-of-Knowledge Representation A-14
EEOs Matched to Items by Panelists A-20

Appendix B. Content Alignment Results: Social Studies B-1
Categorical Concurrence B-1
Depth-of-Knowledge Consistency B-4
Range-of-Knowledge Correspondence B-8
Balance-of-Knowledge Representation B-11
EEOs Matched to Items by Panelists B-15

Appendix C. Panelist Instructions and Extended Evidence Outcome Examples C-1

Appendix D. Item Rating Form Examples D-1

List of Tables

Table 1. Summary Alignment Outcomes on Each Webb Criterion by Grade Level for Science CoAlt ix
Table 2. Summary Alignment Outcomes on Each Webb Criterion by Grade Level for Social studies CoAlt x
Table 2.1. Professional and Demographic Characteristics of Panelists 5
Table 2.2. Number of CoAlt Items Reviewed 6
Table 3.1. Percent Agreement between Panelists and Item Bank on Target Content 9
Table 3.2. Summary of Categorical Concurrence Results for Science CoAlt – Grade 5 10
Table 3.3. Summary of Categorical Concurrence Results for Science CoAlt – Grade 8 11
Table 3.4. Summary of Categorical Concurrence Results for Science CoAlt – High School 12
Table 3.5. Comparison of Panelist Ratings with Test Blueprints 14
Table 3.6. Panelist Ratings on Overall Item Alignment Grade per Forms 15
Table 3.7. Summary of Depth-of-Knowledge Results for Science CoAlt – Grade 5 16
Table 3.8. Summary of Depth-of-Knowledge Results for Science CoAlt – Grade 8 17
Table 3.9. Summary of Depth-of-Knowledge Results for Science CoAlt – High School 17

Table 3.10. Summary of Range-of-Knowledge Results for the Science CoAlt – Grade 5 20
Table 3.11. Summary of Range-of-Knowledge Results for the Science CoAlt – Grade 8 20
Table 3.12. Summary of Range-of-Knowledge Results for the Science CoAlt – High School 21
Table 3.13. Summary of Balance-of-Knowledge Representation Results Science CoAlt – Grade 5 23
Table 3.14. Summary of Balance-of-Knowledge Representation Results Science CoAlt – Grade 8 23
Table 3.15. Summary of Balance-of-Knowledge Representation Results Science CoAlt – High School 24
Table 3.16. Summary Alignment Outcomes on Each Webb Criterion by Grade Level for Science CoAlt 27
Table 4.1. Percent Agreement between Panelists and Item Bank on Target Content 29
Table 4.2. Summary of Categorical Concurrence Results for Social Studies CoAlt – Grade 4 30
Table 4.3. Summary of Categorical Concurrence Results for Social Studies CoAlt – Grade 7 31
Table 4.4. Summary of Categorical Concurrence Results for Social Studies CoAlt – High School 31
Table 4.5. Comparison of Panelist Ratings with Test Blueprints 33
Table 4.6. Panelist Ratings on Overall Item Alignment 33
Table 4.7. Summary of Depth-of-Knowledge Results for Social Studies CoAlt – Grade 4 35
Table 4.8. Summary of Depth-of-Knowledge Results for Social Studies CoAlt– Grade 7 35
Table 4.9. Summary of Depth-of-Knowledge Results for Social Studies CoAlt – High School 36
Table 4.10. Summary of Range-of-Knowledge Results for the Social Studies CoAlt – Grade 4 37
Table 4.11. Summary of Range-of-Knowledge Results for the Social Studies CoAlt– Grade 7 38
Table 4.12. Summary of Range-of-Knowledge Results for the Social Studies CoAlt – High School 38
Table 4.13. Summary of Balance-of-Knowledge Representation Results Social Studies CoAlt – Grade 4 40
Table 4.14. Summary of Balance-of-Knowledge Representation Results Social Studies Coalt – Grade 7 41
Table 4.15. Summary of Balance-of-Knowledge Representation Results Social Studies CoAlt – High School 41
Table 4.16. Summary Alignment Outcomes on Each Webb Criterion by Grade Level for Social studies CoAlt 43
Table A-1. Categorical Concurrence for CoAlt Science, Grade 5: Mean Number of Items per Grade Level Expectation A-1
Table A-2. Categorical Concurrence for CoAlt Science, Grade 8: Mean Number of Items per Grade Level Expectation A-2
Table A-3. Categorical Concurrence for CoAlt Science, High School: Mean Number of Items per Grade Level Expectation A-3

Table A-4. DOK Consistency for CoAlt Science, Grade 5: Mean Percent of Items with DOK Below, At, and Above DOK Level of EOs A-5
Table A-5. DOK Consistency for CoAlt Science, Grade 8: Mean Percent of Items with DOK Below, At, and Above DOK Level of EOs A-6
Table A-6. DOK Consistency for CoAlt Science, High School: Mean Percent of Items with DOK Below, At, and Above DOK Level of EOs A-7
Table A-7. Range-of-Knowledge for CoAlt Science, Grade 5: Mean Percent of EEOs per GLE Linked with Items A-10
Table A-8. Range-of-Knowledge for CoAlt Science, Grade 8: Mean Percent of EEOs per GLE Linked with Items A-11
Table A-9. Range-of-Knowledge for CoAlt Science, High School: Mean Percent of EEOs per GLE Linked with Items A-12
Table A-10. Balance-of-Knowledge Representation for CoAlt Science Grade 5: Mean Balance Index per GLE A-14
Table A-11. Balance-of-Knowledge Representation for CoAlt Science Grade 8: Mean Balance Index per GLE A-15
Table A-12. Balance-of-Knowledge Representation for CoAlt Science High School: Mean Balance Index per GLE A-17
Table A-13. Grade 5 CoAlt Science: EEOs Matched to Items by Panelists A-20
Table A-14. Grade 8 CoAlt Science: EEOs Matched to Items by Panelists A-21
Table A-15. High School CoAlt Science: EEOs Matched to Items by Panelists A-22
Table B-1. Categorical Concurrence for CoAlt Social Studies, Grade 4: Mean Number of Items per Grade Level Expectation B-1
Table B-2. Categorical Concurrence for CoAlt Social Studies, Grade 7: Mean Number of Items per Grade Level Expectation B-2
Table B-3. Categorical Concurrence for CoAlt Social Studies, High School: Mean Number of Items per Grade Level Expectation B-3
Table B-4. DOK Consistency for CoAlt Social Studies, Grade 4: Mean Percent of Items with DOK Below, At, and Above DOK Level of EEOs B-4
Table B-5. DOK Consistency for CoAlt Social Studies, Grade 7: Mean Percent of Items with DOK Below, At, and Above DOK Level of EEOs B-5
Table B-6. DOK Consistency for CoAlt Social Studies, High School: Mean Percent of Items with DOK Below, At, and Above DOK Level of EEOs B-6
Table B-7. Range-of-Knowledge for CoAlt Social Studies, Grade 4: Mean Percent of EEOs per GLE Linked with Items B-8
Table B-8. Range-of-Knowledge for CoAlt Social Studies, Grade 7: Mean Percent of EEOs per GLE Linked with Items B-9
Table B-9. Range-of-Knowledge for CoAlt Social Studies, High School: Mean Percent of EEOs per GLE Linked with Items B-10
Table B-10. Balance-of-Knowledge Representation for CoAlt Social studies Grade 4: Mean Balance Index per GLE B-11
Table B-11. Balance-of-Knowledge Representation for CoAlt Social studies Grade 7: Mean Balance Index per GLE B-12
Table B-12. Balance-of-Knowledge Representation for CoAlt Social studies High School: Mean Balance Index per GLE B-13

Table B-13. Grade 4 CoAlt Social Studies: EEOs Matched to Items by Panelists B-15
Table B-14. Grade 7 CoAlt Social Studies: EEOs Matched to Items by Panelists B-15
Table B-15. High School CoAlt Social Studies: EEOs Matched to Items by Panelists B-16

Independent Alignment Review of the Colorado Alternate Assessment (CoAlt) Science and Social Studies Tests

Executive Summary

Scope of Work

The Human Resources Research Organization (HumRRO) was contracted by Pearson Educational Measurement on behalf of the Colorado Department of Education (CDE) to conduct an external, independent alignment study of the Colorado Alternate Assessments (CoAlt) in science and social studies tests. The alignment study included a review and analysis of the alternate science tests administered at grades 5 and 8 and high school and the social studies tests administered at grades 4 and 7 and high school, to the Colorado Extended Evidence Outcomes for science and social studies, respectively.

CDE requested the alignment study in order to meet both state and federal accountability requirements related to its use of the CoAlt. The federal requirement of the U.S. Department of Education (USDE) stems from the Elementary and Secondary Education Act (ESEA). ESEA challenges each state to establish a coherent assessment system, based on solid academic standards, and encourages the development of alternate assessments for students with significant cognitive difficulties who are not able to access the general assessment. Many states, including Colorado, have developed alternate assessment systems aligned with alternate achievement standards. Similar to general assessments, states should provide independent evidence of the validity of their alternate assessments, including that there is sufficient alignment between alternate achievement standards and alternate assessments. States are required to submit this information as part of the federal peer review process.

An alignment review can provide one form of evidence supporting the validity of the state alternate assessment system. Alignment results should demonstrate that the assessments represent the full range of the extended content standards and that the assessments measure student knowledge in the same manner and at the same level of complexity as specified in the extended content standards.

Methodology

To conduct the study, HumRRO facilitated a review of the alignment between the CoAlt science and social studies items and the Colorado Extended Evidence Outcomes for science and social studies by two panels (one per content area) of Colorado educators. Following the reviews and examination of the alignment, HumRRO analyzed the results for presentation in this report.

Review of Content Alignment

HumRRO convened panels of Colorado educators to review the extent of the alignment between the science and social studies CoAlt tests and the standards they are intended to assess. The review involved two major tasks for panelists to complete: (a) providing depth of knowledge (DOK) ratings for the each of the extended evidence outcomes (EEOs) for science and social studies, and (b) evaluating the science and social studies items by matching them to grade level EEOs, providing an item DOK rating, and selecting a rating of the overall alignment

between item and standard, To maintain the independent and external nature of the study, CDE did not take part in this process. This process was conducted and directed solely by HumRRO.

Pearson recruited the two review panels with the administrative assistance of CDE. Every effort was made to produce panels consisting of teachers reflecting the population of students who take the assessments. Once selected, the panels were convened at the Sheraton Denver West Hotel in Denver, CO on November 12, 2015. Panels included 3-5 reviewers, referred to as panelists.

To conduct the content alignment review, HumRRO applied the Webb (2005) alignment method. This procedure, developed by Dr. Norman Webb, is based on four indicators (or statistics) using the data gathered from the two tasks mentioned above. These statistics describe how well the test items, regardless of item type and point value, cover the content standards in terms of content breadth and depth. The alignment indicators include:

- Categorical concurrence – determines the degree of overall content coverage by the assessment for each content strand. Webb recommends a minimum of six test questions to adequately assess each content strand.
- Range-of-knowledge correspondence – indicates the specific content expectations (i.e., extended evidence outcome) assessed within each strand (i.e., grade level expectation). Webb recommends at least 50% of the GLEs per strand are linked with items.
- Balance-of-knowledge representation – provides a statistical index reflecting the distribution of assessed content within each strand (i.e., how evenly the content is assessed). Webb recommends a minimum index of 70 for a single content strand.
- Depth-of-knowledge consistency – compares the cognitive complexity ratings of the items with the complexity ratings of each content standard. Webb recommends that at least 50% of the items should have complexity ratings at or above the level corresponding to GLEs as determined by panelists.

HumRRO supplemented the Webb criteria with one additional data collection element and two additional analyses. During the alignment workshops, panelists were asked to provide a rating of the quality the match between each item reviewed and the element of the extended content standards to which the item was matched. These ratings were then analyzed and summarized across panelists. Additionally, HumRRO compared panelists' ratings of item content to the content specifications outlines in the test blueprints. These additional analyses were intended to supplement Webb's categorical concurrence criterion, which can be difficult to meet with a small number of operational items on a test form.

Summary of Results

Key Findings and Conclusions

The cumulative results provide validity evidence to support that the content of CoAlt science and social studies test items match the intended content as specified in the extended content standards. Expert panelists from both content areas tended to agree that items were measuring the intended grade level expectations, and to rate items as highly aligned to the Colorado Extended Evidence Outcomes.

The number of items included on an operational form, when considered along with the number of prepared graduate competencies, grade level expectations, and extended evidence outcomes included in the extended content standards, provide important context for interpreting the Webb (1997) criteria. For example, it was essentially impossible for the categorical concurrence correspondence to be fully met given the number of items. Even with this limitation, both content areas were rated as highly or fully aligned on at least three-quarters of the Webb criteria.

Alignment of CoAlt Science to Colorado Extended Evidence Outcomes

Table 1 provides summary conclusions on the alignment of the CoAlt science test to the Colorado Extended Evidence Outcomes per grade tested. The conclusions are based on the following decision criteria (Webb, 2005):

- Fully aligned – assessments align to all content strands (91%–100%),
- Highly aligned – assessments align to the majority of strands (70%–90%),
- Partially aligned – assessments align well to some strands (50%–69%),
- Weakly aligned – assessments align to less than half the strands (below 50%).

Table 1. Summary Alignment Outcomes on Each Webb Criterion by Grade Level for Science CoAlt

Grade Level	Percentage of GLEs that Met Webb Criteria				
	Categorical Concurrence	Depth-of-Knowledge Consistency	Range-of-Knowledge Correspondence		Balance-of-Knowledge Representation
5	Partially aligned (67%)	Fully aligned (100%)	Fully aligned (100%)		Fully aligned (100%)
8	Fully aligned (100%)	Highly aligned (80%)	Highly aligned (90%)		Highly aligned (90%)
High School	Fully aligned (100%)	Highly/fully aligned (82%; 91%)	Highly aligned (90%)	Partially aligned (64%)	Fully aligned (95%; 100%)

Notes. Categorical concurrence is evaluated at the Standard level to reflect score reporting practices. High school criteria with multiple percentages reflect GLEs and PGCs, respectively.

As shown in Table 1 with green highlighting, roughly 92% of the results indicate strong content alignment of the CoAlt science test to the Colorado Extended Evidence Outcomes. Each of the three grade level tests includes a sufficiently even distribution of extended evidence outcomes within the associated grade level expectation and sufficient coverage of the range of extended evidence outcomes within GLEs. The high school and grade 8 science tests include sufficient numbers of items to cover the Colorado Extended Evidence Outcomes at the Standard level. The three grade level tests also include sufficient numbers of items at DOK levels at or above the DOK assigned to the corresponding EEOs.

Additional analyses by HumRRO found that panelists indicated that the CoAlt items reflect the intended content of the test blueprints, and that the large majority of items are highly aligned to the particular extended evidence outcomes to which they were matched.

Alignment of CoAlt Social Studies to Colorado Extended Evidence Outcomes

Table 2 provides summary conclusions on the alignment of the CoAlt social studies test to the Colorado Extended Evidence Outcomes per grade tested, using the same criteria described above.

Table 2. Summary Alignment Outcomes on Each Webb Criterion by Grade Level for Social Studies CoAlt

Grade Level	Percentage of GLEs that Met Webb Criteria			
	Categorical Concurrence	Depth-of-Knowledge Consistency	Range-of-Knowledge Correspondence	Balance-of-Knowledge Representation
4	Weakly aligned (25%)	Highly aligned (75%)	Fully aligned (100%)	Fully aligned (100%)
7	Weakly aligned (0%)	Partially aligned (63%)	Fully aligned (100%)	Fully aligned (100%)
High School	Highly aligned (75%)	Highly aligned (75%)	Fully aligned (100%)	Fully aligned (100%)

Note. High school percentages reflect GLEs and PGCs, respectively.

As shown in Table 2, approximately 75% of the results indicate strong content alignment of the CoAlt social studies test to the Colorado Extended Evidence Outcomes. Each of the three grade level tests includes sufficient coverage of the range of evidence outcomes, and a sufficiently even distribution of evidence outcomes within the associated grade level expectation. There was evidence of high alignment in terms of the numbers of items at DOK levels at or above the DOK assigned to the corresponding EEOs for grade 4 and high school. Only the high school social studies test demonstrated sufficient coverage of the extended content standards at the Standard Level.

Additional analyses by HumRRO found that panelists did indicate that the CoAlt items reflect the intended content of the test blueprints, and that the large majority of items were highly aligned to the particular extended evidence outcomes to which they were matched.

Recommendations

HumRRO makes the following recommendations to strengthen the alignment between the components of the Colorado assessment system:

- ***Review content coverage (categorical concurrence).*** Assessments may not adequately reflect the content that students are expected to know based solely on the number of items on the assessment (not the item type or point value as these are not factors in Webb's (1997) categorical concurrence indicator). From strictly an item count perspective, there are several ways CDE can choose to mitigate this situation such as increase the number of items on the assessment, collapse or otherwise reduce the number of grade level expectations in the extended content standards, or designate some of the grade level expectations for local assessment only. Based on this study, there may not be a sufficient number of items to support standard level scores in grades 4 and 8.
- ***Review Grade 4 social studies item metadata***. Comparisons of panelist's ratings to item bank data and to the content specifications in the test blueprint showed notably

larger discrepancies for the grade 4 social studies test. It may be useful to conduct an internal review to verify that grade 4 social studies item metadata contain no errors.

Independent Alignment Review of the Science Colorado Assessment Program (COALT)

Chapter 1: Introduction

The Human Resources Research Organization (HumRRO) was contracted by Pearson Educational Measurement on behalf of the Colorado Department of Education (CDE) to conduct an external, independent alignment study of the Colorado Alternate Assessments (CoAlt) in science and social studies tests. The alignment study included a review and analysis of the science tests administered at grades 5 and 8 and high school and the social studies tests administered at grades 4 and 7 and high school, to the Colorado Extended Evidence Outcomes for science and social studies, respectively.

CDE requested the alignment study in order to meet both state and federal accountability requirements related to its use of the CoAlt. The federal requirement of the U.S. Department of Education (USDE) stems from the Elementary and Secondary Education Act (ESEA). ESEA challenges each state to establish a coherent assessment system, based on solid academic standards, and encourages the development of alternate assessments for students with significant cognitive difficulties who are not able to access the general assessment. Many states, including Colorado, have developed alternate assessment systems aligned with alternate achievement standards. Similar to general assessments, states should provide independent evidence of the validity of their alternate assessments, including that there is sufficient alignment between extended standards and assessments. States are required to submit this information as part of the federal peer review process.

An alignment review can provide one form of evidence supporting the validity of the state assessment system. Alignment results should demonstrate that the alternate assessments represent the full range of the extended content standards and that the assessments measure student knowledge in the same manner and at the same level of complexity as specified in the extended content standards.

Organization and Contents of the Report

This report contains five chapters. Chapter 2 explains the alignment methodologies used in the study and chapters 3 and 4 provide alignment results for science and social studies, respectively. Chapter 5 summarizes and results and provides recommendations.

Additional information is provided in the appendices of this report. Appendix A contains tables with additional details for each Webb (1997) indicator regarding the content alignment results for each science test, Appendix B contains tables with additional details for each Webb indicator regarding the content alignment results for each social studies test, and Appendix C and Appendix D provide examples of rating forms and training materials used in the alignment workshops.

Chapter 2: Alignment Study Design and Methodology

In this section, we discuss key concepts related to assessment alignment research. This discussion is followed by a description of the alignment evaluations and methods used for this study.

Alignment of Assessments and Standards

Alignment studies, at their heart, answer one vital question related to the validity of an assessment, "Does the assessment content adequately reflect the content that students are expected to learn as provided in the associated content standards?" In general, alignment evaluations for are conducted to document (a) the breadth, or scope, of knowledge and (b) the depth of knowledge, or cognitive processing, expected of students by the content standards that the test is designed to measure. In addition to the question related to assessment validity, alignment analyses help to answer questions such as the following:

- How much and what type of content is covered by the assessment?
- Are students asked to demonstrate this knowledge at the same level of rigor as expected in the content standards?

For this alternate assessment alignment study, HumRRO primarily used the Webb Alignment Method, which is described below. The Webb method was supplemented with one additional data collection element and two additional analyses. During the alignment workshops, panelists were asked to provide a rating of the quality the match between each item reviewed and the element of the extended content standards to which the item was matched. These ratings were then analyzed and summarized across panelists. Additionally, HumRRO compared panelists' ratings of item content to the content specifications outlines in the test blueprints. These additional analyses were intended to supplement Webb's categorical concurrence criterion, which can be difficult to meet with a small number of operational items on a test form.

Webb Alignment Method

HumRRO used the methodology originally developed by Dr. Norman Webb (1997; 1999; 2005). Alignment evidence submitted by states for federal peer review is routinely obtained with Webb's methodology and his approach is supported by the Council of Chief State School Officers (CCSSO).

The Webb method includes four major indicators to evaluate alignment. These indicators link with statistical procedures used to assess how well items on the assessment, regardless of item type and point value, and the extended content standards document actually match. The four alignment indicators are: categorical concurrence, depth-of-knowledge consistency, range-of-knowledge correspondence, and balance-of-knowledge representation.

Categorical concurrence is a basic measure of alignment between extended content standards and test items. This term refers to the proportion of overlap between the content stated in the standards document and that assessed by items on the test.

Range-of-knowledge correspondence examines the range-of-knowledge correspondence between the assessment and extended content standards. The range-of-knowledge correspondence measure looks in greater detail at the breadth of knowledge represented by test items. Categorical concurrence simply notes whether a sufficient number of items on the

test covers each general content topic (i.e., grade level expectations). However, states usually lay out more specific *content objectives* (i.e., extended evidence outcomes) under each strand. The range-of-knowledge correspondence indicates the number of content objectives assessed by items.

Balance-of-knowledge representation focuses on content coverage in yet more detail. In this case, the number of items matched to the content objective does matter. The balance of representation determines whether the assessment measures the content objectives equitably within each standard using only the content objectives identified by panelists and not all content objectives eligible to be assessed. Based on Webb's (1997) method, items should be distributed evenly across the objectives per standard for good balance. The balance-of-knowledge representation is determined by calculating an index, or score, for each standard. Each standard should meet or surpass a minimum index level to demonstrate adequate balance.

Depth of knowledge (DOK) measures the type of cognitive processing required by items and extended content standards. For example, is a student expected to simply identify or recall basic facts or use reason to manipulate information, or to strategize how to best solve a complex problem?

The purpose of using DOK as a measure of alignment is to determine whether a test item and its corresponding extended content standard are written at the same level of cognitive complexity. Panelists make two separate judgments about cognitive complexity, one rating for the extended standard and one rating for the item. These two judgments are compared to determine whether the item is written at the same level as the extended standard to which it is linked. Webb (1997) refers to this comparison as *Depth-of-Knowledge consistency*.

Alignment Workshop Process for Science and Social Studies

The alignment evaluation performed for this study involved a comparison of the CoAlt science and social studies test items to the Colorado Extended Evidence Outcomes. Colorado educators highly familiar with the extended content standards and the assessment provided alignment ratings for the evaluation. To maintain the independent and external nature of the study, CDE did not take part in this process. This process was conducted and directed solely by HumRRO.

Review of Content Alignment

For the content alignment review, HumRRO convened panels of Colorado educators to review grades 5, 8, and high school CoAlt science items, and grades 4, 7, and high school CoAlt social studies items. The review involved two major tasks for panelists to complete: (a) providing depth of knowledge (DOK) ratings for each extended evidence outcome (EEO) for science and social studies, and (b) evaluating the science and social studies items by matching them to a grade level EEO, providing an item DOK rating, and selecting a rating of the quality of alignment between the item and the matched EEO.

Panelists

Pearson recruited the two review panels, science and social studies, with the administrative assistance of CDE. Every effort was made to produce panels consisting of 4-5 teachers or administrators reflecting the population of students who take the assessments. Approximately half of the panels were from suburban school settings while the rest were split between urban

and rural. Table 2.1 presents the characteristics of the panels by content area and grade level. Once selected, the panels were convened at the Sheraton Denver West Hotel in Denver, CO on November 12, 2015.

Table 2.1. Professional and Demographic Characteristics of Panelists

Professional Position	Science									Social Studies								
	# Panelist	Specialty			Education			Gender		# Panelist	Specialty			Education			Gender	
		Content	SpEd	ELL	BA	MA	PhD	M	F		Content	SpEd	ELL	BA	MA	PhD	M	F
Elementary																		
Teacher	4	0	4	0	1	3	0	0	4	3	1	1	0	1	1	0	1	2
Administrator	1	1	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0
Middle																		
Teacher	5	0	4	0	1	2	0	0	5	5	0	3	2	1	2	1	2	3
Administrator	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
High School																		
Teacher	4	0	2	0	1	1	0	0	4	4	0	3	0	0	2	0	0	4
Administrator	0	0	0	0	0	0	0	0	0	1	0	0	1	0	1	0	0	1

Note. Demographic data were not available for a1ll participants.

Training

All panelists received a common introductory training prior to participating in the study. During this large group session, HumRRO provided general alignment study information, roles and responsibilities, key alignment concepts, security and confidentiality concerns, and the alignment workshop procedures. Panelists then moved into content-specific breakout sessions in which they were assigned to a specific grade level. In the breakout rooms, panelists signed non-disclosure agreements and then received additional, targeted training on the process and associated materials prior to beginning their evaluation.

Materials

During the alignment workshop, panelists evaluated the alignment of the CoAlt items with the EEOs by accessing reviewing paper copies of the items and completing electronic rating forms adapted from Webb (2005). All rating forms were completed electronically in Excel®. The item presentation and rating forms are discussed in further detail below.

Test Items. Panelists evaluated CoAlt operational items. Table 2.2 lists the number of items for each grade-level test. The CoAlt test items include directions for test administrators and scoring rubrics. Panelists were given access to all these materials for their review. Because the test items are secure, this report does not include any examples of items or references to specific item content.

Table 2.2. Number of CoAlt Items Reviewed

Subject	Grade	Total Items
Science	5	17
	8	26
	High school	26
Social Studies	4	17
	7	17
	High school	26

Rating Forms and Instructions. Panelists were given instruction sheets describing the rating tasks, the codes to be used, and the excel documents used during their review (see Appendix C). Panelists completed two rating forms, the first was completed as a group (by consensus) to provide depth of knowledge (DOK) ratings for the EEOs and the second form, an item rating form, captures individual ratings for the items (see Appendix C and Appendix D for samples of each).

Procedures

HumRRO conducted the alignment study at Sheraton Denver West Hotel in Denver, CO. The workshop began with a general session that included introductions of staff and observers followed by a brief review of the agenda for the two-day workshop. Panelists then moved to content area breakout rooms to receive more targeted alignment task training before starting to work. Within each breakout room, panelists were seated at grade-specific tables, with 3-5 panelists per group. One HumRRO staff member served as a facilitator in each breakout room. A third HumRRO staff member moved between the rooms and provided assistance as needed. Prior to beginning their review, panelists read and signed affidavits of nondisclosure for the secure materials they would be reviewing during the workshop.

Before each of the rating tasks, a HumRRO staff member trained panelists on the procedures to complete the task, answered questions on the rating criteria, and facilitated a short calibration activity to ensure panelists were comfortable applying ratings. HumRRO staff provided general suggestions and comments when appropriate; however, they emphasized to panelists that staff would not give explicit direction on how to rate standards or items because panelists were valued as content experts. Each panelist was assigned a workstation with rating forms already uploaded on their assigned laptop computer. HumRRO staff provided instructions as needed for working with the electronic rating forms.

Panelists began with DOK evaluations of the content EEOs. Panelists started this process by independently assigning a DOK level to one EEO and then discussing their individual ratings with the group until a consensus rating was reached. When all panelists felt comfortable with the task, groups followed a similar process in which they provided independent ratings for each EEO prior to identifying a group consensus rating. A volunteer scribe within each group recorded these consensus ratings.

Panelists then received specific instructions for rating the items. As a calibration activity, HumRRO staff asked panelists to rate the first two items individually and then discuss the ratings as a group. Once panelists were comfortable using the ratings, they continued the item rating activity on their own. Panelists rated the individual items on the test forms on several dimensions: (a) depth of knowledge required by the item, (b) content match to the EEOs,

(c) and the degree of alignment (i.e., how well the item links to the identified EEO). Within the content match dimension, panelists assigned a *primary EEO* to an item based on a judgment that an item clearly measured this content. Panelists could also assign an *additional EEO* if the item seemed to assess another EEO as well (or nearly as well) as the primary GLE. Again, these were individual ratings, not consensus.

All panelists finished their rating tasks within the 1 day allotted for the workshop. Once panelists finished the review, their session ended.

Chapter 3: Results: Science Content Alignment

The content alignment evaluation analyses discussed in this chapter are based on panelists' ratings of the COALT science items for grades 5 and 8 and high school.

Reliability Results

In this section, we report on the comparison of panelists' ratings of content match to the item bank's documented content match. In other words, do panelists assign the same EEO to an item as the item writer during item development?

Panelist-Test Developer Analyses

This analysis examined the agreement outcomes between the EEO assigned to an item by panelists, and the EEO assigned to an item as noted in the item bank. Table 3.1 presents the agreement outcomes between panelists and the item bank on the content assessed by items. Agreement was analyzed at several levels of specificity. All of the items were analyzed first for 'Exact Match', which indicates that panelists chose the same EEO. If panelists did not show an exact match with the item bank, we determined the percent agreement at the Grade Level Expectation (GLE) level. For high school, we also determine the percent agreement at the Prepared Graduate Competency (PGC) level. Finally, we determined the percent agreement at the standard level (i.e., physical science, life science, earth systems science).The last column in Table 3.2 shows the percentage of ratings by panelists that did not match the item bank coding at all on items.

Table 3.1. Percent Agreement between Panelists and Item Bank on Target Content

Grade	Total Number of Panelist Ratings across Items	Percent Agreement with Item Bank Codes				
		Exact Match	GLE Match	PGC Match	Standard Match	No Match
5	85	91.8%	94.1%	NA	94.1%	5.9%
8	130	86.2%	93.8%	NA	100%	0.0%
High School	104	78.8%	85.6%	93.3%	95.2%	4.8%

As Table 3.1 indicates, panelists were very consistent with the item bank in identifying the content codes of items. Panelists identified an exact match for 79–92% of the items and a match at the GLE level or below for 86–94% of the items. Panelists differed completely from the item bank on content match for only 0–6% of the items. Overall these findings suggest that the majority of science items do measure the intended content.

Webb Alignment Results

In this section, we review the general outcomes of item analyses on the four Webb alignment indicators.

All of Webb's (1997) measures begin with calculations for each panelist and build up to a summary of results across panelists per EEO. First, we calculated the mean ratings across items for each panelist, and then we determined the mean rating across panelists per EEO. Depending on the component under review, results are presented at the broader GLE and

Standard levels (as well as the PGC level for high school). Results at the more specific EEO level are presented in Appendix A.

Categorical Concurrence

Categorical concurrence describes the extent to which the CoAlt items, regardless of item type and point value, cover the content grade level expectations of the Colorado Extended Evidence Outcomes. Webb (1997, 1999, 2005) recommends a minimum of six test questions to adequately assess each grade level expectation. This criterion serves as a guideline for reasonable content coverage based on earlier research on the reliability of tests compared to the number of items (Subkoviak, 1988). Tables 3.2 through 3.4 summarize the CoAlt alignment results for categorical concurrence for each grade level. The GLEs, PGCs, and Standards that meet Webb's indicator criterion are in bold. Tables A-1 through A-3 in Appendix A also contain the standard deviations for each GLE.

Table 3.2. Summary of Categorical Concurrence Results for Science CoAlt – Grade 5

Standard	Grade Level Expectation	Mean Number of Items per GLE	Mean Number of Items per Standard
Physical Science	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.	3.00	3.00
Life Science	All organisms have structures and systems with separate functions.	4.00	**7.00**
	Human body systems have basic structures, functions, and needs.	3.00	
Earth Systems Science	Earth and Sun provide a diversity of renewable and nonrenewable resources.	2.00	**7.00**
	Earth's surface changes constantly through a variety of processes and forces.	2.00	
	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.	3.00	
	GLEs with at Least Six Items	**0 of 6**	
	Standards with at Least Six Items	**2 of 3**	

Table 3.3. Summary of Categorical Concurrence Results for Science CoAlt – Grade 8

Standard	Grade Level Expectation	Mean Number of Items per GLE	Mean Number of Items per Standard
Physical Science	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.	0.00	**6.00**
	There are different forms of energy, and those forms of energy can be changed from one form to another – but total energy is conserved.	1.60	
	Distinguish between physical and chemical changes, noting that mass is conserved during any change.	2.00	
	Recognize that waves such as electromagnetic, sound, seismic, and water have common characteristics and unique properties.	2.40	
	Human activities can deliberately or inadvertently alter ecosystems and their resiliency.	3.00	
Life Science	Organisms reproduce and transmit genetic information (genes) to offspring, which influences individuals' traits in the next generation.	5.00	**8.00**
	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.	2.00	
Earth Systems Science	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.	3.00	**12.00**
	The solar system is comprised of various objects that orbit the Sun and are classified based on their characteristics.	4.20	
	The relative positions and motions of Earth, Moon, and Sun can be used to explain observable effects such as seasons, eclipses, and Moon phases.	2.80	
	GLEs with at Least Six Items	**0 of 10**	
	Standards with at Least Six Items	**3 of 3**	

Table 3.4. Summary of Categorical Concurrence Results for Science CoAlt – High School

Standard	Prepared Graduate Competencies	Grade Level Expectation	Mean N of Items GLE	Mean N of Items PGC	Mean N of Items Standard
Physical Science	Observe, explain, and predict natural phenomena governed by Newton's laws of motion, acknowledging the limitations of their application to very small or very fast objects.	Newton's laws of motion and gravitation describe the relationships among forces acting on and between objects, their masses, and changes in their motion – but have limitations.	1.25	1.25	**7.25**
	Apply an understanding of atomic and molecular structure to explain the properties of matter, and predict outcomes of chemical and nuclear reactions.	Matter has definite structure that determines characteristic physical and chemical properties.	1.50	4.00	
		Matter can change form through chemical or nuclear reactions abiding by the laws of conservation of mass and energy.	1.00		
		Atoms bond in different ways to form molecules and compounds that have definite properties.	1.50		
	Apply an understanding that energy exists in various forms, and its transformation and conservation occur in processes that are predictable and measurable.	Energy exists in many forms such as mechanical, chemical, electrical, radiant, thermal, and nuclear, that can be quantified and experimentally determined.	1.00	2.00	
		When energy changes form, it is neither created not destroyed; however, because some is necessarily lost as heat, the amount of energy available to do work decreases.	1.00		
Life Science	Explain and illustrate with examples how living systems interact with the biotic and abiotic environment.	Matter tends to be cycled within an ecosystem, while energy is transformed and eventually exits an ecosystem.	1.00	1.00	**10.00**
		The size and persistence of populations depend on their interactions with each other and on the abiotic factors in an ecosystem.	0.00		
	Analyze the relationship between structure and function in living systems at a variety of organizational levels, and recognize living systems' dependence on natural selection.	Cellular metabolic activities are carried out by biomolecules produced by organisms.	2.25	5.75	
		The energy for life primarily derives from the interrelated processes of photosynthesis and cellular respiration. Photosynthesis transforms the sun's light energy into the chemical energy of molecular bonds. Cellular respiration allows cells to utilize chemical energy when these bonds are broken.	1.00		
		Cells use passive and active transport of substances across membranes to maintain relatively stable intracellular environments.	1.75		
		Cells, tissues, organs, and organ systems maintain relatively stable	1.50		

Standard	Prepared Graduate Competencies	Grade Level Expectation	Mean N of Items GLE	Mean N of Items PGC	Mean N of Items Standard
		internal environments, even in the face of changing external environments.			
	Analyze how various organisms grow, develop, and differentiate during their lifetimes based on an interplay between genetics and their environment.	Physical and behavioral characteristics of an organism are influenced to varying degrees by heritable genes, many of which encode instructions for the production of proteins.	1.00	2.25	
		Multicellularity makes possible a division of labor at the cellular level through the expression of select genes, but not the entire genome.	1.25		
	Explain how biological evolution accounts for the unity and diversity of living organisms.	Evolution occurs as the heritable characteristics of populations change across generations and can lead populations to become better adapted to their environment.	1.00	1.00	
Earth Systems Science	Describe and interpret how Earth's geologic history and place in space are relevant to our understanding of the processes that have shaped our planet.	The history of the universe, solar system and Earth can be inferred from evidence left from past events.	1.00	2.00	**8.75**
		As part of the solar system, Earth interacts with various extraterrestrial forces and energies such as gravity, solar phenomena, electromagnetic radiation, and impact events that influence the planet's geosphere, atmosphere, and biosphere in a variety of ways.	1.00		
	Evaluate evidence that Earth's geosphere, atmosphere, hydrosphere, and biosphere interact as a complex system.	The theory of plate tectonics helps explain geological, physical, and geographical features of Earth.	2.00	3.50	
		Climate is the result of energy transfer among interactions of the atmosphere, hydrosphere, geosphere, and biosphere.	1.50		
	Describe how humans are dependent on the diversity of resources provided by Earth and Sun.	There are costs, benefits, and consequences of exploration, development, and consumption of renewable and nonrenewable resources.	1.00	1.00	
	Evaluate evidence that Earth's geosphere, atmosphere, hydrosphere, and biosphere interact as a complex system.	The interaction of Earth's surface with water, air, gravity, and biological activity causes physical and chemical changes	1.50	2.50	
		Natural hazards have local, national and global impacts such as volcanoes, earthquakes, tsunamis, hurricanes, and thunderstorms	1.00		
		GLEs with at Least Six Items	**0 of 22**		
		PGCs with at Least Six Items	**0 of 11**		
		Standards with at Least Six Items	**3 of 3**		

As Tables 3.2 through 3.4 indicate, none of the science tests includes a sufficient number of items to meet the minimum requirements for Webb's (1997) categorical concurrence on any of the science GLEs. However, when categorical concurrence is evaluated at the Standard level, 2 of the 3 grade 5 standards and all of the standards for grade 8 and high school are adequately covered by the test items.

These results indicate that the CoAlt science test does not adequately cover all the grade level expectations of the extended science content. However, these results are in part shaped by the number of items reviewed. In high school, for example, in order for there to be a minimum of 6 items representing each of the 22 GLEs, a minimum of 132 items would need to be included. The number of items reviewed at each grade level was too low to allow for the categorical concurrence criterion to be met for all GLEs. Table 3.2 shows that in grade 5 each GLE was matched to at least two items. Tables 3.3 and 3.4 show that only one grade 8 and one high school GLE was not matched to any items. Colorado reports student scores at the Standard level only. These alignment results indicate that there are enough items per standard to support reporting student scores at this level, with the exception of grade 5 physical science.

Because of the limitations inherent in the Webb (1997) criteria due to the minimum item requirements, it is helpful to consider how well the panelists' ratings match the distribution of content as outlined in the test specifications. Table 3.5 presents a comparison of the average number of items matched to each standard as compared to the number of items per standard outlined in the CoAlt test blueprints. Table 3.5 shows that the CoAlt science tests reflect the intended content.

Table 3.5. Comparison of Panelist Ratings with Test Blueprints

Standard	Number of Items per Standard					
	Grade 5		Grade 8		High School	
	Panelists	Blueprint	Panelists	Blueprint	Panelists	Blueprint
Physical Science	3	3	6	7	7.25	7
Life Science	7	7	8	7	10	11
Earth Systems Science	7	7	12	12	8.75	8

In addition to identifying the content assessed by each item, we asked panelists to indicate *how well* the item assessed the content. Panelists subjectively rated the extent of item alignment to the content on a 4-point scale ranging from 'not aligned to any EEO' to 'fully aligned'. Table 3.6 presents the mean number of items (across panelists) at each level of alignment. For each grade level, panelists rated items as well aligned to the EEO matched to that item.

Table 3.6. Panelist Ratings on Overall Item Alignment Grade per Forms

Grade (N items)	Degree of Alignment	Mean Number of Items per Level	*SD*	Percent of Items per Level
5 (N= 17)	Not at all aligned	0.00	0.00	0.00%
	Weakly aligned	1.00	0.00	3.53%
	Highly aligned	3.60	1.14	21.18%
	Fully aligned	12.80	1.48	75.29%
8 (N= 26)	Not at all aligned	0.00	0.00	0.00%
	Weakly aligned	0.00	0.00	0.00%
	Highly aligned	7.00	6.12	26.92%
	Fully aligned	19.00	6.12	73.08%
High School (N= 26)l	Not at all aligned	0.00	0.00	0.00%
	Weakly aligned	1.50	0.71	2.88%
	Highly aligned	9.75	2.22	37.50%
	Fully aligned	15.50	3.00	59.62%

In general, panelists across the three grade levels rated at least 96% of the items as being 'highly aligned' or 'fully aligned'. The grade 5 assessment had the highest percentage of items rated by panelists as being 'weakly aligned' at 3.5%. No items were rated as 'not at all aligned'.

Depth-of-Knowledge Consistency

Analyses of depth-of-knowledge (DOK) measure the type of cognitive processing required of students by content standards. The DOK requirements implied by the EEOs should be matched by assessment items. To confirm this match, panelists were asked to rate the EEOs and the science items separately. Webb (1997) includes an alignment indicator that directly compares panelists' DOK ratings of content standards and test items, which he refers to as *depth-of-knowledge consistency*.

To make their ratings of the extended content standards and test items, panelists used a modified version of a cognitive complexity rating scale developed for evaluating the depth of knowledge of alternate assessments (see Flowers, Wakeman, Browder, & Karvonen, 2007). However, during analysis, panelists' DOK ratings of the extended standards and items were collapsed into a three-point classification scheme to better correspond with the DOK guidance for the alternate standards provided in the test blueprints and with the DOK classification scheme used in the item bank.

The rating categories assigned by panelists and the recoded DOK value (in parentheses) included:

- Level 0 None: No content clearly measured; too vague. (*not assigned by any panelist*)
- Level 1 Attention: Requires students to display ability to acknowledge, reply, and respond to text or related subject features. (*recoded Level 1: Recall and Reproduction*)
- Level 2 Memorize/recall: Requires the ability to recite or recall facts or information. It involves the ability to distinguish between simple text-based and one-step procedures. (*recoded Level 1: Recall and Reproduction*)

- Level 3 Performance: Requires students to use recalled facts or information for simple tasks. (*recoded Level 1: Recall and Reproduction*)
- Level 4 Comprehension: Requires processing beyond recall and observation and may require both understanding and subsequent processing of text. It involves ordering, classifying, estimating text or numbers as well as identifying patterns, main points, or two-step procedures. (*recoded Level 2: Skills and Concepts*)
- Level 5 Application: Show ability to go beyond text; to reason, plan, or use of evidence to connect ideas. Students will use text, data, or observations to draw conclusions or solve non-routine problems. (*recoded Level 2: Skills and Concepts*)
- Level 6 Analysis, Synthesis, Evaluation: Requires extended higher order processing. It typically requires extended time to complete a task, but the time is not spent on repetitive tasks. It involves taking information and applying this information to a new task; which may require generating a hypothesis, perform complex analyses, or make connections among different texts. (*recoded Level 3: Strategic Thinking and Reasoning*)

Tables 3.7 through 3.9 summarize the depth-of-knowledge consistency results for each grade level of the CoAlt science test. Because panelists evaluated depth of knowledge at the most specific level of the standards document (EEOs), the table refers to consistency between the items and the EEOs to which they were matched. Results are summarized at the GLE level for ease of presentation. Tables A-4 through A-6 in Appendix A contain the means and standard deviations for DOK ratings at all levels.

Webb's (1997) suggested criterion for this alignment indicator is that at least 50% of the items should have complexity ratings at or above the level of the corresponding EEO. The percentages on strands that reach the 50% criterion are bolded.

Table 3.7. Summary of Depth-of-Knowledge Results for Science CoAlt – Grade 5

Standard	Grade Level Expectations	Percent of Items with DOK At or Above the Level of the EEOs
Physical Science	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.	**100.00**
Life Science	All organisms have structures and systems with separate functions.	**60.00**
	Human body systems have basic structures, functions, and needs.	**93.33**
Earth Systems Science	Earth and Sun provide a diversity of renewable and nonrenewable resources.	**60.00**
	Earth's surface changes constantly through a variety of processes and forces.	**100.00**
	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.	**66.67**
	Number of GLEs with item DOK at or above EEO DOK	**6 of 6**

Table 3.8. Summary of Depth-of-Knowledge Results for Science CoAlt – Grade 8

Standard	Grade Level Expectations	Percent of Items with DOK At or Above the Level of the EEOs
Physical Science	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.	--
	There are different forms of energy, and those forms of energy can be changed from one form to another – but total energy is conserved.	46.67
	Distinguish between physical and chemical changes, noting that mass is conserved during any change.	**80.00**
	Recognize that waves such as electromagnetic, sound, seismic, and water have common characteristics and unique properties.	**70.00**
Life Science	Human activities can deliberately or inadvertently alter ecosystems and their resiliency.	**73.33**
	Organisms reproduce and transmit genetic information (genes) to offspring, which influences individuals' traits in the next generation.	**84.00**
Earth Systems Science	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.	**90.00**
	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.	**60.00**
	The solar system is comprised of various objects that orbit the Sun and are classified based on their characteristics.	**77.33**
	The relative positions and motions of Earth, Moon, and Sun can be used to explain observable effects such as seasons, eclipses, and Moon phases.	**90.00**
	Number of GLEs with item DOK at or above EEO DOK	**8 of 10**

Table 3.9. Summary of Depth-of-Knowledge Results for Science CoAlt – High School

Standard	Prepared Graduate Competencies	Grade Level Expectations	Percent of Items with DOK At or Above the Level of the EEOs	
			GLE	PGC
Physical Science	Observe, explain, and predict natural phenomena governed by Newton's laws of motion, acknowledging the limitations of their application to very small or very fast objects.	Newton's laws of motion and gravitation describe the relationships among forces acting on and between objects, their masses, and changes in their motion – but have limitations.	**87.50**	**87.50**
	Apply an understanding of atomic and molecular structure to explain the properties of matter, and predict outcomes of chemical and nuclear reactions.	Matter has definite structure that determines characteristic physical and chemical properties.	**100.00**	**87.50**
		Matter can change form through chemical or nuclear reactions abiding by the laws of conservation of mass and energy.	**100.00**	
		Atoms bond in different ways to form molecules and compounds that have definite properties.	**75.00**	

Standard	Prepared Graduate Competencies	Grade Level Expectations	Percent of Items with DOK At or Above the Level of the EEOs	
			GLE	PGC
	Apply an understanding that energy exists in various forms, and its transformation and conservation occur in processes that are predictable and measurable.	Energy exists in many forms such as mechanical, chemical, electrical, radiant, thermal, and nuclear, that can be quantified and experimentally determined.	**100.00**	**100.00**
		When energy changes form, it is neither created not destroyed; however, because some is necessarily lost as heat, the amount of energy available to do work decreases.	**100.00**	
Life Science	Explain and illustrate with examples how living systems interact with the biotic and abiotic environment.	Matter tends to be cycled within an ecosystem, while energy is transformed and eventually exits an ecosystem.	**50.00**	**50.00**
		The size and persistence of populations depend on their interactions with each other and on the abiotic factors in an ecosystem.	--	
	Analyze the relationship between structure and function in living systems at a variety of organizational levels, and recognize living systems' dependence on natural selection.	Cellular metabolic activities are carried out by biomolecules produced by organisms.	**54.17**	**59.17**
		The energy for life primarily derives from the interrelated processes of photosynthesis and cellular respiration. Photosynthesis transforms the sun's light energy into the chemical energy of molecular bonds. Cellular respiration allows cells to utilize chemical energy when these bonds are broken.	0.00	
		Cells use passive and active transport of substances across membranes to maintain relatively stable intracellular environments.	**100.00**	
		Cells, tissues, organs, and organ systems maintain relatively stable internal environments, even in the face of changing external environments.	**100.00**	
	Analyze how various organisms grow, develop, and differentiate during their lifetimes based on an interplay between genetics and their environment.	Physical and behavioral characteristics of an organism are influenced to varying degrees by heritable genes, many of which encode instructions for the production of proteins.	**100.00**	**100.00**
		Multicellularity makes possible a division of labor at the cellular level through the expression of select genes, but not the entire genome.	**100.00**	
	Explain how biological evolution accounts for the unity and diversity of living organisms.	Evolution occurs as the heritable characteristics of populations change across generations and can lead populations to become better adapted to their environment.	**100.00**	**100.00**
Earth Systems Science	Describe and interpret how Earth's geologic history and place in space are relevant to	The history of the universe, solar system and Earth can be inferred from evidence left from past events.	**100.00**	**50.00**

Standard	Prepared Graduate Competencies	Grade Level Expectations	Percent of Items with DOK At or Above the Level of the EEOs	
			GLE	PGC
	our understanding of the processes that have shaped our planet.	As part of the solar system, Earth interacts with various extraterrestrial forces and energies such as gravity, solar phenomena, electromagnetic radiation, and impact events that influence the planet's geosphere, atmosphere, and biosphere in a variety of ways.	0.00	
	Evaluate evidence that Earth's geosphere, atmosphere, hydrosphere, and biosphere interact as a complex system.	The theory of plate tectonics helps explain geological, physical, and geographical features of Earth.	**100.00**	**93.75**
		Climate is the result of energy transfer among interactions of the atmosphere, hydrosphere, geosphere, and biosphere.	**87.50**	
	Describe how humans are dependent on the diversity of resources provided by Earth and Sun.	There are costs, benefits, and consequences of exploration, development, and consumption of renewable and nonrenewable resources.	**100.00**	**100.00**
	Evaluate evidence that Earth's geosphere, atmosphere, hydrosphere, and biosphere interact as a complex system.	The interaction of Earth's surface with water, air, gravity, and biological activity causes physical and chemical changes	**75.00**	45.83
		Natural hazards have local, national and global impacts such as volcanoes, earthquakes, tsunamis, hurricanes, and thunderstorms	0.00	
	Number of GLEs with item DOK at or above EEO DOK		**18 of 22**	
	Number of PGCs with item DOK at or above EEO DOK		**10 of 11**	

In grade 5, panelists indicated that for all the GLEs, at least 50% of items assess students at the appropriate level of cognitive complexity. In grade 8, 80% of the GLEs met this criterion. At the high school level, 82% of the GLEs and 91% of PGCs met Webb's (1997) DOK criterion.

Range of Knowledge Correspondence

The *range-of-knowledge correspondence* measure examines in greater detail the breadth of knowledge covered by the assessment. In addition to evaluating which grade level expectations are assessed, we must look at how many of the EEOs within a GLE are represented by items. The EEOs should be linked with at least one item. Webb's (1997) minimum level of acceptability for range-of-knowledge correspondence is that at least 50% of EEOs per GLE link with items. Tables 3.10 through 3.12 summarize the range-of-knowledge results for each grade level COALT science test per GLE. The strands that meet Webb's indicator criterion are in bold.

Table 3.10. Summary of Range-of-Knowledge Results for the Science CoAlt – Grade 5

Standard	Grade Level Expectations	Percent of EEOs per GLE Matched to at Least One Item
Physical Science	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.	**100.00**
Life Science	All organisms have structures and systems with separate functions.	**100.00**
	Human body systems have basic structures, functions, and needs.	**100.00**
Earth Systems Science	Earth and Sun provide a diversity of renewable and nonrenewable resources.	**67.67**
	Earth's surface changes constantly through a variety of processes and forces.	**66.67**
	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.	**93.33**
	Number of GLEs Assessed Adequately	**6 of 6**

Table 3.11. Summary of Range-of-Knowledge Results for the Science CoAlt – Grade 8

Standard	Grade Level Expectations	Percent of EEOs per GLE Matched to at Least One Item
Physical Science	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.	0.00
	There are different forms of energy, and those forms of energy can be changed from one form to another – but total energy is conserved.	**100.00**
	Distinguish between physical and chemical changes, noting that mass is conserved during any change.	**80.00**
	Recognize that waves such as electromagnetic, sound, seismic, and water have common characteristics and unique properties.	**80.00**
Life Science	Human activities can deliberately or inadvertently alter ecosystems and their resiliency.	**100.00**
	Organisms reproduce and transmit genetic information (genes) to offspring, which influences individuals' traits in the next generation.	**100.00**
Earth Systems Science	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.	**66.67**
	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.	**100.00**
	The solar system is comprised of various objects that orbit the Sun and are classified based on their characteristics.	**75.00**
	The relative positions and motions of Earth, Moon, and Sun can be used to explain observable effects such as seasons, eclipses, and Moon phases.	**56.00**
	Number of GLEs Assessed Adequately	**9 of 10**

Table 3.12. Summary of Range-of-Knowledge Results for the Science CoAlt – High School

Standard	Prepared Graduate Competencies	Grade Level Expectations	Percent of EOs per GLE Matched to at Least One Item	
			GLE	PGC
Physical Science	Observe, explain, and predict natural phenomena governed by Newton's laws of motion, acknowledging the limitations of their application to very small or very fast objects.	Newton's laws of motion and gravitation describe the relationships among forces acting on and between objects, their masses, and changes in their motion – but have limitations.	33.33	33.33
	Apply an understanding of atomic and molecular structure to explain the properties of matter, and predict outcomes of chemical and nuclear reactions.	Matter has definite structure that determines characteristic physical and chemical properties.	**50.00**	**62.5**
		Matter can change form through chemical or nuclear reactions abiding by the laws of conservation of mass and energy.	**100.00**	
		Atoms bond in different ways to form molecules and compounds that have definite properties.	**62.5**	
	Apply an understanding that energy exists in various forms, and its transformation and conservation occur in processes that are predictable and measurable.	Energy exists in many forms such as mechanical, chemical, electrical, radiant, thermal, and nuclear, that can be quantified and experimentally determined.	33.33	40.00
		When energy changes form, it is neither created not destroyed; however, because some is necessarily lost as heat, the amount of energy available to do work decreases.	**50.00**	
Life Science	Explain and illustrate with examples how living systems interact with the biotic and abiotic environment.	Matter tends to be cycled within an ecosystem, while energy is transformed and eventually exits an ecosystem.	**100.00**	33.33
		The size and persistence of populations depend on their interactions with each other and on the abiotic factors in an ecosystem.	0.00	
	Analyze the relationship between structure and function in living systems at a variety of organizational levels, and recognize living systems' dependence on natural selection.	Cellular metabolic activities are carried out by biomolecules produced by organisms.	**58.33**	**55.56**
		The energy for life primarily derives from the interrelated processes of photosynthesis and cellular respiration. Photosynthesis transforms the sun's light energy into the chemical energy of molecular bonds. Cellular respiration allows cells to utilize chemical energy when these bonds are broken.	**50.00**	
		Cells use passive and active transport of substances across membranes to maintain relatively stable intracellular environments.	**87.50**	
		Cells, tissues, organs, and organ systems maintain relatively stable internal environments, even in the face of changing external environments.	**50.00**	

Standard	Prepared Graduate Competencies	Grade Level Expectations	Percent of EOs per GLE Matched to at Least One Item	
			GLE	PGC
	Analyze how various organisms grow, develop, and differentiate during their lifetimes based on an interplay between genetics and their environment.	Physical and behavioral characteristics of an organism are influenced to varying degrees by heritable genes, many of which encode instructions for the production of proteins.	**50.00**	**66.67**
		Multicellularity makes possible a division of labor at the cellular level through the expression of select genes, but not the entire genome.	**100.00**	
	Explain how biological evolution accounts for the unity and diversity of living organisms.	Evolution occurs as the heritable characteristics of populations change across generations and can lead populations to become better adapted to their environment.	**100.00**	**100.00**
Earth Systems Science	Describe and interpret how Earth's geologic history and place in space are relevant to our understanding of the processes that have shaped our planet.	The history of the universe, solar system and Earth can be inferred from evidence left from past events.	**50.00**	**50.00**
		As part of the solar system, Earth interacts with various extraterrestrial forces and energies such as gravity, solar phenomena, electromagnetic radiation, and impact events that influence the planet's geosphere, atmosphere, and biosphere in a variety of ways.	**50.00**	
	Evaluate evidence that Earth's geosphere, atmosphere, hydrosphere, and biosphere interact as a complex system.	The theory of plate tectonics helps explain geological, physical, and geographical features of Earth.	**50.00**	41.67
		Climate is the result of energy transfer among interactions of the atmosphere, hydrosphere, geosphere, and biosphere.	37.50	
	Describe how humans are dependent on the diversity of resources provided by Earth and Sun.	There are costs, benefits, and consequences of exploration, development, and consumption of renewable and nonrenewable resources.	**50.00**	**50.00**
	Evaluate evidence that Earth's geosphere, atmosphere, hydrosphere, and biosphere interact as a complex system.	The interaction of Earth's surface with water, air, gravity, and biological activity causes physical and chemical changes	**75.00**	**62.5**
		Natural hazards have local, national and global impacts such as volcanoes, earthquakes, tsunamis, hurricanes, and thunderstorms	**50.00**	
		Number of GLEs Assessed Adequately	**18 of 22**	
		Number of PGCs Assessed Adequately	**7 of 11**	

The grade 5 and grade 8 science tests met the minimum range-of-knowledge criterion for 90% and 100% of the GLEs, respectively. At the high school level, this criterion was met for 82% of GLEs and 64% of PGCs. This difference is a direct result of the larger number of EEOs available to be assessed at the high school level compared to the other grades. Tables A-7 through A-9 in Appendix A contain the means and standard deviations for each GLE and the number of assessable EEOs per GLE.

Balance-of-Knowledge Representation

The fourth measure of alignment included in the Webb (1997) method is *balance-of-knowledge representation*. This measure describes the distribution of items linked to each EEO within each GLE. The number of items should be distributed rather evenly between the EEOs to achieve good balance.

The content balance is determined by calculating an index, or score, for each GLE[1]. According to Webb (1997), the minimum acceptable index for a single strand is 70 (on a scale of 0 to 100 with 100 representing perfect balance). An index of 70 or higher suggests that items broadly assess the EEOs for a GLE instead of clustering around a subset of EEOs.

Two cautions should be noted regarding the balance index when interpreting the results. First, only those EEOs actually matched to items by the panelists are included in calculations of the balance index. A given GLE may include more EEOs than are actually linked to items by panelists. For example, if a particular GLE includes four EEOs in the standards document but panelists found items matching to just three EEOs, only these three EEOs are evaluated for item distribution. Recognizing this feature of the balance index is important in cases when the range measure and balance measure produce seemingly contrasting results.

Tables 3.13 through 3.15 summarize the results on balance-of-content representation per grade for the CoAlt science tests. All of the grades assessed surpassed the minimum level of acceptability (index of 70) for demonstrating good content balance among those EEOs matched to items for each GLE. The GLEs that meet Webb's (1997) indicator criterion are in bold. Tables A-10 through A-12 contain means associated with the calculation of the balance index.

Table 3.13. Summary of Balance-of-Knowledge Representation Results Science CoAlt – Grade 5

Standard	Grade Level Expectations	Balance Index
Physical Science	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.	**83.33**
Life Science	All organisms have structures and systems with separate functions.	**83.33**
	Human body systems have basic structures, functions, and needs.	**83.33**
Earth Systems Science	Earth and Sun provide a diversity of renewable and nonrenewable resources.	**100.00**
	Earth's surface changes constantly through a variety of processes and forces.	**100.00**
	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.	**96.67**
	Number of GLEs Assessed Adequately	**6 of 6**

Table 3.14. Summary of Balance-of-Knowledge Representation Results Science CoAlt – Grade 8

Standard	Grade Level Expectations	Balance Index

[1] The exact formula for calculating the balance index is explained in detail in Webb's (2005) alignment training manual: http://www.wcer.wisc.edu/WAT/index.aspx.

Physical Science	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.	0.00
	There are different forms of energy, and those forms of energy can be changed from one form to another – but total energy is conserved.	**100.00**
	Distinguish between physical and chemical changes, noting that mass is conserved during any change.	**100.00**
	Recognize that waves such as electromagnetic, sound, seismic, and water have common characteristics and unique properties.	**100.00**
Life Science	Human activities can deliberately or inadvertently alter ecosystems and their resiliency.	**100.00**
	Organisms reproduce and transmit genetic information (genes) to offspring, which influences individuals' traits in the next generation.	**86.00**
Earth Systems Science	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.	**100.00**
	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.	**83.33**
	The solar system is comprised of various objects that orbit the Sun and are classified based on their characteristics.	**85.67**
	The relative positions and motions of Earth, Moon, and Sun can be used to explain observable effects such as seasons, eclipses, and Moon phases.	**100.00**
	Number of GLEs Assessed Adequately	**9 of 10**

Table 3.15. Summary of Balance-of-Knowledge Representation Results Science CoAlt – High School

Standard	Prepared Graduate Competencies	Grade Level Expectations	Balance Index	
			GLE	PGC
Physical Science	Observe, explain, and predict natural phenomena governed by Newton's laws of motion, acknowledging the limitations of their application to very small or very fast objects.	Newton's laws of motion and gravitation describe the relationships among forces acting on and between objects, their masses, and changes in their motion – but have limitations.	**100.00**	**100.00**
	Apply an understanding of atomic and molecular structure to explain the properties of matter, and predict outcomes of chemical and nuclear reactions.	Matter has definite structure that determines characteristic physical and chemical properties.	**100.00**	**95.83**
		Matter can change form through chemical or nuclear reactions abiding by the laws of conservation of mass and energy.	**100.00**	
		Atoms bond in different ways to form molecules and compounds that have definite properties.	**100.00**	
	Apply an understanding that energy exists in various forms, and its transformation and conservation occur in processes that are predictable and measurable.	Energy exists in many forms such as mechanical, chemical, electrical, radiant, thermal, and nuclear, that can be quantified and experimentally determined.	**100.00**	**100.00**
		When energy changes form, it is neither created not destroyed; however, because	**100.00**	

Standard	Prepared Graduate Competencies	Grade Level Expectations	Balance Index	
			GLE	PGC
		some is necessarily lost as heat, the amount of energy available to do work decreases.		
Life Science	Explain and illustrate with examples how living systems interact with the biotic and abiotic environment.	Matter tends to be cycled within an ecosystem, while energy is transformed and eventually exits an ecosystem.	**100.00**	**100.00**
		The size and persistence of populations depend on their interactions with each other and on the abiotic factors in an ecosystem.	0.00	
	Analyze the relationship between structure and function in living systems at a variety of organizational levels, and recognize living systems' dependence on natural selection.	Cellular metabolic activities are carried out by biomolecules produced by organisms.	**91.67**	**89.58**
		The energy for life primarily derives from the interrelated processes of photosynthesis and cellular respiration. Photosynthesis transforms the sun's light energy into the chemical energy of molecular bonds. Cellular respiration allows cells to utilize chemical energy when these bonds are broken.	**100.00**	
		Cells use passive and active transport of substances across membranes to maintain relatively stable intracellular environments.	**100.00**	
		Cells, tissues, organs, and organ systems maintain relatively stable internal environments, even in the face of changing external environments.	**100.00**	
	Analyze how various organisms grow, develop, and differentiate during their lifetimes based on an interplay between genetics and their environment.	Physical and behavioral characteristics of an organism are influenced to varying degrees by heritable genes, many of which encode instructions for the production of proteins.	**100.00**	**95.83**
		Multicellularity makes possible a division of labor at the cellular level through the expression of select genes, but not the entire genome.	**100.00**	
	Explain how biological evolution accounts for the unity and diversity of living organisms.	Evolution occurs as the heritable characteristics of populations change across generations and can lead populations to become better adapted to their environment.	**100.00**	**100.00**
Earth Systems Science	Describe and interpret how Earth's geologic history and place in space are relevant to our understanding of the processes that have shaped our planet.	The history of the universe, solar system and Earth can be inferred from evidence left from past events.	**100.00**	**100.00**
		As part of the solar system, Earth interacts with various extraterrestrial forces and energies such as gravity, solar phenomena, electromagnetic radiation, and impact events that influence the planet's geosphere, atmosphere, and biosphere in a variety of ways.	**100.00**	
	Evaluate evidence that Earth's geosphere, atmosphere, hydrosphere,	The theory of plate tectonics helps explain geological, physical, and geographical features of Earth.	**100.00**	**83.33**

Standard	Prepared Graduate Competencies	Grade Level Expectations	Balance Index	
			GLE	PGC
	and biosphere interact as a complex system.	Climate is the result of energy transfer among interactions of the atmosphere, hydrosphere, geosphere, and biosphere.	**100.00**	
	Describe how humans are dependent on the diversity of resources provided by Earth and Sun.	There are costs, benefits, and consequences of exploration, development, and consumption of renewable and nonrenewable resources.	**100.00**	**100.00**
	Evaluate evidence that Earth's geosphere, atmosphere, hydrosphere, and biosphere interact as a complex system.	The interaction of Earth's surface with water, air, gravity, and biological activity causes physical and chemical changes	**100.00**	**100.00**
		Natural hazards have local, national and global impacts such as volcanoes, earthquakes, tsunamis, hurricanes, and thunderstorms	**100.00**	
		Number of GLEs Assessed Adequately	**21 of 22**	
		Number of PGCs Assessed Adequately	**11 of 11**	

Summary and Discussion on Webb Alignment Indicators

The overall alignment results provide generally positive support for the content validity of the CoAlt science tests. Summary alignment judgments are based on Webb (2005). These summary judgments focus on the percentage of content GLEs represented well by the assessment. Webb outlined a scale with a range of potential alignment outcomes applied to each of the four indicators:

- Fully aligned – assessments align to all GLEs (91%–100%),
- Highly aligned – assessments align to the majority of GLEs (70%–90%),
- Partially aligned – assessments align well to some GLEs (50%–69%),
- Weakly aligned – assessments align to less than half the GLEs (below 50%).

Webb's (1997) alignment method does not allow for a *single* judgment of overall alignment across the four alignment indicators. However, one can get a sense of overall alignment between the assessments and standards by looking at all of the alignment indicators together.

Table 3.16 presents the summary alignment outcomes for the CoAlt science tests based on the above scale. The table includes a summary judgment for each Webb alignment indicator per grade level based on the percentage of GLEs that met the minimum alignment criteria. This summary table is linked to the bottom row of Tables A-1 through A-12 in Appendix A. Thus, these summary judgments reflect a final evaluation of each grade assessment per Webb indicator criteria *across* the GLEs.

As shown in Table 3.16 with green highlighting, roughly 92% of the results indicate strong content alignment of the CoAlt science test to the Colorado Extended Evidence Outcomes. Each of the three grade level tests includes a sufficiently even distribution of extended evidence outcomes within the associated grade level expectation and sufficient coverage of the range of extended evidence outcomes within GLEs. The high school and grade 8 science tests include sufficient numbers of items to cover the Colorado Extended Evidence Outcomes at the

Standard level. The three grade level tests also include sufficient numbers of items at DOK levels at or above the DOK assigned to the corresponding EEOs.

Additional analyses by HumRRO found that panelists indicated that the CoAlt items reflect the intended content of the test blueprints, and that the large majority of items are highly aligned to the particular extended evidence outcomes to which they were matched.

Table 3.16. Summary Alignment Outcomes on Each Webb Criterion by Grade Level for Science CoAlt

Grade Level	Percentage of GLEs that Met Webb Criteria				
	Categorical Concurrence	Depth-of-Knowledge Consistency	Range-of-Knowledge Correspondence		Balance-of-Knowledge Representation
5	Partially aligned (67%)	Fully aligned (100%)	Fully aligned (100%)		Fully aligned (100%)
8	Fully aligned (100%)	Highly aligned (80%)	Highly aligned (90%)		Highly aligned (90%)
High School	Fully aligned (100%)	Highly/fully aligned (82%; 91%)	Highly aligned (90%)	Partially aligned (64%)	Fully aligned (95%; 100%)

Note. Categorical concurrence is evaluated at the Standard level to reflect score reporting practices. High school criteria with multiple percentages reflect GLEs and PGCs, respectively.

Tables A-13 through A-15 in Appendix A present the mean number of items matched to each EEO and the number of panelists represented.

Suggestions for improving the alignment between the CoAlt science tests and Colorado Extended Evidence Outcomes are discussed in Chapter 5, Summary and Recommendations.

Chapter 4: Results: Social Studies Content Alignment

The content alignment evaluation analyses discussed in this chapter are based on panelists' ratings of the CoAlt social studies items for grades 4 and 7 and high school.

Reliability Results

In this section, we report on the comparison of panelists' ratings of content match to the item bank's documented content match. In other words, do panelists assign the same EEO to an item as the item writer during item development?

Panelist-Test Developer Analyses

This analysis examined the agreement outcomes between the EEO assigned to an item by panelists, and the EEO assigned to an item as noted in the item bank. Table 4.1 presents the agreement outcomes between panelists and the item bank on the content assessed by items. Agreement was analyzed at several levels of specificity. All of the items were analyzed first for 'exact Match', which indicates that panelists chose the same EEO. If panelists did not show an exact match with the item bank, we determined the percent agreement at the Grade Level Expectation (GLE) level. For high school, we also determine the percent agreement at the Prepared Graduate Competency (PGC) level. Finally, we determined the percent agreement at the standard level (i.e., history, geography, economics, and civics).The last column in Table 4.1 shows the percentage of ratings by panelists that did not match the item bank coding at all on items.

Table 4.1. Percent Agreement between Panelists and Item Bank on Target Content

Grade	Total Number of Panelist Ratings across Items	Percent Agreement with Item Bank Codes				
		Exact Match	GLE Match	PGC Match	Standard Match	No Match
4	51	82.4%	82.4%	NA	84.3%	15.7%
7	85	88.2%	96.5%	NA	100.0%	0.0%
High School	130	84.6%	90.8%	93.1%	94.6%	5.4%

As Table 4.1 indicates, panelists were highly consistent with the item bank in identifying the content codes of items. Panelists identified an exact match for 82–85% of the ratings and a match at the GLE level or below for 82–97% of the ratings. Panelists differed completely from the item bank on content match for 0–16% of the ratings. Overall these findings suggest that the majority of social studies items do measure the intended content.

Webb Alignment Results

In this section, we review the general outcomes of item analyses on the four Webb (1997) alignment indicators.

All of Webb's measures begin with calculations for each panelist and build up to a summary of results across panelists per EEO. First, we calculated the mean ratings across items for each panelist, and then we determined the mean rating across panelists per EEO. Depending on the component under review, results are presented at the broader GLE and Standard levels (as well

as the PGC level for high school). Results at the more specific EEO level are presented in Appendix B.

Categorical Concurrence

Categorical concurrence describes the extent to which the CoAlt items, regardless of item type and point value, cover the content grade level expectations of the Colorado Extended Evidence Outcomes. Webb (1997) recommends a minimum of six test questions to adequately assess each grade level expectation. This criterion serves as a guideline for reasonable content coverage based on earlier research on the reliability of tests compared to the number of items (Subkoviak, 1988). Tables 4.2 through 4.4 summarize the CoAlt alignment results for categorical concurrence for each grade level. The GLEs, PGCs, and Standards that meet Webb's indicator criterion are in bold. Tables B-1 through B-3 in Appendix B also contain the standard deviations for each grade level expectation.

Table 4.2. Summary of Categorical Concurrence Results for Social Studies CoAlt – Grade 4

Standard	Grade Level Expectation	Mean Number of Items per GLE	Mean Number of Items per Standard
History	Organize and sequence events to understand the concepts of chronology and cause and effect in the history of Colorado.	1.50	2.00
	The historical eras, individuals, groups, ideas and themes in Colorado history and their relationships to key events in the United States.	1.00	
Geography	Use several types of geographic tools to answer questions about the geography of Colorado.	2.00	4.33
	Connections within and across human and physical systems are developed.	2.33	
Economics	People respond to positive and negative incentives.	3.33	**6.33**
	The relationship between choice and opportunity cost (PFL).	3.00	
Civics	Analyze and debate multiple perspectives on an issue.	2.00	4.00
	The origins, structure, and functions of the Colorado government	2.00	
	GLEs with at Least Six Items	**0 of 8**	
	Standards with at Least Six Items	**1 of 4**	

Table 4.3. Summary of Categorical Concurrence Results for Social Studies CoAlt – Grade 7

Standard	Grade Level Expectation	Mean Number of Items per GLE	Mean Number of Items per Standard
History	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.	2.80	**5.00**
	The historical eras, individuals, groups, ideas and themes within regions of the Eastern Hemisphere and their relationships with one another	2.20	
Geography	Use geographic tools to gather data and make geographic inferences and predictions.	2.00	**4.00**
	Regions have different issues and perspectives.	2.00	
Economics	Supply and demand influence price and profit in a market economy.	2.00	**3.00**
	The distribution of resources influences economic production and individual choices (Economics and PFL).	1.00	
Civics	Compare how various nations define the rights, responsibilities, and roles of citizens.	3.40	**5.00**
	Different forms of government and international organizations and their influence in the world community.	2.00	
	GLEs with at Least Six Items	**0 of 8**	
	Standards with at Least Six Items	**0 of 4**	

Table 4.4. Summary of Categorical Concurrence Results for Social Studies CoAlt – High School

Standard	Prepared Graduate Competency	Grade Level Expectations	Mean *N* of Items GLE	Mean *N* of Items PGC	Mean *N* of Items Standard
History	Develop an understanding of how people view, construct, and interpret history.	Use the historical method of inquiry to ask questions, evaluate primary and secondary sources, critically analyze and interpret data, and develop interpretations defended by evidence.	2.00	2.00	**6.60**
	Analyze key historical periods and patterns of change over time within and across nations and cultures	The key concepts of continuity and change, cause and effect, complexity, unity and diversity over time.	3.60	4.60	
		The significance of ideas as powerful forces throughout history.	1.00		
Geography	Develop spatial understanding, perspectives, and personal connections to the world	Use different types of maps and geographic tools to analyze features on Earth to investigate and solve geographic questions.	3.20	5.20	**7.00**
		Explain and interpret geographic variables that influence the interactions of people, places and environments.	2.00		

Standard	Prepared Graduate Competency	Grade Level Expectations	Mean *N* of Items GLE	Mean *N* of Items PGC	Mean *N* of Items Standard
	Examine places and regions and the connections among them	The interconnected nature of the world, its people and places.	1.80	1.80	
Economics	Understand the allocation of scarce resources in societies through analysis of individual choice, market interaction, and public policy.	Productive resources – natural, human, capital – are scarce; therefore, choices are made about how individuals, businesses, governments, and societies allocate these resources.	1.00	3.20	**7.20**
		Economic policies affect markets.	1.20		
		Government and competition affect markets.	1.00		
	Acquire the knowledge and economic reasoning skills to make sound financial decisions (PFL).	Design, analyze, and apply a financial plan based on short- and long-term financial goals (PFL).	1.00	4.00	
		Analyze strategic spending, saving, and investment options to achieve the objectives of diversification, liquidity, income, and growth (PFL).	1.00		
		The components of personal credit to manage credit and debt (PFL).	1.00		
		Identify, develop, and evaluate risk-management strategies (PFL).	1.00		
Civics	Analyze and practice rights, roles, and responsibilities of citizens.	Research, formulate positions, and engage in appropriate civic participation to address local, state, and national issues or policies.	1.50	1.50	5.20
	Analyze origins, structure, and functions of governments and their impacts on societies and citizens.	Purposes of and limitations on the foundations, structures and functions of government.	2.20	4.00	
		Analyze how public policy - domestic and foreign - is developed at the local, state, and national levels and compare how policy-making occurs in other forms of government.	1.80		
		GLEs with at Least Six Items	**0 of 16**		
		PGCs with at Least Six Items	**0 of 8**		
		Standards with at Least Six Items	**3 of 4**		

As Tables 4.2 through 4.4 indicate, none of the social studies tests include a sufficient number of items to meet the minimum requirements for categorical concurrence on any social studies GLEs. It is important to note that these results are in large part shaped by the number of items reviewed. In high school, for example, in order for there to be a minimum of 6 items representing each of the 16 GLEs, a minimum of 96 items would need to be included. Tables 4.2 through 4.4 show that each GLE was linked, on average, to at least 1 item. If we group the content at the standard level only, the minimum content requirements are met for grade 4 economics and high school history, geography, and economics.

Because of the limitations inherent in the Webb (1997) criteria due to the minimum item requirements, it is helpful to consider how well the panelists' ratings match the distribution of content as outlined in the test specifications. Table 4.5 presents a comparison of the average number of items matched to each standard as compared to the number of items per standard outlined in the CoAlt test blueprints. Table 4.5 shows that the CoAlt social studies tests generally reflect the intended content. Grade 4 panelists did identify fewer history items and more economics items than indicated in the test blueprint.

Table 4.5. Comparison of Panelist Ratings with Test Blueprints

Standard	Number of Items per Standard					
	Grade 4		Grade 7		High School	
	Panelists	Blueprint	Panelists	Blueprint	Panelists	Blueprint
History	2	4	5	5	6.6	7
Geography	4.3	5	4	4	7	7
Economics	6.3	4	3	3	7.2	7
Civics	4	4	5	5	5.2	5

In addition to identifying the content assessed by each item, we asked panelists to indicate *how well* the item assessed the content. Panelists subjectively rated the extent of item alignment to the content on a 4-point scale ranging from 'not aligned to any EEO' to 'fully aligned'. Table 4.6 presents the mean number of items (across panelists) at each level of alignment. For each grade level, panelists rated items as well aligned to the EEO matched to that item.

Table 4.6. Panelist Ratings on Overall Item Alignment

Grade (N items)	Degree of Alignment	Mean Number of Items per Level	*SD*	Percent of Items per Level
4 (N=17)	Not at all aligned	0.00	0.00	0.00
	Weakly aligned	1.00	0.00	3.92
	Highly aligned	2.50	0.71	9.80
	Fully aligned	14.67	2.08	86.27
7 (N=17)	Not at all aligned	0.00	0.00	0.00
	Weakly aligned	2.00	1.73	7.06
	Highly aligned	5.40	2.70	31.76
	Fully aligned	10.40	2.70	61.18
High School (N=26)	Not at all aligned	0.00	0.00	0.00
	Weakly aligned	2.00	0.00	1.54
	Highly aligned	11.25	10.40	34.62
	Fully aligned	20.75	4.03	63.85

In general, panelists across the three grade levels rated at least 93% of the items as being 'highly aligned' or 'fully aligned'. The grade 7 assessment had the highest percentage of items rated by panelists as being 'Weakly aligned' or 'Not at all aligned' at 7%. No items were rated as 'not aligned' to an EEO.

Depth-of-Knowledge Consistency

Analyses of depth-of-knowledge (DOK) measure the type of cognitive processing required of students by content standards. The DOK requirements implied by the EEOs should be matched by assessment items. To confirm this match, panelists were asked to rate the EEOs and the social studies items separately. Webb (1997) includes an alignment indicator that directly compares panelists' DOK ratings of content standards and test items, which he refers to as *depth-of-knowledge consistency*.

To make their ratings of the extended content standards and test items, panelists used a modified version of a cognitive complexity rating scale developed for evaluating the depth of knowledge of alternate assessments (see Flowers, Wakeman, Browder, & Karvonen, 2007). However, during analysis, panelists' DOK ratings of the extended standards and items were collapsed into a three-point classification scheme to better correspond with the DOK guidance for the alternate standards provided in the test blueprints and with the DOK classification scheme used in the item bank.

The rating categories assigned by panelists and the recoded DOK value (in parentheses) included:

- Level 0 None: No content clearly measured; too vague. (*not assigned by any panelist*)
- Level 1 Attention: Requires students to display ability to acknowledge, reply, and respond to text or related subject features. (*recoded Level 1: Recall and Reproduction*)
- Level 2 Memorize/recall: Requires the ability to recite or recall facts or information. It involves the ability to distinguish between simple text-based and one-step procedures. (*recoded Level 1: Recall and Reproduction*)
- Level 3 Performance: Requires students to use recalled facts or information for simple tasks. (*recoded Level 1: Recall and Reproduction*)
- Level 4 Comprehension: Requires processing beyond recall and observation and may require both understanding and subsequent processing of text. It involves ordering, classifying, estimating text or numbers as well as identifying patterns, main points, or two-step procedures. (*recoded Level 2: Skills and Concepts*)
- Level 5 Application: Show ability to go beyond text; to reason, plan, or use of evidence to connect ideas. Students will use text, data, or observations to draw conclusions or solve non-routine problems. (*recoded Level 2: Skills and Concepts*)
- Level 6 Analysis, Synthesis, Evaluation: Requires extended higher order processing. It typically requires extended time to complete a task, but the time is not spent on repetitive tasks. It involves taking information and applying this information to a new task; which may require generating a hypothesis, perform complex analyses, or make connections among different texts. (*recoded Level 3: Strategic Thinking and Reasoning*)

Tables 4.7 through 4.9 summarize the depth-of-knowledge consistency results for each grade level of the CoAlt social studies test. Because panelists evaluated depth of knowledge at the most specific level of the standards document (EEOs), the table refers to consistency between the items and the EEOs to which they were matched. Results are summarized at the GLE level for ease of presentation. Tables B-4 through B-6 in Appendix B contain the means and standard deviations for DOK ratings at all levels.

Webb's (1997) suggested criterion for this alignment indicator is that at least 50% of the items should have complexity ratings at or above the level of the corresponding EEO. The percentages on strands that reach the 50% criterion are bolded.

Table 4.7. Summary of Depth-of-Knowledge Results for Social Studies CoAlt – Grade 4

Standard	Grade Level Expectations	Percent of Items with DOK At or Above the Level of the EEOs
History	Organize and sequence events to understand the concepts of chronology and cause and effect in the history of Colorado.	**100.00**
	The historical eras, individuals, groups, ideas and themes in Colorado history and their relationships to key events in the United States.	**100.00**
Geography	Use several types of geographic tools to answer questions about the geography of Colorado.	**100.00**
	Connections within and across human and physical systems are developed.	**100.00**
Economics	People respond to positive and negative incentives.	30.56
	The relationship between choice and opportunity cost (PFL).	**88.89**
Civics	Analyze and debate multiple perspectives on an issue.	**100.00**
	The origins, structure, and functions of the Colorado government	33.33
	Number of GLEs with item DOK at or above EEO DOK	**6 of 8**

Table 4.8. Summary of Depth-of-Knowledge Results for Social Studies CoAlt – Grade 7

Standard	Grade Level Expectations	Percent of Items with DOK At or Above the Level of the EEOs
History	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.	6.67
	The historical eras, individuals, groups, ideas and themes within regions of the Eastern Hemisphere and their relationships with one another.	**100.00**
Geography	Use geographic tools to gather data and make geographic inferences and predictions.	40.00
	Regions have different issues and perspectives.	**100.00**
Economics	Supply and demand influence price and profit in a market economy.	**100.00**
	The distribution of resources influences economic production and individual choices (Economics and PFL).	**60.00**
Civics	Compare how various nations define the rights, responsibilities, and roles of citizens.	**86.67**
	Different forms of government and international organizations and their influence in the world community.	25.00
	Number of GLEs with item DOK at or above EEO DOK	**5 of 8**

Table 4.9. Summary of Depth-of-Knowledge Results for Social Studies CoAlt – High School

Standard	Prepared Graduate Competency	Grade Level Expectation	Percent of Items with DOK At or Above the Level of the EEOs	
			GLE	PGC
History	Develop an understanding of how people view, construct, and interpret history	Use the historical method of inquiry to ask questions, evaluate primary and secondary sources, critically analyze and interpret data, and develop interpretations defended by evidence.	30.00	30.00
	Analyze key historical periods and patterns of change over time within and across nations and cultures	The key concepts of continuity and change, cause and effect, complexity, unity and diversity over time.	30.00	46.00
		The significance of ideas as powerful forces throughout history.	**100.00**	
Geography	Develop spatial understanding, perspectives, and personal connections to the world	Use different types of maps and geographic tools to analyze features on Earth to investigate and solve geographic questions.	31.67	**58.00**
		Explain and interpret geographic variables that influence the interactions of people, places and environments.	**100.00**	
	Examine places and regions and the connections among them	The interconnected nature of the world, its people and places.	**100.00**	**100.00**
Economics	Understand the allocation of scarce resources in societies through analysis of individual choice, market interaction, and public policy	Productive resources – natural, human, capital – are scarce; therefore, choices are made about how individuals, businesses, governments, and societies allocate these resources.	**100.00**	**100.00**
		Economic policies affect markets.	**100.00**	
		Government and competition affect markets.	**100.00**	
	Acquire the knowledge and economic reasoning skills to make sound financial decisions (PFL)	Design, analyze, and apply a financial plan based on short- and long-term financial goals (PFL).	40.00	**70.00**
		Analyze strategic spending, saving, and investment options to achieve the objectives of diversification, liquidity, income, and growth (PFL).	**80.00**	
		The components of personal credit to manage credit and debt (PFL).	**60.00**	
		Identify, develop, and evaluate risk-management strategies (PFL).	**100.00**	
Civics	Analyze and practice rights, roles, and responsibilities of citizens	Research, formulate positions, and engage in appropriate civic participation to address local, state, and national issues or policies.	**100.00**	**100.00**
	Analyze origins, structure, and functions of governments and their impacts on societies and citizens	Purposes of and limitations on the foundations, structures and functions of government.	**100.00**	**90.00**
		Analyze how public policy - domestic and foreign - is developed at the local, state, and national levels and compare how policy-making occurs in other forms of government.	**80.00**	
	Number of GLEs with item DOK at or above EEO DOK		**12 of 16**	
	Number of PGCs with item DOK at or above EEO DOK		**6 of 8**	

In grade 4, panelists' ratings using Webb (1997) DOK levels indicate that items on 75% of the grade level expectations assess students at the appropriate cognitive complexity. At grade 7, 63% of the grade level expectations met the Webb criterion. At the high school level, 75% of the GLEs and PGCs met Webb's DOK criterion.

Range of Knowledge Correspondence

The *range-of-knowledge correspondence* measure examines in greater detail the breadth of knowledge covered by the assessment. In addition to evaluating which content strands are assessed, we must look at how many of the EEOs within a GLE are represented by items. The EEOs should be linked with at least one item. Webb's (1997) minimum level of acceptability for range-of-knowledge correspondence is that at least 50% of EOs per GLE link with items. Tables 4.10 through 4.12 summarize the range-of-knowledge results for each grade level CoAlt social studies test per GLE. The GLEs that meet Webb's indicator criterion are in bold.

Table 4.10. Summary of Range-of-Knowledge Results for the Social Studies CoAlt – Grade 4

Standard	Grade Level Expectations	Percent of EEOs per GLE Matched to at Least One Item
History	Organize and sequence events to understand the concepts of chronology and cause and effect in the history of Colorado.	**100.00**
	The historical eras, individuals, groups, ideas and themes in Colorado history and their relationships to key events in the United States.	**100.00**
Geography	Use several types of geographic tools to answer questions about the geography of Colorado.	**50.00**
	Connections within and across human and physical systems are developed.	**83.33**
Economics	People respond to positive and negative incentives.	**100.00**
	The relationship between choice and opportunity cost (PFL).	**100.00**
Civics	Analyze and debate multiple perspectives on an issue.	**100.00**
	The origins, structure, and functions of the Colorado government.	**100.00**
	Number of GLEs Assessed Adequately	**8 of 8**

Table 4.11. Summary of Range-of-Knowledge Results for the Social Studies CoAlt– Grade 7

Standard	Grade Level Expectations	Percent of EEOs per GLE Matched to at Least One Item
History	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.	**80.00**
	The historical eras, individuals, groups, ideas and themes within regions of the Eastern Hemisphere and their relationships with one another.	**100.00**
Geography	Use geographic tools to gather data and make geographic inferences and predictions.	**100.00**
	Regions have different issues and perspectives.	**100.00**
Economics	Supply and demand influence price and profit in a market economy.	**90.00**
	The distribution of resources influences economic production and individual choices (Economics and PFL).	**50.00**
Civics	Compare how various nations define the rights, responsibilities, and roles of citizens.	**90.00**
	Different forms of government and international organizations and their influence in the world community.	**100.00**
	Number of GLEs Assessed Adequately	**8 of 8**

Table 4.12. Summary of Range-of-Knowledge Results for the Social Studies CoAlt – High School

Standard	Prepared Graduate Competencies	Grade Level Expectations	Percent of EEOs per GLE Matched to at Least One Item	
			GLE	PGC
History	Develop an understanding of how people view, construct, and interpret history.	Use the historical method of inquiry to ask questions, evaluate primary and secondary sources, critically analyze and interpret data, and develop interpretations defended by evidence.	**100.00**	**100.00**
	Analyze key historical periods and patterns of change over time within and across nations and cultures	The key concepts of continuity and change, cause and effect, complexity, unity and diversity over time.	**73.33**	**80.00**
		The significance of ideas as powerful forces throughout history.	**100.00**	
Geography	Develop spatial understanding, perspectives, and personal connections to the world	Use different types of maps and geographic tools to analyze features on Earth to investigate and solve geographic questions.	**100.00**	**100.00**
		Explain and interpret geographic variables that influence the interactions of people, places and environments.	**100.00**	
	Examine places and regions and the connections among them	The interconnected nature of the world, its people and places.	**100.00**	**100.00**

Standard	Prepared Graduate Competencies	Grade Level Expectations	Percent of EEOs per GLE Matched to at Least One Item	
			GLE	PGC
Economics	Understand the allocation of scarce resources in societies through analysis of individual choice, market interaction, and public policy.	Productive resources – natural, human, capital – are scarce; therefore, choices are made about how individuals, businesses, governments, and societies allocate these resources.	**100.00**	**100.00**
		Economic policies affect markets.	**100.00**	
		Government and competition affect markets.	**100.00**	
	Acquire the knowledge and economic reasoning skills to make sound financial decisions (PFL).	Design, analyze, and apply a financial plan based on short- and long-term financial goals (PFL).	**100.00**	**100.00**
		Analyze strategic spending, saving, and investment options to achieve the objectives of diversification, liquidity, income, and growth (PFL).	**100.00**	
		The components of personal credit to manage credit and debt (PFL).	**100.00**	
		Identify, develop, and evaluate risk-management strategies (PFL).	**100.00**	
Civics	Analyze and practice rights, roles, and responsibilities of citizens.	Research, formulate positions, and engage in appropriate civic participation to address local, state, and national issues or policies.	**62.50**	**62.50**
	Analyze origins, structure, and functions of governments and their impacts on societies and citizens.	Purposes of and limitations on the foundations, structures and functions of government.	**100.00**	**80.00**
		Analyze how public policy - domestic and foreign - is developed at the local, state, and national levels and compare how policy-making occurs in other forms of government.	**60.00**	
	Number of GLEs Assessed Adequately		**16 of 16**	
	Number of PGCs Assessed Adequately		**8 of 8**	

The CoAlt social studies tests at all three grade levels met the minimum range-of-knowledge criterion for all of the GLEs. Tables B-7 through B-9 in Appendix B contain the means and standard deviations for each strand and the number of assessable EEOs per GLE.

Balance-of-Knowledge Representation

The fourth measure of alignment included in the Webb (1997) method is *balance-of-knowledge representation*. This measure describes the distribution of items linked to each EEO within each GLE. The number of items should be distributed rather evenly between the EEOs to achieve good balance.

The content balance is determined by calculating an index, or score, for each GLE[2]. According to Webb, the minimum acceptable index for a single strand is 70 (on a scale of 0 to 100 with 100 representing perfect balance). An index of 70 or higher suggests that items broadly assess the EEOs for a GLE instead of clustering around a subset of EEOs.

Two cautions should be noted regarding the balance index when interpreting the results. First, only those EEOs actually matched to items by the panelists are included in calculations of the balance index. A given GLE may include more EEOs than are actually linked to items by panelists. For example, if a particular GLE includes four EEOs in the extended content standards document but panelists found items matching to just three EEOs, only these three EEOs are evaluated for item distribution. Recognizing this feature of the balance index is important in cases when the range measure and balance measure produce seemingly contrasting results.

Tables 4.13 through 4.15 summarize the results on balance-of-content representation per grade for the CoAlt social studies tests. All of the grades assessed surpassed the minimum level of acceptability (index of 70) for demonstrating good content balance among those EEOs matched to items for each GLE, with the exception of two high school GLEs to which no items were matched. The GLEs that meet Webb's (1997) indicator criterion are in bold. Tables B-10 through B-12 contain means associated with the calculation of the balance index.

Table 4.13. Summary of Balance-of-Knowledge Representation Results Social Studies CoAlt – Grade 4

Standard	Grade Level Expectations	Balance Index
History	Organize and sequence events to understand the concepts of chronology and cause and effect in the history of Colorado.	**100.00**
	The historical eras, individuals, groups, ideas and themes in Colorado history and their relationships to key events in the United States.	**100.00**
Geography	Use several types of geographic tools to answer questions about the geography of Colorado.	**100.00**
	Connections within and across human and physical systems are developed.	**94.44**
Economics	People respond to positive and negative incentives.	**100.00**
	The relationship between choice and opportunity cost (PFL).	**100.00**
Civics	Analyze and debate multiple perspectives on an issue.	**100.00**
	The origins, structure, and functions of the Colorado government	**100.00**
	Number of GLEs Assessed Adequately	**8 of 8**

[2] The exact formula for calculating the balance index is explained in detail in Webb's (2005) alignment training manual: http://www.wcer.wisc.edu/WAT/index.aspx.

Table 4.14. Summary of Balance-of-Knowledge Representation Results Social Studies CoAlt – Grade 7

Standard	Grade Level Expectations	Balance Index
History	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.	**90.00**
	The historical eras, individuals, groups, ideas and themes within regions of the Eastern Hemisphere and their relationships with one another	**100.00**
Geography	Use geographic tools to gather data and make geographic inferences and predictions.	**100.00**
	Regions have different issues and perspectives.	**100.00**
Economics	Supply and demand influence price and profit in a market economy.	**100.00**
	The distribution of resources influences economic production and individual choices (Economics and PFL).	**100.00**
Civics	Compare how various nations define the rights, responsibilities, and roles of citizens.	**88.00**
	Different forms of government and international organizations and their influence in the world community.	**100.00**
	Number of GLEs Assessed Adequately	**8 of 8**

Table 4.15. Summary of Balance-of-Knowledge Representation Results Social Studies CoAlt – High School

Standard	Prepared Graduate Competency	Grade Level Expectations	Balance Index	
			GLE	PGC
History	Develop an understanding of how people view, construct, and interpret history.	Use the historical method of inquiry to ask questions, evaluate primary and secondary sources, critically analyze and interpret data, and develop interpretations defended by evidence.	**100.00**	**100.00**
	Analyze key historical periods and patterns of change over time within and across nations and cultures	The key concepts of continuity and change, cause and effect, complexity, unity and diversity over time.	**80.00**	**79.67**
		The significance of ideas as powerful forces throughout history.	**100.00**	
Geography	Develop spatial understanding, perspectives, and personal connections to the world	Use different types of maps and geographic tools to analyze features on Earth to investigate and solve geographic questions.	**100.00**	**88.67**
		Explain and interpret geographic variables that influence the interactions of people, places and environments.	**100.00**	
	Examine places and regions and the connections among them	The interconnected nature of the world, its people and places.	**100.00**	**100.00**
Economics	Understand the allocation of scarce resources in societies through analysis of individual choice, market interaction, and public policy.	Productive resources – natural, human, capital – are scarce; therefore, choices are made about how individuals, businesses, governments, and societies allocate these resources.	**100.00**	**96.67**
		Economic policies affect markets.	**100.00**	

Standard	Prepared Graduate Competency	Grade Level Expectations	Balance Index	
			GLE	PGC
		Government and competition affect markets.	**100.00**	
	Acquire the knowledge and economic reasoning skills to make sound financial decisions (PFL).	Design, analyze, and apply a financial plan based on short- and long-term financial goals (PFL).	**100.00**	**100.00**
		Analyze strategic spending, saving, and investment options to achieve the objectives of diversification, liquidity, income, and growth (PFL).	**100.00**	
		The components of personal credit to manage credit and debt (PFL).	**100.00**	
		Identify, develop, and evaluate risk-management strategies (PFL).	**100.00**	
Civics	Analyze and practice rights, roles, and responsibilities of citizens.	Research, formulate positions, and engage in appropriate civic participation to address local, state, and national issues or policies.	**100.00**	**100.00**
	Analyze origins, structure, and functions of governments and their impacts on societies and citizens.	Purposes of and limitations on the foundations, structures and functions of government.	**96.67**	**86.67**
		Analyze how public policy - domestic and foreign - is developed at the local, state, and national levels and compare how policy-making occurs in other forms of government.	**100.00**	
	Number of GLEs Assessed Adequately		**16 of 16**	
	Number of PGCs Assessed Adequately		**8 of 8**	

Summary and Discussion on Webb Alignment Indicators

The overall alignment results provide generally positive support for the content validity of the CoAlt social studies tests. Summary alignment judgments are based on Webb (2005). These summary judgments focus on the percentage of content GLEs represented well by the assessment. Webb outlined a scale with a range of potential alignment outcomes applied to each of the four indicators:

- Fully aligned – assessments align to all GLEs (91%–100%),
- Highly aligned – assessments align to the majority of GLEs (70%–90%),
- Partially aligned – assessments align well to some GLEs (50%–69%),
- Weakly aligned – assessments align to less than half the GLEs (below 50%).

Webb's (1997) alignment method does not allow for a *single* judgment of overall alignment across the four alignment indicators. However, one can get a sense of overall alignment between the assessments and standards by looking at all of the alignment indicators together.

Table 4.16 presents the summary alignment outcomes for the CoAlt social studies tests based on the above scale. The table includes a summary judgment for each Webb (1997) alignment indicator per grade level based on the percentage of GLEs that met the minimum alignment criteria. This summary table is linked to the bottom row of Tables B-1 through B-12 in

Appendix B. Thus, these summary judgments reflect a final evaluation of each grade assessment per Webb indicator criteria *across* the GLEs.

As shown in Table 4.16 with green highlighting, approximately 75% of the results indicate strong content alignment of the CoAlt social studies test to the Colorado Extended Evidence Outcomes. Each of the three grade level tests includes sufficient coverage of the range of evidence outcomes, and a sufficiently even distribution of evidence outcomes within the associated grade level expectation. There was evidence of high alignment in terms of the numbers of items at DOK levels at or above the DOK assigned to the corresponding EEOs for grade 4 and high school. Only the high school social studies test demonstrated sufficient coverage of the extended content standards at the Standard Level.

Additional analyses by HumRRO found that panelists did indicate that the CoAlt items reflect the intended content of the test blueprints, and that the large majority of items were highly aligned to the particular extended evidence outcomes to which they were matched.

Table 4.16. Summary Alignment Outcomes on Each Webb Criterion by Grade Level for Social studies CoAlt

Grade Level	Percentage of GLEs that Met Webb Criteria			
	Categorical Concurrence	Depth-of-Knowledge Consistency	Range-of-Knowledge Correspondence	Balance-of-Knowledge Representation
4	Weakly aligned (25%)	Highly aligned (75%)	Fully aligned (100%)	Fully aligned (100%)
7	Weakly aligned (0%)	Partially aligned (63%)	Fully aligned (100%)	Fully aligned (100%)
High School	Highly aligned (75%)	Highly aligned (75%)	Fully aligned (100%)	Fully aligned (100%)

Note. High school percentages reflect GLEs and PGCs, respectively.

Tables B-13 and B-15 in Appendix B present the mean number of items matched to each EEO and the number of panelists represented.

Suggestions for improving the alignment between the CoAlt social studies tests and Colorado Extended Evidence Outcomes are discussed in Chapter 5, Summary and Recommendations.

Chapter 5: Summary and Recommendations

HumRRO conducted a review of the CoAlt science and social studies tests to examine the content alignment to the Colorado Extended Evidence Outcomes. Alignment of assessments to the content standards they are designed to measure is a requirement of the No Child Left Behind Act (2002).

The cumulative results provide validity evidence to support that the content of CoAlt science and social studies test items match the intended content as specified in the standards. Expert panelists from both content areas tended to agree that items were measuring the intended grade level expectations, and to rate items as highly aligned to the Colorado Extended Evidence Outcomes.

The number of items included on an operational form, when considered along with the number of prepared graduate competencies, grade level expectations, and extended evidence outcomes included in the extended content standards, provide important context for interpreting the Webb (1997) criteria. For example, it was essentially impossible for the categorical concurrence correspondence to be fully met given the number of items. Even with this limitation, both content areas were rated as highly or fully aligned on at least three quarters of the Webb criteria.

As with most reviews of state assessment systems, these findings point to areas where the alignment between assessments and content standards could be strengthened. For this reason, HumRRO makes the following recommendations to Colorado on ways in which alignment of the CoAlt might be improved:

- ***Review content coverage (categorical concurrence).*** Assessments may not adequately reflect the content that students are expected to know based solely on the number of items on the assessment (not the item type or point value as these are not factors in Webb's (1997) categorical concurrence indicator). From strictly an item count perspective, there are several ways CDE can choose to mitigate this situation such as increase the number of items on the assessment, collapse or otherwise reduce the number of grade level expectations in the extended content standards, or designate some of the grade level expectations for local assessment only. Based on this study, there may not be a sufficient number of items to support standard level scores in grades 4 and 8.
- ***Review grade 4 social studies item metadata***. Comparisons of panelist's ratings to item bank data and to the content specifications in the test blueprint showed notably larger discrepancies for the grade 4 social studies test. It may be useful to conduct an internal review to verify that grade 4 social studies item metadata contain no errors.

References

Flowers, C., Wakeman, S., Browder, D., & Karvonen, M. (2007). *Links for academic learning: An alignment protocol for alternate assessments based on alternate achievement standards.* Charlotte, NC: University of North Carolina at Charlotte. Retrieved from: http://www.naacpartners.org/LAL/documents/NAAC_AlignmentManualVer8_3.pdf

No Child Left Behind Act of 2001. Public Law 107-110.

Subkoviak, M. J. (1988). A practitioner's guide to computation and interpretation of reliability indices for mastery tests. *Journal of Educational Measurement*, *25*(1), 47–55.

U.S. Department of Education. (April, 2004). *Standards and assessments peer review guidance: Information and examples for meeting requirements of the No Child Left Behind Act of 2001*. Washington, DC: U.S. Department of Education, Office of Elementary and Secondary Education. Retrieved from http://www.ed.gov/policy/elsec/guid/saaprguidance.doc

Webb, N. L. (1999). *Alignment of Science and Science standards and assessments in four states. (Research Monograph 18)*. Madison, WI: National Institute for Science Education and Council of Chief State School Officers. (ERIC Document Reproduction Service No. ED440852).

Webb, N. L. (1997). *Research Monograph No. 6: Criteria for alignment of expectations and assessments in Science and Science education*. Washington, D.C.: Council of Chief State Schools Officers.

Webb, N. L. (2005). *Webb alignment tool: Training manual*. Madison, WI: Wisconsin Center for Education Research. Available: http://www.wcer.wisc.edu/WAT/index.aspx

Appendix A.
Content Alignment Results: Science

The following tables include complete statistical results on the Webb (1997) alignment indicators, including means and standard deviations per strand for each CoAlt science test.

Categorical Concurrence

The categorical concurrence results for the grades 5 and 8 and high school CoAlt science tests are presented below. Each table includes: the mean number of items matched by panelists; the standard deviation among panelists' ratings; and, the final alignment conclusion (Yes or No). The bottom row indicates the percentage of strands that met the minimum alignment indicator criterion.

Table A-1. Categorical Concurrence for CoAlt Science, Grade 5: Mean Number of Items per Grade Level Expectation

Grade Level Expectation	Number of Items per GLE		At Least Six Items per GLE
	Mean Items Matched	*SD*	
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.	3.00	0.00	No
All organisms have structures and systems with separate functions.	4.00	0.00	No
Human body systems have basic structures, functions, and needs.	3.00	0.00	No
Earth and Sun provide a diversity of renewable and nonrenewable resources.	2.00	0.00	No
Earth's surface changes constantly through a variety of processes and forces.	2.00	0.00	No
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.	3.00	0.00	No
Percentage of GLEs with at least six items: 0%			

Table A-2. Categorical Concurrence for CoAlt Science, Grade 8: Mean Number of Items per Grade Level Expectation

Title of Strand	Number of Items per GLE		At Least Six Items per GLE
	Mean Items Matched	*SD*	
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.	0.00	0.00	No
There are different forms of energy, and those forms of energy can be changed from one form to another – but total energy is conserved.	1.60	0.89	No
Distinguish between physical and chemical changes, noting that mass is conserved during any change.	2.00	0.00	No
Recognize that waves such as electromagnetic, sound, seismic, and water have common characteristics and unique properties.	2.40	0.89	No
Human activities can deliberately or inadvertently alter ecosystems and their resiliency.	3.00	0.00	No
Organisms reproduce and transmit genetic information (genes) to offspring, which influences individuals' traits in the next generation.	5.00	0.00	No
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.	2.00	0.00	No
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.	3.00	0.00	No
The solar system is comprised of various objects that orbit the Sun and are classified based on their characteristics.	4.20	1.10	No
The relative positions and motions of Earth, Moon, and Sun can be used to explain observable effects such as seasons, eclipses, and Moon phases.	2.80	1.10	No
Percentage of GLEs with at least six items: 0%			

Table A-3. Categorical Concurrence for CoAlt Science, High School: Mean Number of Items per Grade Level Expectation

Grade Level Expectation	Number of Items per GLE		At Least Six Items per GLE
	Mean Items Matched	*SD*	
Newton's laws of motion and gravitation describe the relationships among forces acting on and between objects, their masses, and changes in their motion – but have limitations.	1.25	0.50	No
Matter has definite structure that determines characteristic physical and chemical properties.	1.50	0.58	No
Matter can change form through chemical or nuclear reactions abiding by the laws of conservation of mass and energy.	1.00	0.00	No
Atoms bond in different ways to form molecules and compounds that have definite properties.	1.50	0.58	No
Energy exists in many forms such as mechanical, chemical, electrical, radiant, thermal, and nuclear, that can be quantified and experimentally determined.	1.00	0.00	No
When energy changes form, it is neither created not destroyed; however, because some is necessarily lost as heat, the amount of energy available to do work decreases.	1.00	0.00	No
Matter tends to be cycled within an ecosystem, while energy is transformed and eventually exits an ecosystem.	1.00	0.00	No
The size and persistence of populations depend on their interactions with each other and on the abiotic factors in an ecosystem.	2.25	0.96	No
Cellular metabolic activities are carried out by biomolecules produced by organisms.	1.00	0.00	No
The energy for life primarily derives from the interrelated processes of photosynthesis and cellular respiration. Photosynthesis transforms the sun's light energy into the chemical energy of molecular bonds. Cellular respiration allows cells to utilize chemical energy when these bonds are broken.	1.75	0.50	No
Cells use passive and active transport of substances across membranes to maintain relatively stable intracellular environments.	1.50	0.71	No
Cells, tissues, organs, and organ systems maintain relatively stable internal environments, even in the face of changing external environments.	1.00	0.00	No
Physical and behavioral characteristics of an organism are influenced to varying degrees by heritable genes, many of which encode instructions for the production of proteins.	1.25	0.50	No

Grade Level Expectation	Number of Items per GLE		At Least Six Items per GLE
	Mean Items Matched	*SD*	
Multicellularity makes possible a division of labor at the cellular level through the expression of select genes, but not the entire genome.	1.00	0.00	No
Evolution occurs as the heritable characteristics of populations change across generations and can lead populations to become better adapted to their environment.	1.00	0.00	No
The history of the universe, solar system and Earth can be inferred from evidence left from past events.	1.00	0.00	No
As part of the solar system, Earth interacts with various extraterrestrial forces and energies such as gravity, solar phenomena, electromagnetic radiation, and impact events that influence the planet's geosphere, atmosphere, and biosphere in a variety of ways.	2.00	0.00	No
The theory of plate tectonics helps explain geological, physical, and geographical features of Earth.	1.50	0.58	No
Climate is the result of energy transfer among interactions of the atmosphere, hydrosphere, geosphere, and biosphere.	1.00	0.00	No
There are costs, benefits, and consequences of exploration, development, and consumption of renewable and nonrenewable resources.	1.50	0.58	No
The interaction of Earth's surface with water, air, gravity, and biological activity causes physical and chemical changes	1.00	0.00	No
Natural hazards have local, national and global impacts such as volcanoes, earthquakes, tsunamis, hurricanes, and thunderstorms	1.25	0.50	No
Percentage of GLEs with at least six items: 0%			

Depth-of-Knowledge Consistency

The Depth-of-Knowledge (DOK) consistency results for the grades 5 and 8 and high school CoAlt science tests are presented below. The tables present the results from the comparison between the depth-of-knowledge expected in the matched evidence outcome and the depth-of-knowledge assessed by items. The tables include the mean percentage of items rated as below, at the same level, or above the DOK level of the EOs along with the corresponding standard deviations. GLEs with at least 50% of items at the same (or above) DOK level of the matched EEO met the minimum indicator criterion.

Table A-4. DOK Consistency for CoAlt Science, Grade 5: Mean Percent of Items with DOK Below, At, and Above DOK Level of EOs

Grade Level Expectation	Mean Items per GLE	Depth-of-Knowledge Consistency						DOK Consistency Target Met
		% Items Below		% Items Same Level		% Items Above		
		M	*SD*	*M*	*SD*	*M*	*SD*	
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.	3.00	0.00	0.00	66.67	0.00	33.33	0.00	Yes
All organisms have structures and systems with separate functions.	4.00	40.00	22.36	35.00	22.36	25.00	0.00	Yes
Human body systems have basic structures, functions, and needs.	3.00	6.67	14.91	80.00	18.26	13.33	18.26	Yes
Earth and Sun provide a diversity of renewable and nonrenewable resources.	2.00	40.00	22.36	10.00	22.36	50.00	35.36	Yes
Earth's surface changes constantly through a variety of processes and forces.	2.00	0.00	0.00	50.00	0.00	50.00	0.00	Yes
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.	3.00	33.33	0.00	46.67	18.26	20.00	18.26	Yes
Percentage of GLEs with 50% of item DOK at or above objective DOK: 100%								

Table A-5. DOK Consistency for CoAlt Science, Grade 8: Mean Percent of Items with DOK Below, At, and Above DOK Level of EOs

Grade Level Expectation	Mean Items per GLE	Depth-of-Knowledge Consistency						DOK Consistency Target Met
		% Items Below		% Items Same Level		% Items Above		
		M	*SD*	*M*	*SD*	*M*	*SD*	
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.	0.00	--	--	--	--	--	--	No
There are different forms of energy, and those forms of energy can be changed from one form to another – but total energy is conserved.	1.60	53.33	50.55	46.67	50.55	0.00	0.00	No
Distinguish between physical and chemical changes, noting that mass is conserved during any change.	2.00	20.00	27.39	70.00	27.39	10.00	22.36	Yes
Recognize that waves such as electromagnetic, sound, seismic, and water have common characteristics and unique properties.	2.40	30.00	29.81	33.33	47.14	36.67	41.50	Yes
Human activities can deliberately or inadvertently alter ecosystems and their resiliency.	3.00	26.67	36.51	73.33	36.51	0.00	0.00	Yes
Organisms reproduce and transmit genetic information (genes) to offspring, which influences individuals' traits in the next generation.	5.00	16.00	8.94	60.00	20.00	24.00	16.73	Yes
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.	2.00	10.00	22.36	80.00	27.39	10.00	22.36	Yes
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.	3.00	40.00	27.89	60.00	27.89	0.00	0.00	Yes
The solar system is comprised of various objects that orbit the Sun and are classified based on their characteristics.	4.20	22.67	20.87	65.33	25.99	12.00	10.95	Yes
The relative positions and motions of Earth, Moon, and Sun can be used to explain observable effects such as seasons, eclipses, and Moon phases.	2.80	10.00	22.36	70.00	27.39	20.00	27.39	Yes
Percentage of GLEs with 50% of item DOK at or above objective DOK: 80%								

Table A-6. DOK Consistency for CoAlt Science, High School: Mean Percent of Items with DOK Below, At, and Above DOK Level of EOs

Grade Level Expectation	Mean Items per GLE	Depth-of-Knowledge Consistency						DOK Consistency Target Met
		% Items Below		% Items Same Level		% Items Above		
		M	*SD*	*M*	*SD*	*M*	*SD*	
Newton's laws of motion and gravitation describe the relationships among forces acting on and between objects, their masses, and changes in their motion – but have limitations.	1.25	12.50	25.00	87.50	25.00	0.00	0.00	Yes
Matter has definite structure that determines characteristic physical and chemical properties.	1.50	0.00	0.00	50.00	40.82	50.00	40.82	Yes
Matter can change form through chemical or nuclear reactions abiding by the laws of conservation of mass and energy.	1.00	0.00	0.00	0.00	0.00	100.00	0.00	Yes
Atoms bond in different ways to form molecules and compounds that have definite properties.	1.50	25.00	28.87	50.00	40.82	25.00	50.00	Yes
Energy exists in many forms such as mechanical, chemical, electrical, radiant, thermal, and nuclear, that can be quantified and experimentally determined.	1.00	0.00	0.00	100.00	0.00	0.00	0.00	Yes
When energy changes form, it is neither created not destroyed; however, because some is necessarily lost as heat, the amount of energy available to do work decreases.	1.00	0.00	0.00	25.00	50.00	75.00	50.00	Yes
Matter tends to be cycled within an ecosystem, while energy is transformed and eventually exits an ecosystem.	1.00	50.00	57.74	50.00	57.74	0.00	0.00	Yes
The size and persistence of populations depend on their interactions with each other and on the abiotic factors in an ecosystem.	0.00	--	--	--	--	--	--	No
Cellular metabolic activities are carried out by biomolecules produced by organisms.	2.25	45.83	41.67	33.33	47.14	20.83	25.00	Yes
The energy for life primarily derives from the interrelated processes of photosynthesis and cellular respiration. Photosynthesis transforms the sun's light energy into the	1.00	100.00	0.00	0.00	0.00	0.00	0.00	No

Grade Level Expectation	Mean Items per GLE	Depth-of-Knowledge Consistency						DOK Consistency Target Met
		% Items Below		% Items Same Level		% Items Above		
		M	*SD*	*M*	*SD*	*M*	*SD*	
chemical energy of molecular bonds. Cellular respiration allows cells to utilize chemical energy when these bonds are broken.								
Cells use passive and active transport of substances across membranes to maintain relatively stable intracellular environments.	1.75	0.00	0.00	37.50	25.00	62.50	25.00	Yes
Cells, tissues, organs, and organ systems maintain relatively stable internal environments, even in the face of changing external environments.	1.50	0.00	0.00	25.00	35.36	75.00	35.36	Yes
Physical and behavioral characteristics of an organism are influenced to varying degrees by heritable genes, many of which encode instructions for the production of proteins.	1.00	0.00	0.00	25.00	50.00	75.00	50.00	Yes
Multicellularity makes possible a division of labor at the cellular level through the expression of select genes, but not the entire genome.	1.25	0.00	0.00	100.00	0.00	0.00	0.00	Yes
Evolution occurs as the heritable characteristics of populations change across generations and can lead populations to become better adapted to their environment.	1.00	0.00	0.00	25.00	50.00	75.00	50.00	Yes
The history of the universe, solar system and Earth can be inferred from evidence left from past events.	1.00	0.00	0.00	100.00	0.00	0.00	0.00	Yes
As part of the solar system, Earth interacts with various extraterrestrial forces and energies such as gravity, solar phenomena, electromagnetic radiation, and impact events that influence the planet's geosphere, atmosphere, and biosphere in a variety of ways.	1.00	100.00	0.00	0.00	0.00	0.00	0.00	No
The theory of plate tectonics helps explain geological,	2.00	0.00	0.00	75.00	28.87	25.00	28.87	Yes

Grade Level Expectation	Mean Items per GLE	Depth-of-Knowledge Consistency						DOK Consistency Target Met
		% Items Below		% Items Same Level		% Items Above		
		M	*SD*	*M*	*SD*	*M*	*SD*	
physical, and geographical features of Earth.								
Climate is the result of energy transfer among interactions of the atmosphere, hydrosphere, geosphere, and biosphere.	1.50	12.50	25.00	50.00	47.82	37.50	47.87	Yes
There are costs, benefits, and consequences of exploration, development, and consumption of renewable and nonrenewable resources.	1.00	0.00	0.00	100.00	0.00	0.00	0.00	Yes
The interaction of Earth's surface with water, air, gravity, and biological activity causes physical and chemical changes	1.50	25.00	50.00	37.50	47.87	37.50	47.87	Yes
Natural hazards have local, national and global impacts such as volcanoes, earthquakes, tsunamis, hurricanes, and thunderstorms	1.00	100.00	0.00	0.00	0.00	0.00	0.00	No
Percentage of GLEs with 50% of item DOK at or above objective DOK:82%								

Range-of-Knowledge Correspondence

The results for Range-of-Knowledge correspondence for the grades 5 and 8 and high school CoAlt science tests are presented below. The tables include the mean number, standard deviation, and percentage of EEOs by GLE. For acceptable range-of-knowledge correspondence, a minimum of 50% of EEOs within each GLE should be matched to at least one item.

Table A-7. Range-of-Knowledge for CoAlt Science, Grade 5: Mean Percent of EEOs per GLE Linked with Items

Grade Level Expectation	Number of EEOs	Mean Items per GLE	Range of EEOs			Range-of-Knowledge Target Met
			EEOs with At Least One Item		% of Total EEOs per GLE	
			M	*SD*		
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.	2	3.00	2.00	0.00	100.00	Yes
All organisms have structures and systems with separate functions.	3	4.00	3.00	0.00	100.00	Yes
Human body systems have basic structures, functions, and needs.	2	3.00	2.00	0.00	100.00	Yes
Earth and Sun provide a diversity of renewable and nonrenewable resources.	3	2.00	2.00	0.00	66.67	Yes
Earth's surface changes constantly through a variety of processes and forces.	3	2.00	2.00	0.00	66.67	Yes
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.	3	3.00	2.80	0.45	93.33	Yes
Percentage of GLEs with 50% of EEOs linked to at least one item: 100%						

Table A-8. Range-of-Knowledge for CoAlt Science, Grade 8: Mean Percent of EEOs per GLE Linked with Items

Grade Level Expectation	Number of EOs	Mean Items per GLE	Range of EEOs			Range-of-Knowledge Target Met
			EEOs with At Least One Item		% of Total EEOs per GLE	
			M	*SD*		
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.	2	0.00	0.00	0.00	0.00	No
There are different forms of energy, and those forms of energy can be changed from one form to another – but total energy is conserved.	1	1.60	1.00	0.00	100.00	Yes
Distinguish between physical and chemical changes, noting that mass is conserved during any change.	2	2.00	1.60	0.55	80.00	Yes
Recognize that waves such as electromagnetic, sound, seismic, and water have common characteristics and unique properties.	3	2.40	2.40	0.89	80.00	Yes
Human activities can deliberately or inadvertently alter ecosystems and their resiliency.	1	3.00	1.00	0.00	100.00	Yes
Organisms reproduce and transmit genetic information (genes) to offspring, which influences individuals' traits in the next generation.	2	5.00	2.00	0.00	100.00	Yes
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.	3	2.00	2.00	0.00	66.67	Yes
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.	2	3.00	2.00	0.00	100.00	Yes
The solar system is comprised of various objects that orbit the Sun and are classified based on their characteristics.	4	4.20	3.00	0.71	75.00	Yes
The relative positions and motions of Earth, Moon, and Sun can be used to explain observable effects such as seasons, eclipses, and Moon phases.	5	2.80	2.80	1.10	56.00	Yes
Percentage of GLEs with 50% of EEOs linked to at least one item:90%						

Table A-9. Range-of-Knowledge for CoAlt Science, High School: Mean Percent of EEOs per GLE Linked with Items

Grade Level Expectation	Number of EEOs	Mean Items per GLE	Range of EEOs			Range-of-Knowledge Target Met
			EEOs with At Least One Item		% of Total EEOs per GLE	
			M	*SD*		
Newton's laws of motion and gravitation describe the relationships among forces acting on and between objects, their masses, and changes in their motion – but have limitations.	3	1.25	1.00	0.00	33.33	No
Matter has definite structure that determines characteristic physical and chemical properties.	3	1.50	1.50	0.58	50.00	Yes
Matter can change form through chemical or nuclear reactions abiding by the laws of conservation of mass and energy.	1	1.00	1.00	0.00	100.00	Yes
Atoms bond in different ways to form molecules and compounds that have definite properties.	2	1.50	1.25	0.50	62.50	Yes
Energy exists in many forms such as mechanical, chemical, electrical, radiant, thermal, and nuclear, that can be quantified and experimentally determined.	3	1.00	1.00	0.00	33.33	No
When energy changes form, it is neither created not destroyed; however, because some is necessarily lost as heat, the amount of energy available to do work decreases.	2	1.00	1.00	0.00	50.00	Yes
Matter tends to be cycled within an ecosystem, while energy is transformed and eventually exits an ecosystem.	1	1.00	1.00	0.00	100.00	Yes
The size and persistence of populations depend on their interactions with each other and on the abiotic factors in an ecosystem.	2	0.00	0.00	0.00	0.00	No
Cellular metabolic activities are carried out by biomolecules produced by organisms.	3	2.25	1.75	0.50	58.33	Yes
The energy for life primarily derives from the interrelated processes of photosynthesis and cellular respiration. Photosynthesis transforms the sun's light energy into the chemical energy of molecular bonds. Cellular respiration allows cells to utilize chemical energy when these bonds are broken.	2	1.00	1.00	0.00	50.00	Yes
Cells use passive and active transport of substances across membranes to maintain relatively stable intracellular environments.	2	1.75	1.75	0.50	87.50	Yes
Cells, tissues, organs, and organ systems maintain relatively stable internal environments, even in the face of changing external environments.	2	1.50	1.00	0.00	50.00	Yes

Grade Level Expectation	Number of EEOs	Mean Items per GLE	Range of EEOs			Range-of-Knowledge Target Met
			EEOs with At Least One Item		% of Total EEOs per GLE	
			M	*SD*		
Physical and behavioral characteristics of an organism are influenced to varying degrees by heritable genes, many of which 4 encode instructions for the production of proteins.	2	1.00	1.00	0.00	50.00	Yes
Multicellularity makes possible a division of labor at the cellular level through the expression of select genes, but not the entire genome.	1	1.25	1.00	0.00	100.00	Yes
Evolution occurs as the heritable characteristics of populations change across generations and can lead populations to become better adapted to their environment.	1	1.00	1.00	0.00	100.00	Yes
The history of the universe, solar system and Earth can be inferred from evidence left from past events.	2	1.00	1.00	0.00	50.00	Yes
As part of the solar system, Earth interacts with various extraterrestrial forces and energies such as gravity, solar phenomena, electromagnetic radiation, and impact events that influence the planet's geosphere, atmosphere, and biosphere in a variety of ways.	2	1.00	1.00	0.00	50.00	Yes
The theory of plate tectonics helps explain geological, physical, and geographical features of Earth.	2	2.00	1.00	0.00	50.00	Yes
Climate is the result of energy transfer among interactions of the atmosphere, hydrosphere, geosphere, and biosphere.	4	1.50	1.50	0.58	37.50	No
There are costs, benefits, and consequences of exploration, development, and consumption of renewable and nonrenewable resources.	2	1.00	1.00	0.00	50.00	Yes
The interaction of Earth's surface with water, air, gravity, and biological activity causes physical and chemical changes	2	1.50	1.50	0.58	75.00	Yes
Natural hazards have local, national and global impacts such as volcanoes, earthquakes, tsunamis, hurricanes, and thunderstorms	2	1.00	1.00	0.00	50.00	Yes
Percentage of GLEs with 50% of EEOs linked to at least one item:82%						

Balance-of-Knowledge Representation

The results for Balance-of-Knowledge representation for the grades 5 and 8 and high school CoAlt science tests are presented below. The tables also include the percentage of items linked to each strand. The minimum acceptable balance index is 70 out of 100.

Table A-10. Balance-of-Knowledge Representation for CoAlt Science Grade 5: Mean Balance Index per GLE

Grade Level Expectation	EEOs per GLE	Balance-of-Knowledge Representation					Balance Index Target Met
		Mean EEOs Linked with Items	Mean Items per GLE	Mean % of Items (of total) Linked to GLE	Mean Balance Index		
		M	*M*	*M*	*M*	*SD*	
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.	2	2.00	3.00	17.65	83.33	0.00	Yes
All organisms have structures and systems with separate functions.	3	3.00	4.00	23.53	83.33	0.00	Yes
Human body systems have basic structures, functions, and needs.	2	2.00	3.00	17.65	83.33	0.00	Yes
Earth and Sun provide a diversity of renewable and nonrenewable resources.	3	2.00	2.00	11.76	100.00	0.00	Yes
Earth's surface changes constantly through a variety of processes and forces.	3	2.00	2.00	11.76	100.00	0.00	Yes
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.	3	2.80	3.00	17.65	96.67	7.45	Yes
Total	16						
Percentage of GLEs with a balance of representation index of 70 or greater: 100%							

Table A-11. Balance-of-Knowledge Representation for CoAlt Science Grade 8: Mean Balance Index per GLE

Grade Level Expectation	EEOs per GLE	Balance-of-Knowledge Representation					Balance Index Target Met
		Mean EEOs Linked with Items	Mean Items per GLE	Mean % of Items (of total) Linked to GLE	Mean Balance Index		
		M	*M*	*M*	*M*	*SD*	
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.	2	0.00	0.00	0.00	0.00	0.00	No
There are different forms of energy, and those forms of energy can be changed from one form to another – but total energy is conserved.	1	1.00	1.60	6.25	100.00	0.00	Yes
Distinguish between physical and chemical changes, noting that mass is conserved during any change.	2	1.60	1.80	6.95	100.00	0.00	Yes
Recognize that waves such as electromagnetic, sound, seismic, and water have common characteristics and unique properties.	3	2.40	2.40	9.26	100.00	0.00	Yes
Human activities can deliberately or inadvertently alter ecosystems and their resiliency.	1	1.00	3.00	11.63	100.00	0.00	Yes
Organisms reproduce and transmit genetic information (genes) to offspring, which influences individuals' traits in the next generation.	2	2.00	5.00	19.38	86.00	8.94	Yes
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.	3	2.00	2.00	7.75	100.00	0.00	Yes
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.	2	2.00	3.00	11.63	83.33	0.00	Yes
The solar system is comprised of various objects that orbit the Sun and are classified based on their characteristics.	4	3.00	4.20	16.31	85.67	9.55	Yes

Grade Level Expectation	EEOs per GLE	Balance-of-Knowledge Representation					Balance Index Target Met
		Mean EEOs Linked with Items	Mean Items per GLE	Mean % of Items (of total) Linked to GLE	Mean Balance Index		
		M	*M*	*M*	*M*	*SD*	
The relative positions and motions of Earth, Moon, and Sun can be used to explain observable effects such as seasons, eclipses, and Moon phases.	5	2.80	2.80	10.83	100.00	0.00	Yes
Total	25						
Percentage of GLEs with a balance of representation index of 70 or greater: 100%							

Table A-12. Balance-of-Knowledge Representation for CoAlt Science High School: Mean Balance Index per GLE

Grade Level Expectation	EEOs per GLE	Balance-of-Knowledge Representation					Balance Index Target Met
		Mean EEOs Linked with Items	Mean Items per GLE	Mean % of Items (of total) Linked to GLE	Mean Balance Index		
		M	*M*	*M*	*M*	*SD*	
Newton's laws of motion and gravitation describe the relationships among forces acting on and between objects, their masses, and changes in their motion – but have limitations.	3	1.00	1.25	4.81	100.00	0.00	Yes
Matter has definite structure that determines characteristic physical and chemical properties.	3	1.50	1.50	5.77	100.00	0.00	Yes
Matter can change form through chemical or nuclear reactions abiding by the laws of conservation of mass and energy.	1	1.00	1.00	3.85	100.00	0.00	Yes
Atoms bond in different ways to form molecules and compounds that have definite properties.	2	1.25	1.50	5.77	100.00	0.00	Yes
Energy exists in many forms such as mechanical, chemical, electrical, radiant, thermal, and nuclear, that can be quantified and experimentally determined.	3	1.00	1.00	3.85	100.00	0.00	Yes
When energy changes form, it is neither created not destroyed; however, because some is necessarily lost as heat, the amount of energy available to do work decreases.	2	1.00	1.00	3.85	100.00	0.00	Yes
Matter tends to be cycled within an ecosystem, while energy is transformed and eventually exits an ecosystem.	1	1.00	1.00	3.85	100.00	0.00	Yes
The size and persistence of populations depend on their interactions with each other and on the abiotic factors in an ecosystem.	2	0.00	0.00	0.00	0.00	0.00	No
Cellular metabolic activities are carried out by biomolecules produced by organisms.	3	1.75	2.25	8.65	91.67	9.62	Yes

Grade Level Expectation	EEOs per GLE	Balance-of-Knowledge Representation					Balance Index Target Met
		Mean EEOs Linked with Items	Mean Items per GLE	Mean % of Items (of total) Linked to GLE	Mean Balance Index		
		M	*M*	*M*	*M*	*SD*	
The energy for life primarily derives from the interrelated processes of photosynthesis and cellular respiration. Photosynthesis transforms the sun's light energy into the chemical energy of molecular bonds. Cellular respiration allows cells to utilize chemical energy when these bonds are broken.	2	1.00	1.00	3.85	100.00	0.00	Yes
Cells use passive and active transport of substances across membranes to maintain relatively stable intracellular environments.	2	1.75	1.75	6.73	100.00	0.00	Yes
Cells, tissues, organs, and organ systems maintain relatively stable internal environments, even in the face of changing external environments.	2	1.00	1.50	5.77	100.00	0.00	Yes
Physical and behavioral characteristics of an organism are influenced to varying degrees by heritable genes, many of which 4encode instructions for the production of proteins.	2	1.00	1.00	3.85	100.00	0.00	Yes
Multicellularity makes possible a division of labor at the cellular level through the expression of select genes, but not the entire genome.	1	1.00	1.25	4.81	100.00	0.00	Yes
Evolution occurs as the heritable characteristics of populations change across generations and can lead populations to become better adapted to their environment.	1	1.00	1.00	3.85	100.00	0.00	Yes
The history of the universe, solar system and Earth can be inferred from evidence left from past events.	2	1.00	1.00	3.85	100.00	0.00	Yes

Grade Level Expectation	EEOs per GLE	Balance-of-Knowledge Representation					Balance Index Target Met
		Mean EEOs Linked with Items	Mean Items per GLE	Mean % of Items (of total) Linked to GLE	Mean Balance Index		
		M	*M*	*M*	*M*	*SD*	
As part of the solar system, Earth interacts with various extraterrestrial forces and energies such as gravity, solar phenomena, electromagnetic radiation, and impact events that influence the planet's geosphere, atmosphere, and biosphere in a variety of ways.	2	1.00	1.00	3.85	100.00	0.00	Yes
The theory of plate tectonics helps explain geological, physical, and geographical features of Earth.	2	1.00	2.00	7.69	100.00	0.00	Yes
Climate is the result of energy transfer among interactions of the atmosphere, hydrosphere, geosphere, and biosphere.	4	1.50	1.50	5.77	100.00	0.00	Yes
There are costs, benefits, and consequences of exploration, development, and consumption of renewable and nonrenewable resources.	2	1.00	1.00	3.85	100.00	0.00	Yes
The interaction of Earth's surface with water, air, gravity, and biological activity causes physical and chemical changes	2	1.50	1.50	5.77	100.00	0.00	Yes
Natural hazards have local, national and global impacts such as volcanoes, earthquakes, tsunamis, hurricanes, and thunderstorms	2	1.00	1.00	3.85	100.00	0.00	Yes
Total	46						
Percentage of GLEs with a balance of representation index of 70 or greater: 95%							

EEOs Matched to Items by Panelists

Tables A-13 through A-15 present the EEOs, along with the mean number of items, matched by panelists. Column 1 presents the HumRRO code corresponding to each of the EEOs. One note of caution when reading these tables, the same items may not be represented by the mean number of items. For example, EEO code '1.1.a' in the first row shows that 7 panelists matched a mean number of 7.14 items to this EEO. This does not mean/assume that the items matched to the EEO by the panelists were the same items across panelists.

Table A-13. Grade 5 CoAlt Science: EEOs Matched to Items by Panelists

HumRRO EEO Code	Number of Panelists	Mean Number of Items per EEO	*SD*
1.1.a	5	2.00	0.00
1.1.b	5	1.00	0.00
2.1.a	5	1.20	0.45
2.1.b	5	1.80	0.45
2.1.c	5	1.00	0.00
2.2.a	5	2.00	0.00
2.2.b	5	1.00	0.00
3.1.a	1	1.00	0.00
3.1.b	5	1.00	0.00
3.1.c	4	1.00	0.00
3.2.a	5	1.00	0.00
3.2.b	0	0.00	0.00
3.2.c	5	1.00	0.00
3.3.a	4	1.00	0.00
3.3.b	5	1.00	0.00
3.3.c	5	1.20	0.45

Table A-14. Grade 8 CoAlt Science: EEOs Matched to Items by Panelists

HumRRO EEO	Number of Panelists	Mean Number of Items per EEO	*SD*
1.1.a	0	0.00	0.00
1.1.b	0	0.00	0.00
1.2.a	5	1.60	0.89
1.3.a	4	1.25	0.50
1.3.b	4	1.00	0.00
1.3.c	1	1.00	0.00
1.4.a	3	1.00	0.00
1.4.b	4	1.00	0.00
1.4.c	5	1.00	0.00
2.1.a	5	3.00	0.00
2.2.a	5	1.80	0.45
2.2.b	5	3.20	0.45
3.1.a	4	1.00	0.00
3.1.b	2	1.00	0.00
3.1.c	4	1.00	0.00
3.2.a	5	2.00	0.00
3.2.b	5	1.00	0.00
3.3.a	2	1.00	0.00
3.3.b	3	1.33	0.58
3.3.c	5	2.00	0.71
3.3.d	5	1.00	0.00
3.4.a	2	1.00	0.00
3.4.c	4	1.00	0.00
3.4.d	3	1.00	0.00
3.4.e	5	1.00	0.00

Table A-15. High School CoAlt Science: EEOs Matched to Items by Panelists

HumRRO EEO Code	Number of Panelists	Mean Number of Items per EEO	*SD*
1.1.a	1	1.00	0.00
1.1.b	0	0.00	0.00
1.1.c	3	1.33	0.58
1.2.a	3	1.00	0.00
1.2.b	2	1.00	0.00
1.2.c	1	1.00	0.00
1.3.a	4	1.00	0.00
1.4.a	3	1.33	0.58
1.4.b	2	1.00	0.00
1.5.a	0	0.00	0.00
1.5.b	3	1.00	0.00
1.5.c	1	1.00	0.00
1.6.a	1	1.00	0.00
1.6.b	3	1.00	0.00
2.1.a	4	1.00	0.00
2.2.a	0	0.00	0.00
2.2.b	0	0.00	0.00
2.3.a	1	1.00	0.00
2.3.b	3	1.00	0.00
2.3.c	3	1.67	0.58
2.4.a	0	0.00	0.00
2.4.b	4	1.00	0.00
2.5.a	4	1.00	0.00
2.5.b	3	1.00	0.00
2.6.a	0	0.00	0.00
2.6.b	2	1.50	0.71
2.7.a	4	1.00	0.00
2.7.b	0	0.00	0.00
2.8.a	4	1.25	0.50
2.9.a	4	1.00	0.00
3.1.a	0	0.00	0.00
3.1.b	4	1.00	0.00
3.2.a	0	0.00	0.00
3.2.b	4	1.00	0.00
3.3.a	0	0.00	0.00
3.3.b	4	2.00	0.00
3.4.a	0	0.00	0.00
3.4.b	2	1.00	0.00
3.4.c	4	1.00	0.00
3.4.d	0	0.00	0.00
3.5.a	3	1.00	0.00
3.5.b	0	0.00	0.00
3.6.a	3	1.00	0.00
3.6.b	3	1.00	0.00
3.7.a	0	0.00	0.00
3.7.b	4	1.00	0.00

Appendix B.
Content Alignment Results: Social Studies

The following tables include complete statistical results on the Webb alignment indicators, including means and standard deviations per strand for each CoAlt social studies test.

Categorical Concurrence

The categorical concurrence results for the grades 4 and 7 and high school CoAlt social studies tests are presented below. Each table includes: the mean number of items matched by panelists; the standard deviation among panelists' ratings; and, the final alignment conclusion (Yes or No). The bottom row indicates the percentage of strands that met the minimum alignment indicator criterion.

Table B-1. Categorical Concurrence for CoAlt Social Studies, Grade 4: Mean Number of Items per Grade Level Expectation

Grade Level Expectation	Number of Items per GLE		At Least Six Items per GLE
	Mean Items Matched	*SD*	
Organize and sequence events to understand the concepts of chronology and cause and effect in the history of Colorado.	1.50	0.71	No
The historical eras, individuals, groups, ideas and themes in Colorado history and their relationships to key events in the United States.	1.00	0.00	No
Use several types of geographic tools to answer questions about the geography of Colorado.	2.00	0.00	No
Connections within and across human and physical systems are developed.	2.33	0.58	No
People respond to positive and negative incentives.	3.33	0.58	No
The relationship between choice and opportunity cost (PFL).	3.00	0.00	No
Analyze and debate multiple perspectives on an issue.	2.00	0.00	No
The origins, structure, and functions of the Colorado government	2.00	0.00	No
Percentage of GLEs with at least six items: 0%			

Table B-2. Categorical Concurrence for CoAlt Social Studies, Grade 7: Mean Number of Items per Grade Level Expectation

Grade Level Expectation	Number of Items per GLE		At Least Six Items per GLE
	Mean Items Matched	*SD*	
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.	2.80	0.45	No
The historical eras, individuals, groups, ideas and themes within regions of the Eastern Hemisphere and their relationships with one another	2.20	0.45	No
Use geographic tools to gather data and make geographic inferences and predictions.	2.00	0.00	No
Regions have different issues and perspectives.	2.00	0.00	No
Supply and demand influence price and profit in a market economy.	2.00	0.00	No
The distribution of resources influences economic production and individual choices (Economics and PFL).	1.00	0.00	No
Compare how various nations define the rights, responsibilities, and roles of citizens.	3.40	0.89	No
Different forms of government and international organizations and their influence in the world community.	2.00	0.00	No
Percentage of GLEs with at least six items: 0%			

Table B-3. Categorical Concurrence for CoAlt Social Studies, High School: Mean Number of Items per Grade Level Expectation

Grade Level Expectation	Number of Items per GLE		At Least Six Items per GLE
	Mean Items Matched	*SD*	
Use the historical method of inquiry to ask questions, evaluate primary and secondary sources, critically analyze and interpret data, and develop interpretations defended by evidence.	2.00	0.00	No
The key concepts of continuity and change, cause and effect, complexity, unity and diversity over time.	3.60	0.55	No
The significance of ideas as powerful forces throughout history.	1.00	0.00	No
Use different types of maps and geographic tools to analyze features on Earth to investigate and solve geographic questions.	3.20	0.45	No
Explain and interpret geographic variables that influence the interactions of people, places and environments.	2.00	0.00	No
The interconnected nature of the world, its people and places.	1.80	0.45	No
Productive resources – natural, human, capital – are scarce; therefore, choices are made about how individuals, businesses, governments, and societies allocate these resources.	1.00	0.00	No
Economic policies affect markets.	1.20	0.45	No
Government and competition affect markets.	1.00	0.00	No
Design, analyze, and apply a financial plan based on short- and long-term financial goals (PFL).	1.00	0.00	No
Analyze strategic spending, saving, and investment options to achieve the objectives of diversification, liquidity, income, and growth (PFL).	1.00	0.00	No
The components of personal credit to manage credit and debt (PFL).	1.00	0.00	No
Identify, develop, and evaluate risk-management strategies (PFL).	1.00	0.00	No
Research, formulate positions, and engage in appropriate civic participation to address local, state, and national issues or policies.	1.50	0.58	No
Purposes of and limitations on the foundations, structures and functions of government.	2.20	0.45	No
Analyze how public policy - domestic and foreign - is developed at the local, state, and national levels and compare how policy-making occurs in other forms of government.	1.80	0.45	No
Percentage of GLEs with at least six items: 0%			

Depth-of-Knowledge Consistency

The Depth-of-Knowledge (DOK) consistency results for the grades 4 and 7 and high school CoAlt social studies tests are presented below. The tables present the results from the comparison between the depth-of-knowledge expected in the matched evidence outcome and the depth-of-knowledge assessed by items. The tables include the mean percentage of items rated as below, at the same level, or above the DOK level of the EEOs along with the corresponding standard deviations. GLEs with at least 50% of items at the same (or above) DOK level of the matched EEO met the minimum indicator criterion.

Table B-4. DOK Consistency for CoAlt Social Studies, Grade 4: Mean Percent of Items with DOK Below, At, and Above DOK Level of EEOs

Grade Level Expectation	Mean Items per GLE	Depth-of-Knowledge Consistency						DOK Consistency Target Met
		% Items Below		% Items Same Level		% Items Above		
		M	*SD*	*M*	*SD*	*M*	*SD*	
Organize and sequence events to understand the concepts of chronology and cause and effect in the history of Colorado.	1.50	0.00	0.00	100.00	0.00	0.00	0.00	Yes
The historical eras, individuals, groups, ideas and themes in Colorado history and their relationships to key events in the United States.	1.00	0.00	0.00	33.33	57.74	66.67	57.74	Yes
Use several types of geographic tools to answer questions about the geography of Colorado.	2.00	0.00	0.00	33.33	28.87	66.67	28.87	Yes
Connections within and across human and physical systems are developed.	2.33	0.00	0.00	61.11	34.69	38.89	34.69	Yes
People respond to positive and negative incentives.	3.33	69.44	33.68	30.56	33.68	0.00	0.00	No
The relationship between choice and opportunity cost (PFL).	3.00	11.11	19.25	33.33	0.00	55.56	19.25	Yes
Analyze and debate multiple perspectives on an issue.	2.00	0.00	0.00	100.00	0.00	0.00	0.00	Yes
The origins, structure, and functions of the Colorado government	2.00	66.67	28.87	33.33	28.87	0.00	0.00	No
Percentage of strands with 50% of item DOK at or above objective DOK: 75%								

Table B-5. DOK Consistency for CoAlt Social Studies, Grade 7: Mean Percent of Items with DOK Below, At, and Above DOK Level of EEOs

Grade Level Expectation	Mean Items per GLE	Depth-of-Knowledge Consistency						DOK Consistency Target Met
		% Items Below		% Items Same Level		% Items Above		
		M	*SD*	*M*	*SD*	*M*	*SD*	
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.	2.80	93.33	14.91	6.67	14.91	0.00	0.00	No
The historical eras, individuals, groups, ideas and themes within regions of the Eastern Hemisphere and their relationships with one another	2.20	0.00	0.00	73.33	25.28	26.67	25.28	Yes
Use geographic tools to gather data and make geographic inferences and predictions.	2.00	0.00	0.00	90.00	22.36	10.00	22.36	No
Regions have different issues and perspectives.	2.00	60.00	41.83	40.00	41.83	0.00	0.00	Yes
Supply and demand influence price and profit in a market economy.	2.00	0.00	0.00	50.00	50.00	50.00	50.00	Yes
The distribution of resources influences economic production and individual choices (Economics and PFL).	1.00	40.00	54.77	40.00	54.77	20.00	44.72	Yes
Compare how various nations define the rights, responsibilities, and roles of citizens.	3.40	13.33	18.26	52.00	17.26	34.67	23.76	Yes
Different forms of government and international organizations and their influence in the world community.	2.00	75.00	50.00	25.00	50.00	0.00	0.00	No
Percentage of strands with 50% of item DOK at or above objective DOK: 63%								

Table B-6. DOK Consistency for CoAlt Social Studies, High School: Mean Percent of Items with DOK Below, At, and Above DOK Level of EEOs

Grade Level Expectation	Mean Items per GLE	Depth-of-Knowledge Consistency						DOK Consistency Target Met
		% Items Below		% Items Same Level		% Items Above		
		M	*SD*	*M*	*SD*	*M*	*SD*	
Use the historical method of inquiry to ask questions, evaluate primary and secondary sources, critically analyze and interpret data, and develop interpretations defended by evidence.	2.00	70.00	44.72	30.00	44.72	00.00	0.00	No
The key concepts of continuity and change, cause and effect, complexity, unity and diversity over time.	3.60	70.00	41.08	15.00	33.54	15.00	13.69	No
The significance of ideas as powerful forces throughout history.	1.00	0.00	0.00	20.00	44.72	80.00	44.72	Yes
Use different types of maps and geographic tools to analyze features on Earth to investigate and solve geographic questions.	3.20	68.33	33.54	31.67	33.54	0.00	0.00	No
Explain and interpret geographic variables that influence the interactions of people, places and environments.	2.00	0.00	0.00	70.00	27.39	30.00	27.39	Yes
The interconnected nature of the world, its people and places.	1.80	0.00	0.00	10.00	22.36	90.00	22.36	Yes
Productive resources – natural, human, capital – are scarce; therefore, choices are made about how individuals, businesses, governments, and societies allocate these resources.	1.00	0.00	0.00	0.00	0.00	100.00	0.00	Yes
Economic policies affect markets.	1.20	0.00	0.00	70.00	44.72	30.00	44.72	Yes
Government and competition affect markets.	1.00	0.00	0.00	20.00	44.72	80.00	44.72	Yes
Design, analyze, and apply a financial plan based on short- and long-term financial goals (PFL).	1.00	60.00	54.77	40.00	54.77	0.00	0.00	No
Analyze strategic spending, saving, and investment options to achieve the objectives of diversification, liquidity, income, and growth (PFL).	1.00	20.00	44.72	60.00	54.77	20.00	44.72	Yes

Grade Level Expectation	Mean Items per GLE	Depth-of-Knowledge Consistency						DOK Consistency Target Met
		% Items Below		% Items Same Level		% Items Above		
		M	*SD*	*M*	*SD*	*M*	*SD*	
The components of personal credit to manage credit and debt (PFL).	1.00	40.00	54.77	40.00	54.77	20.00	44.72	Yes
Identify, develop, and evaluate risk-management strategies (PFL).	1.00	0.00	0.00	20.00	44.72	80.00	44.72	Yes
Research, formulate positions, and engage in appropriate civic participation to address local, state, and national issues or policies.	1.50	0.00	0.00	100.00	0.00	0.00	0.00	Yes
Purposes of and limitations on the foundations, structures and functions of government.	2.20	0.00	0.00	46.67	50.55	53.33	50.55	Yes
Analyze how public policy - domestic and foreign - is developed at the local, state, and national levels and compare how policy-making occurs in other forms of government.	1.80	20.00	27.39	60.00	41.83	20.00	27.39	Yes
Percentage of strands with 50% of item DOK at or above objective DOK: 75%								

Range-of-Knowledge Correspondence

The results for Range-of-Knowledge correspondence for the grades 4 and 7 and high school CoAlt social studies tests are presented below. The tables include the mean number, standard deviation, and percentage of EEOs by GLE. For acceptable range-of-knowledge correspondence, a minimum of 50% of EEOs within each GLE should be matched to at least one item.

Table B-7. Range-of-Knowledge for CoAlt Social Studies, Grade 4: Mean Percent of EEOs per GLE Linked with Items

Grade Level Expectation	Number of EEOs	Mean Items per GLE	Range of EEOs			Range-of-Knowledge Target Met
			EEOs with At Least One Item		% of Total EEOs per GLE	
			M	*SD*		
Organize and sequence events to understand the concepts of chronology and cause and effect in the history of Colorado.	1.00	1.50	0.00	0.71	100.00	Yes
The historical eras, individuals, groups, ideas and themes in Colorado history and their relationships to key events in the United States.	1.00	1.00	0.00	0.00	100.00	Yes
Use several types of geographic tools to answer questions about the geography of Colorado.	1.00	2.00	0.00	0.00	50.00	Yes
Connections within and across human and physical systems are developed.	1.67	2.33	0.58	0.58	83.33	Yes
People respond to positive and negative incentives.	1.00	3.33	0.00	0.58	100.00	Yes
The relationship between choice and opportunity cost (PFL).	1.00	3.00	0.00	0.00	100.00	Yes
Analyze and debate multiple perspectives on an issue.	1.00	2.00	0.00	0.00	100.00	Yes
The origins, structure, and functions of the Colorado government	1.00	2.00	0.00	0.00	100.00	Yes
Percentage of GLEs with 50% of EEOs linked to at least one item: 100%						

Table B-8. Range-of-Knowledge for CoAlt Social Studies, Grade 7: Mean Percent of EEOs per GLE Linked with Items

Grade Level Expectation	Number of EEOs	Mean Items per GLE	Range of EOs			Range-of-Knowledge Target Met
			EEOs with At Least One Item		% of Total EEOs per GLE	
			M	*SD*		
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.	1.60	2.80	0.55	0.45	80.00	Yes
The historical eras, individuals, groups, ideas and themes within regions of the Eastern Hemisphere and their relationships with one another	1.00	2.20	0.00	0.45	100.00	Yes
Use geographic tools to gather data and make geographic inferences and predictions.	1.00	2.00	0.00	0.00	100.00	Yes
Regions have different issues and perspectives.	1.00	2.00	0.00	0.00	100.00	Yes
Supply and demand influence price and profit in a market economy.	1.80	2.00	0.45	0.00	90.00	Yes
The distribution of resources influences economic production and individual choices (Economics and PFL).	1.00	1.00	0.00	0.00	50.00	Yes
Compare how various nations define the rights, responsibilities, and roles of citizens.	1.80	3.40	0.45	0.89	90.00	Yes
Different forms of government and international organizations and their influence in the world community.	1.00	2.00	0.00	0.00	100.00	Yes
Percentage of GLEs with 50% of EEOs linked to at least one item:100%						

Table B-9. Range-of-Knowledge for CoAlt Social Studies, High School: Mean Percent of EEOs per GLE Linked with Items

Grade Level Expectation	Number of EEOs	Mean Items per GLE	Range of EEOs			Range-of-Knowledge Target Met
			EEOs with At Least One Item		% of Total EEOs per GLE	
			M	*SD*		
Use the historical method of inquiry to ask questions, evaluate primary and secondary sources, critically analyze and interpret data, and develop interpretations defended by evidence.	1.00	2.00	0.00	0.00	100.00	Yes
The key concepts of continuity and change, cause and effect, complexity, unity and diversity over time.	2.20	3.60	0.45	0.55	73.33	Yes
The significance of ideas as powerful forces throughout history.	1.00	1.00	0.00	0.00	100.00	Yes
Use different types of maps and geographic tools to analyze features on Earth to investigate and solve geographic questions.	1.00	3.20	0.00	0.45	100.00	Yes
Explain and interpret geographic variables that influence the interactions of people, places and environments.	1.00	2.00	0.00	0.00	100.00	Yes
The interconnected nature of the world, its people and places.	1.00	1.80	0.00	0.45	100.00	Yes
Productive resources – natural, human, capital – are scarce; therefore, choices are made about how individuals, businesses, governments, and societies allocate these resources.	1.00	1.00	0.00	0.00	100.00	Yes
Economic policies affect markets.	1.00	1.20	0.00	0.45	100.00	Yes
Government and competition affect markets.	1.00	1.00	0.00	0.00	100.00	Yes
Design, analyze, and apply a financial plan based on short- and long-term financial goals (PFL).	1.00	1.00	0.00	0.00	100.00	Yes
Analyze strategic spending, saving, and investment options to achieve the objectives of diversification, liquidity, income, and growth (PFL).	1.00	1.00	0.00	0.00	100.00	Yes
The components of personal credit to manage credit and debt (PFL).	1.00	1.00	0.00	0.00	100.00	Yes
Identify, develop, and evaluate risk-management strategies (PFL).	1.00	1.00	0.00	0.00	100.00	Yes
Research, formulate positions, and engage in appropriate civic participation to address local, state, and national issues or policies.	1.25	1.25	0.50	0.50	62.50	Yes
Purposes of and limitations on the foundations, structures and functions of government.	2.00	2.20	0.00	0.45	100.00	Yes
Analyze how public policy - domestic and foreign - is developed at the local, state, and national levels and compare how policy-making occurs in other forms of government.	1.20	1.80	0.45	0.45	60.00	Yes
Percentage of GLEs with 50% of EEOs linked to at least one item:100%						

Balance-of-Knowledge Representation

The results for Balance-of-Knowledge representation for the grades 4 and 7 and high school CoAlt social studies tests are presented below. The tables also include the percentage of items linked to each strand. The minimum acceptable balance index is 70 out of 100.

Table B-10. Balance-of-Knowledge Representation for CoAlt Social studies Grade 4: Mean Balance Index per GLE

Grade Level Expectation	EEOs per GLE	Balance-of-Knowledge Representation					Balance Index Target Met
		Mean EEOs Linked with Items	Mean Items per GLE	Mean % of Items (of total) Linked to GLE	Mean Balance Index		
		M	*M*	*M*	*M*	*SD*	
Organize and sequence events to understand the concepts of chronology and cause and effect in the history of Colorado.	1	1.00	1.50	9.01	100.00	0.00	Yes
The historical eras, individuals, groups, ideas and themes in Colorado history and their relationships to key events in the United States.	1	1.00	1.00	6.00	100.00	0.00	Yes
Use several types of geographic tools to answer questions about the geography of Colorado.	2	1.00	2.00	12.01	100.00	0.00	Yes
Connections within and across human and physical systems are developed.	2	1.67	2.33	13.97	94.44	9.62	Yes
People respond to positive and negative incentives.	1	1.00	3.33	19.98	100.00	0.00	Yes
The relationship between choice and opportunity cost (PFL).	1	1.00	3.00	18.01	100.00	0.00	Yes
Analyze and debate multiple perspectives on an issue.	1	1.00	2.00	12.01	100.00	0.00	Yes
The origins, structure, and functions of the Colorado government	1	1.00	2.00	12.01	100.00	0.00	Yes
Total	10						
Percentage of GLEs with a balance of representation index of 70 or greater: 100%							

Table B-11. Balance-of-Knowledge Representation for CoAlt Social studies Grade 7: Mean Balance Index per GLE

Grade Level Expectation	EEOs per GLE	Balance-of-Knowledge Representation					Balance Index Target Met
		Mean EEOs Linked with Items	Mean Items per GLE	Mean % of Items (of total) Linked to GLE	Mean Balance Index		
		M	*M*	*M*	*M*	*SD*	
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.	2	1.60	2.80	16.47	90.00	9.13	Yes
The historical eras, individuals, groups, ideas and themes within regions of the Eastern Hemisphere and their relationships with one another	1	1.00	2.20	12.94	100.00	0.00	Yes
Use geographic tools to gather data and make geographic inferences and predictions.	1	1.00	2.00	11.76	100.00	0.00	Yes
Regions have different issues and perspectives.	1	1.00	2.00	11.76	100.00	0.00	Yes
Supply and demand influence price and profit in a market economy.	2	1.80	2.00	11.76	100.00	0.00	Yes
The distribution of resources influences economic production and individual choices (Economics and PFL).	2	1.00	1.00	5.88	100.00	0.00	Yes
Compare how various nations define the rights, responsibilities, and roles of citizens.	2	1.80	3.40	20.00	88.00	7.30	Yes
Different forms of government and international organizations and their influence in the world community.	1	1.00	2.00	11.76	100.00	0.00	Yes
Total	12						
Percentage of GLEs with a balance of representation index of 70 or greater: 100%							

Table B-12. Balance-of-Knowledge Representation for CoAlt Social studies High School: Mean Balance Index per GLE

Grade Level Expectation	EEOs per GLE	Balance-of-Knowledge Representation					Balance Index Target Met
		Mean EEOs Linked with Items	Mean Items per GLE	Mean % of Items (of total) Linked to GLE	Mean Balance Index		
		M	*M*	*M*	*M*	*SD*	
Use the historical method of inquiry to ask questions, evaluate primary and secondary sources, critically analyze and interpret data, and develop interpretations defended by evidence.	1	1.00	2.00	7.75	100.00	0.00	Yes
The key concepts of continuity and change, cause and effect, complexity, unity and diversity over time.	3	2.20	3.60	13.94	80.00	4.56	Yes
The significance of ideas as powerful forces throughout history.	1	1.00	1.00	3.88	100.00	0.00	Yes
Use different types of maps and geographic tools to analyze features on Earth to investigate and solve geographic questions.	1	1.00	3.20	12.40	100.00	0.00	Yes
Explain and interpret geographic variables that influence the interactions of people, places and environments.	1	1.00	2.00	7.75	100.00	0.00	Yes
The interconnected nature of the world, its people and places.	1	1.00	1.80	6.98	100.00	0.00	Yes
Productive resources – natural, human, capital – are scarce; therefore, choices are made about how individuals, businesses, governments, and societies allocate these resources.	1	1.00	1.00	3.88	100.00	0.00	Yes
Economic policies affect markets.	1	1.00	1.20	4.65	100.00	0.00	Yes
Government and competition affect markets.	1	1.00	1.00	3.88	100.00	0.00	Yes
Design, analyze, and apply a financial plan based on short- and long-term financial goals (PFL).	1	1.00	1.00	3.88	100.00	0.00	Yes
Analyze strategic spending, saving, and investment options to achieve the objectives of diversification, liquidity, income, and growth (PFL).	1	1.00	1.00	3.88	100.00	0.00	Yes
The components of personal credit to manage credit and debt (PFL).	1	1.00	1.00	3.88	100.00	0.00	Yes

Grade Level Expectation	EEOs per GLE	Balance-of-Knowledge Representation					Balance Index Target Met
		Mean EEOs Linked with Items	Mean Items per GLE	Mean % of Items (of total) Linked to GLE	Mean Balance Index		
		M	*M*	*M*	*M*	*SD*	
Identify, develop, and evaluate risk-management strategies (PFL).	1	1.00	1.00	3.88	100.00	0.00	Yes
Research, formulate positions, and engage in appropriate civic participation to address local, state, and national issues or policies.	2	1.25	1.25	4.85	100.00	0.00	Yes
Purposes of and limitations on the foundations, structures and functions of government.	2	2.00	2.20	8.52	96.67	7.45	Yes
Analyze how public policy - domestic and foreign - is developed at the local, state, and national levels and compare how policy-making occurs in other forms of government.	2	1.20	1.80	6.98	100.00	0.00	Yes
Total	21						
Percentage of GLEs with a balance of representation index of 70 or greater: 100%							

EEOs Matched to Items by Panelists

Tables B-13 through B-15 present the EEOs, along with the mean number of items, matched by panelists. Column 1 presents the HumRRO code corresponding to each of the EEOs. One note of caution when reading these tables, the same items may not be represented by the mean number of items. For example, EEO code '1.1.a' in the first row shows that 7 panelists matched a mean number of 2.14 items to this EEO. This does not mean/assume that the items matched to the EEO by the panelists were the same items across panelists.

Table B-13. Grade 4 CoAlt Social Studies: EEOs Matched to Items by Panelists

HumRRO EEO Code	Number of Panelists	Mean Number of Items per EEO	*SD*
1.1.a	2	1.50	0.71
1.2.a	3	1.00	0.00
2.1.a	3	2.00	0.00
2.1.b	0	0.00	0.00
2.2.a	2	1.00	0.00
2.2.b	3	1.67	0.58
3.1.a	3	3.33	0.58
3.2.a	3	3.00	0.00
4.1.a	3	2.00	0.00
4.2.a	3	2.00	0.00

Table B-14. Grade 7 CoAlt Social Studies: EEOs Matched to Items by Panelists

HumRRO EEO	Number of Panelists	Mean Number of Items per EEO	*SD*
1.1.a	5	2.20	0.45
1.1.b	3	1.00	0.00
1.2.a	5	2.20	0.45
2.1.a	5	2.00	0.00
2.2.a	5	2.00	0.00
3.1.a	5	1.20	0.45
3.1.b	4	1.00	0.00
3.2.a	5	1.00	0.00
3.2.b	0	0.00	0.00
4.1.a	4	1.75	0.50
4.1.b	5	2.00	1.00
4.2.a	4	2.00	0.00

Table B-15. High School CoAlt Social Studies: EEOs Matched to Items by Panelists

HumRRO EEO	Number of Panelists	Mean Number of Items per EEO	*SD*
1.1.a	5	2.00	0.00
1.2.a	4	1.75	0.96
1.2.b	3	1.00	0.00
1.2.c	4	2.00	0.82
1.3.a	5	1.00	0.00
2.1.a	5	3.20	0.45
2.2.a	5	2.00	0.00
2.3.a	5	1.80	0.45
3.1.a	5	1.00	0.00
3.2.a	5	1.20	0.45
3.3.a	5	1.00	0.00
3.4.a	5	1.00	0.00
3.5.a	5	1.00	0.00
3.6.a	5	1.00	0.00
3.7.a	5	1.00	0.00
4.1.a	2	1.00	0.00
4.1.b	3	1.00	0.00
4.2.a	5	1.20	0.45
4.2.b	5	1.00	0.00
4.3.a	4	1.50	0.58
4.3.b	2	1.50	0.71

Appendix C.
Panelist Instructions and Extended Evidence Outcome Examples

Panelists received the following Panelist Instruction and Extended Evidence Outcomes sheets as reference materials corresponding with verbal instructions from HumRRO facilitators. Panelist materials were provided for each grade level; however, only example of both documents from Science and Social Studies are included in this appendix.

CoAlt Science Alignment Process
Panelist Instructions for Grade 5

	Rating Task	Documents Needed	File Format
1	CoAlt Science Extended Evidence Outcomes (EEOs) (Consensus)	Science G5 Panelist Instructions	Print copy
		Science G5 EEO Consensus	Print copy
		Science G5 EEO Consensus	Excel
2	CoAlt Science Items (Individual)	Science G5 Panelist Instructions	Print copy
		Science G5 Extended Evidence Outcomes	Print copy
		Grade 5 Science Items	Print copy
		Science G5 Item Rating	Excel

Prior to alignment steps, train:

(1) Review handouts, particularly the CoAlt Panelist Instructions
(2) Access HumRRO item rating forms:
    a. Locate form on desktop, double click to open.
    b. "Save As" the file name and add **underscore and your 3 initials** (e.g., Science G5 Item Rating_***eas***).

1 Review CoAlt EEOs and provide Depth of Knowledge (DOK) rating (Consensus)

Train Task:
(1) Receive the Science G5 EEO Consensus paper copy.
    a. You will handwrite your DOK rating on this form.
(2) Make DOK ratings
    a. The facilitator will discuss the 4 DOK levels and will ask for a volunteer to record the panel's ratings in the G5 EEO Consensus Excel form. See the Support Materials section in this document for the information. Refer to this section as needed.

Conduct Task:
(1) Provide individual ratings on the paper copy.
(2) Determine if everyone provided the same rating. If not, share your reasons for your rating.
(3) The group will come to a consensus on the rating and majority will rule if necessary.
(4) The volunteer will enter the group's consensus rating in the Science G5 EEO Consensus Excel form.

2 Rate CoAlt Science Items

Train Task:

(1) You will review CoAlt test items, assign a DOK level, select the EEO that the item is targeting, and provide ratings regarding the linkage.
(2) The facilitator will discuss the columns in the Excel form.
    a. Columns B and C: The item sequence number and UIN
    b. Column D: Assign the DOK level
    c. Column E-H: Item Linkage and Overall Alignment
        - E: Select the grade level EEO that best covers the content measured by the item
        - F: Indicate how well the content measured by the item aligns (matches or links) with the selected EEO using the following rating scale.

Rating	Overall Alignment for Item and EEO Rating Descriptions
1	Not aligned to any EEO (No EEO was entered in column C)
2	Weakly aligned (item does not assess the content of the EEO well)
3	Highly aligned (item assesses EEO core content reasonably well)
4	Fully aligned (item assesses content that clearly matches with the EEO)

        - G and H: If you rate the overall alignment as 1 or 2, describe exactly what content in the item is not covered by the EEO. Provide a secondary EEO if you feel the item equally assesses another EEO.

Conduct the Task:

(1) Save the Science G5 Item Rating file on desktop with your 3 initials.
(2) Rate 2 or so (facilitator will determine) items independently (DOK, alignment), then conduct calibration discussion.
(3) Conduct individual ratings for each item in order. No consensus discussions.
(4) Save the file regularly!!!

Support Materials

DOK Definitions

Level	DOK Description
1	**None**: No content clearly measured; too vague
2	**Attention**: Requires students to display ability to acknowledge, reply, and respond to text or related subject features. Examples: Attends to pictures/symbols pertinent to a story or attends while teacher reads subject related text. (touch, look, vocalize, respond, attend)
3	**Memorize/recall**: Requires the ability to recite or recall facts or information. It involves the ability to distinguish between simple text-based and one-step procedures. Examples: Indicates understanding of new words or recalls basic ideas in passages via speech, writing, or signs. (list, describe (facts), identify, state, define, label, recognize, record, match, recall, relate)
4	**Performance**: Requires students to use recalled facts or information for simple tasks. Example: Retell information taken from printed materials. (perform, demonstrate, follow, count, locate, read)
5	**Comprehension**: Requires processing beyond recall and observation and may require both understanding and subsequent processing of text. It involves ordering, classifying, estimating text or numbers as well as identifying patterns, main points, or two-step procedures. Example: Draw a line through parts of passage with errors (capitalization or grammar) (explain, conclude, group/categorize, restate, review, translate, describe (concepts), paraphrase, infer, summarize, illustrate)
6	**Application**: Show ability to go beyond text; to reason, plan, or use of evidence to connect ideas. Students will use text, data, or observations to draw conclusions or solve non-routine problems. Example: Which of the following conclusions is best supported by information from the passage? (compute, organize, collect, apply, classify, construct, solve, use, order, develop, generate, interact with text, implement)
7	**Analysis, Synthesis, Evaluation**: Requires extended higher order processing. It typically requires extended time to complete a task, but the time is not spent on repetitive tasks. It involves taking information and applying this information to a new task; which may require generating a hypothesis, perform complex analyses, or make connections among different texts. Example: You will become a storyteller and will research and write the story of a Southerner who has moved to the North after the Civil War. (pattern, analyze, compare, contrast, compose, predict, extend, plan, judge, evaluate, interpret, cause/effect, investigate, examine, distinguish, differentiate, generate)

CoAlt Social Studies Alignment Process
Panelist Instructions for Grade 7

	Rating Task	Documents Needed	File Format
1	CoAlt Social Studies Extended Evidence Outcomes (EEOs) (Consensus)	Social Studies G7 Panelist Instructions	Print copy
		Social Studies G7 EEO Consensus	Print copy
		Social Studies G7 EEO Consensus	Excel
2	CoAlt Social Studies Items (Individual)	Social Studies G7 Panelist Instructions	Print copy
		Social Studies G7 Extended Evidence Outcomes	Print copy
		Grade 7 Social Studies Items	Print copy
		Social Studies G7 Item Rating	Excel

Prior to alignment steps, train:

(3) Review handouts, particularly the CoAlt Panelist Instructions
(4) Access HumRRO item rating forms:
   a. Locate form on desktop, double click to open.
   b. "Save As" the file name and add **underscore and your 3 initials** (e.g., Social Studies G7 Item Rating_***eas***).

1 Review CoAlt EEOs and provide Depth of Knowledge (DOK) rating (Consensus)

Train Task:

(3) Receive the Social Studies G7 EEO Consensus paper copy.
   a. You will handwrite your DOK rating on this form.
(4) Make DOK ratings
   a. The facilitator will discuss the 4 DOK levels and will ask for a volunteer to record the panel's ratings in the G7 EEO Consensus Excel form. See the Support Materials section in this document for the information. Refer to this section as needed.

Conduct Task:

(5) Provide individual ratings on the paper copy.
(6) Determine if everyone provided the same rating. If not, share your reasons for your rating.
(7) The group will come to a consensus on the rating and majority will rule if necessary.
(8) The volunteer will enter the group's consensus rating in the Social Studies G7 EEO Consensus Excel form.

2 Rate CoAlt Social Studies Items

Train Task:

(3) You will review CoAlt test items, assign a DOK level, select the EEO that the item is targeting, and provide ratings regarding the linkage.
(4) The facilitator will discuss the columns in the Excel form.
   a. Columns B and C: The item sequence number and UIN
   b. Column D: Assign the DOK level
   c. Column E-H: Item Linkage and Overall Alignment

- E: Select the grade level EEO that best covers the content measured by the item
- F: Indicate how well the content measured by the item aligns (matches or links) with the selected EEO using the following rating scale.

Rating	Overall Alignment for Item and EEO Rating Descriptions
1	Not aligned to any EEO (No EEO was entered in column C)
2	Weakly aligned (item does not assess the content of the EEO well)
3	Highly aligned (item assesses EEO core content reasonably well)
4	Fully aligned (item assesses content that clearly matches with the EEO)

- G and H: If you rate the overall alignment as 1 or 2, describe exactly what content in the item is not covered by the EEO. Provide a secondary EEO if you feel the item equally assesses another EEO.

Conduct the Task:

(5) Save the Social Studies G7 Item Rating file on desktop with your 3 initials.
(6) Rate 2 or so (facilitator will determine) items independently (DOK, alignment), then conduct calibration discussion.
(7) Conduct individual ratings for each item in order. No consensus discussions.
(8) Save the file regularly!!!

Support Materials

DOK Definitions

Level	DOK Description
0	**None**: No content clearly measured; too vague
1	**Attention**: Requires students to display ability to acknowledge, reply, and respond to text or related subject features. Examples: Attends to pictures/symbols pertinent to a story or attends while teacher reads subject related text. (touch, look, vocalize, respond, attend)
2	**Memorize/recall**: Requires the ability to recite or recall facts or information. It involves the ability to distinguish between simple text-based and one-step procedures. Examples: Indicates understanding of new words or recalls basic ideas in passages via speech, writing, or signs. (list, describe (facts), identify, state, define, label, recognize, record, match, recall, relate)
3	**Performance**: Requires students to use recalled facts or information for simple tasks. Example: Retell information taken from printed materials. (perform, demonstrate, follow, count, locate, read)
4	**Comprehension**: Requires processing beyond recall and observation and may require both understanding and subsequent processing of text. It involves ordering, classifying, estimating text or numbers as well as identifying patterns, main points, or two-step procedures. Example: Draw a line through parts of passage with errors (capitalization or grammar) (explain, conclude, group/categorize, restate, review, translate, describe (concepts), paraphrase, infer, summarize, illustrate)
5	**Application**: Show ability to go beyond text; to reason, plan, or use of evidence to connect ideas. Students will use text, data, or observations to draw conclusions or solve non-routine problems. Example: Which of the following conclusions is best supported by information from the passage? (compute, organize, collect, apply, classify, construct, solve, use, order, develop, generate, interact with text, implement)
6	**Analysis, Synthesis, Evaluation**: Requires extended higher order processing. It typically requires extended time to complete a task, but the time is not spent on repetitive tasks. It involves taking information and applying this information to a new task; which may require generating a hypothesis, perform complex analyses, or make connections among different texts. Example: You will become a storyteller and will research and write the story of a Southerner who has moved to the North after the Civil War. (pattern, analyze, compare, contrast, compose, predict, extend, plan, judge, evaluate, interpret, cause/effect, investigate, examine, distinguish, differentiate, generate)

Panelists received a copy of the Colorado Extended Evidence Outcomes (EEOs) for either Science or Social Studies as a reference for data entry into rating forms. Only a portion of the coded EEOs for Science grade 5 and Social Studies grade 4 are provided as examples.

Standard	Concepts & Skills	Extended Evidence Outcomes	HumRRO ID
Physical Science	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.	Separate simple mixtures based on physical properties	1.1.a
		Demonstrate that the weight of a mixture of solid objects before and after it is separated into parts is the same	1.1.b
Life Science	All organisms have structures and systems with separate functions.	Compare and contrast physical characteristics in plants and animals (plant/plant, animal/animal)	2.1.a
		Sort animals by observable characteristics based on a given group (birds, reptiles, insects and mammals)	2.1.b
		Identify how living organisms attain basic needs for survival	2.1.c
	Human body systems have basic structures, functions, and needs.	Identify the function of the main internal organs of the body	2.2.a
		Describe ways to maintain a healthy body	2.2.b
Earth Systems Science	Earth and Sun provide a diversity of renewable and nonrenewable resources.	Identify Earth's resources (water, wind, and some fossil fuels such as coal, gas, solar)	3.1.a
		Identify ways to conserve resources (turn off lights, turn off water when brushing teeth)	3.1.b
		Distinguish between renewable and nonrenewable resources	3.1.c

Extended Evidence Outcomes for Social Studies Grade 4 Example

Standard	Concepts & Skills	Extended Evidence Outcomes	HumRRO ID
History	Organize and sequence events to understand the concepts of chronology and cause and effect in the history of Colorado.	Indicate one to three factors that affected the growth of Colorado (i.e. mining, farming, transportation, natural resources)	1.1.a
	The historical eras, individuals, groups, ideas and themes in Colorado history and their relationships to key events in the United States.	Identify significant artifacts related to Colorado history (e.g. cliff dwellings, covered wagons, mining tools, trains)	1.2.a
Geography	Use several types of geographic tools to answer questions about the geography of Colorado.	Identify features on a map of Colorado (i.e. mountains, river, plains, lakes)	2.1.a
		Create or illustrate features on a map of Colorado	2.1.b
	Connections within and across human and physical systems are developed.	Recognize that people live together in the same location (settlement)	2.2.a
		Identify how Colorado communities are connected (e.g. roads, railroads, airways, waterways)	2.2.b
Economics	People respond to positive and negative incentives.	Identify types of goods and services native to Colorado (e.g., tourism, steel, mining, agriculture, etc.)	3.1.a
	The relationship between choice and opportunity cost (PFL).	Demonstrate an understanding of the value of items (e.g. $1 can buy gum not a car)	3.2.a

Appendix D.
Item Rating Form Examples

Panelists used laptops for data entry into rating forms. Examples of each are provided.

CoAlt Science Item Review for Grade 5

			Item DOK	Item Linkage and Overall Alignment			
	Item Number	UIN	Depth Of Knowledge	Linked EEO	Overall Alignment	Explanation	EEO 2
			1-None 2-Attention 3-Memorize/Recall 4-Performance 5-Comprehension 6-Application 7-Analysis, Synthesis, Evaluation	Enter EEO ID Code	1- Not aligned 2- Weakly aligned 3- Highly aligned 4- Fully aligned	If not highly or fully aligned, describe what the item measures that does not match with the EEO	Enter Secondary EEO ID Code
1	1	COSC05S223					
2	2	COSC05S100					
3	4	COSC05S109					

CoAlt Social Studies Item Review for Grade 7

			Item DOK	Item Linkage and Overall Alignment			
	Item Number	UIN	Depth Of Knowledge	Linked EEO	Overall Alignment	Explanation	EEO 2
			1-None 2-Attention 3-Memorize/Recall 4-Performance 5-Comprehension 6-Application 7-Analysis, Synthesis, Evaluation	Enter EEO ID Code	1- Not aligned 2- Weakly aligned 3- Highly aligned 4- Fully aligned	If not highly or fully aligned, describe what the item measures that does not match with the EEO	Enter Secondary EEO ID Code
1	1	COSS07S108					
2	2	COSS07S114					
3	4	COSS07S120					