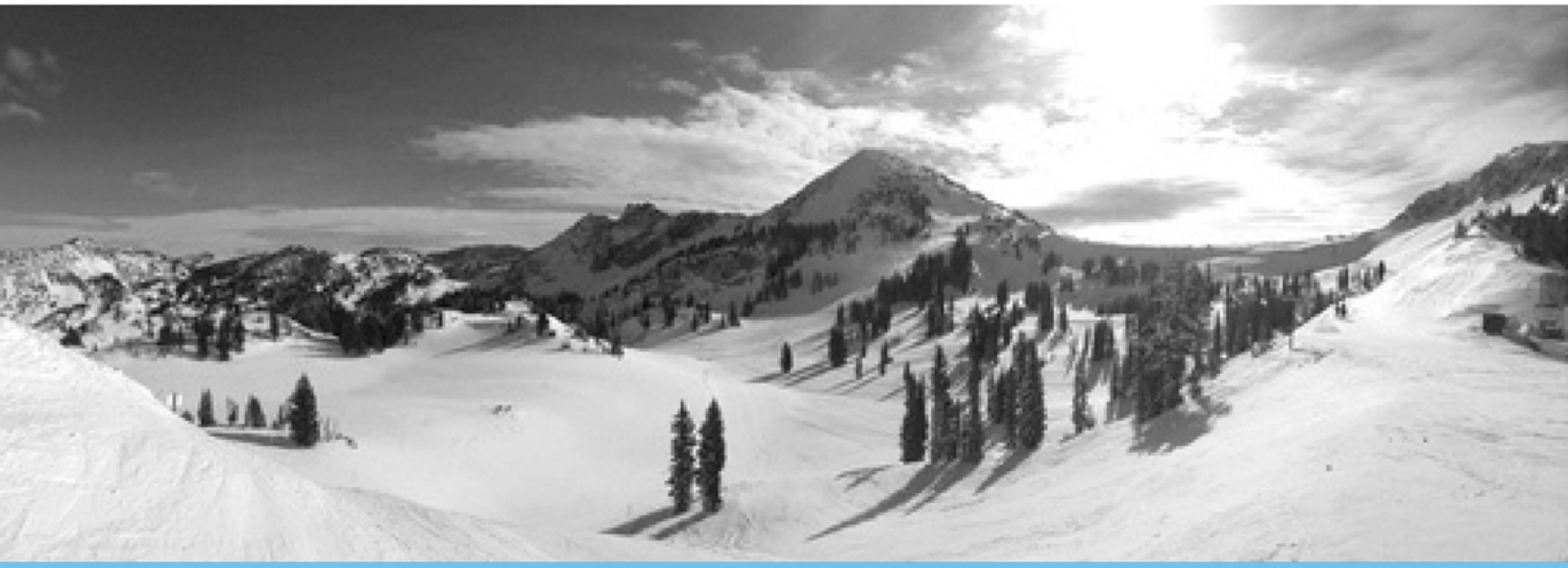
Colorado Alternate Assessment Program



Technical Report

Science

2021

Colorado Alternate Assessment (CoAlt): Science Technical Report 2020–2021

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 *4*
Overview of CoAlt: Science *5*
Assessment Development Partners *7*
CHAPTER 2: ITEM DEVELOPMENT AND ITEM BANKING 9
Item Development Process *9*
Item Reviews *10*
Item Banking Process *12*
CHAPTER 3: TEST CONSTRUCTION 13
CHAPTER 4: TEST ADMINISTRATION PROCEDURES 15
Manuals *15*
Training *15*
Accessibility and Accommodations *16*
Test Security *17*
Score Monitoring *18*
CHAPTER 5: SCORING THE ASSESSMENTS 19
CHAPTER 6: STANDARD SETTING 21
CHAPTER 7: REPORTING 22
Description of Scores *22*
Score Reports *23*
CHAPTER 8: CALIBRATION, EQUATING, AND SCALING 24
Calibration *24*
Equating and Scaling *24*
Steps in the Calibration and Scaling Process *25*
CHAPTER 9: RELIABILITY 27
Cronbach's Alpha *27*
Standard Error of Measurement *28*
Conditional Standard Error of Measurement *28*
Decision Consistency and Accuracy *28*
Inter-Rater Agreement *29*
CHAPTER 10: VALIDITY 31
Sources of Validity Evidence *31*
Fairness *35*
CHAPTER 1: OPERATIONAL ITEMS 38
Administration Summary *38*
Calibration Results *38*
Performance Results *38*
Validity Evidence *39*

REFERENCES **40**

CoAlt: Science Tables 7–28 ..... **42**

Appendices **73**

APPENDIX A: COALT: SCIENCE ELIGIBILITY GUIDELINES 74
APPENDIX B: COALT: SCIENCE TEST BLUEPRINTS 77
APPENDIX C: COALT: SCIENCE SAMPLE SCORE REPORT 79

PART I: HISTORICAL OVERVIEW AND SUMMARY OF PROCESSES

CHAPTER 1: INTRODUCTION AND BACKGROUND

Requirements

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 ability or language background, must be provided each year 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 eligible students, these end-of-year 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. The EEOs were adopted and incorporated into the CAS in August 2011.

In 2015, Colorado passed legislation (C.R.S. §22-7-1013 (8) (a-c)) that allows for parents to excuse their child(ren) from testing.

During the 2020-2021 school year, Colorado received a partial waiver of spring 2021 federal assessment requirements from the U.S. Department of Education (USED) due to COVID-19 conditions in Colorado. The partial waiver solely addressed CMAS/CoAlt English language arts (ELA), mathematics and Science. Colorado administers the Dynamic Learning Maps consortium assessments for the CoAlt ELA and mathematics which are not covered by this technical report. CoAlt Science assessments were administered to students in grade 8 and high school in spring 2021.

The USED waived the requirement to administer the CoAlt (and CMAS) Science assessments to students in grade 5. The waiver included a requirement for all grades and content areas to be publicly reported as long as the minimum number of students and student data privacy requirements are met. These adjustments to testing, along with the suspension of Colorado's social studies assessments, were also made for spring 2021 by the Colorado legislature. The table below includes the CoAlt Science grade levels that were administered across Colorado in spring 2021.

Content Area	2021 Required Test	2021 Not Administered
Science	Grade 8 and High School	Grade 5
Social Studies		Grades 4 and 7

Intended Population

The CoAlt: Science assessments are designed for students with the most significant cognitive disabilities. These students have 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—Colorado Measures of Academic Success (CMAS) Science—even with accommodations may take CoAlt: Science. 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 assessments is determined by a student's Individualized Education Program (IEP) team. The IEP team will determine whether a student should participate in CoAlt: Science or CMAS Science 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 the CoAlt: Science assessments are the most appropriate assessments for the student if the student meets all 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: Science 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/cdesped/accommodationsmanual_participationguidelinesworksheet.

Purpose of the Document

The purpose of the *CoAlt: Science 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 2021 CoAlt: Science assessments, including content, assessment development, administration, scoring, and technical attributes.

The *CoAlt: Science 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 2021 administration operational items.

Overview of CoAlt: Science

Purpose of CoAlt: Science

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 area of science, Colorado developed the CoAlt: Science assessments.

The goals of the Colorado Assessment System, including the CoAlt: Science 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: Science promotes improved instruction toward grade-level expectations, growth over time toward independent performance, and high expectations toward achievement in the content areas.

Description of CoAlt: Science

The CoAlt: Science assessments are academic, standards-based assessments 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 area of science. The EEOs are alternate academic standards that describe what students taking CoAlt: Science are expected to know and be able to demonstrate at each grade level in Science.

The test design of CoAlt: Science was developed to provide this unique population of students with an opportunity to demonstrate their knowledge of the EEOs. The CoAlt: Science assessments include paper-based test books used by the Test Examiner to administer test items to the students. The test books are designed to sit on the table allowing the Test Examiner to read the question and answer choices to the student while at the same time allowing the student to view the answer choices. The test books include 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 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 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 (i.e., placing option cards in designated boxes within a chart or diagram). Students may manipulate the option cards independently or indicate the desired placement to the Test Examiner through their preferred mode of expressive communication, such as verbal directions or eye gaze. 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. This item type allows students to demonstrate their knowledge and skills to create a product, revealing a different level of understanding of specific concepts and skills than that which are demonstrated through SR items alone.

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

The CoAlt: Science assessments were administered in the following grades in spring 2021:

- Science: grades 8 and 11

The CoAlt: Science assessments are administered during an assessment window. There is no minimum or maximum testing time for administration of the assessments. Testing may extend over multiple days for a student. The assessment may be stopped and restarted at any time; however, once an item is presented, the item should be completed before stopping. The amount of time it takes the student to complete the assessment is recorded by the Test Examiner and entered into PearsonAccess^{next} after testing is complete.

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 Science EEOs can be found online at the following location:
http://www.cde.state.co.us/CoExtendedEO/StateStandards

CoAlt: Science Assessment Frameworks were developed to better identify the content standards that may be assessed on the CoAlt: Science 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 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 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.

Item development for the CoAlt: Science assessments began in summer 2012. The newly-developed items were then administered in a stand-alone field test in spring 2013 for Elementary/Middle School and in fall 2013 for High School. 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 on future operational assessments.

Following the first operational administration of the Elementary/Middle School assessments in spring 2014, performance standards (i.e., cut scores) were set and the final cut scores were approved and used for reporting purposes. The same process was undertaken for the High School Science assessment following the first operational administration in fall 2014. The Colorado State Board of Education formally adopted CoAlt: Science performance standards between 2014 and 2016.

Assessment Development Partners

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

Colorado Department of Education

CDE is responsible for implementing state and federal education laws. CDE's Assessment Unit works closely with Colorado school districts, educators, community stakeholders, and assessment development partners to develop and administer the state assessments. CDE focuses on creating assessments that serve students, schools, districts, and the community while complying with state and federal legal requirements. CDE content, assessment administration, special populations, technology, data and psychometric staff works closely with Pearson on each facet of the assessment. CDE serves as the ultimate approver of services and products provided.

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

As the primary contractor responsible for end-to-end of assessment cycle services and products, Pearson worked closely with CDE throughout the CoAlt: Science development and administration processes. This included item and test development, forms creation, enrollment, packaging and distribution, test delivery, scoring, customer service, standard setting, score reporting, and psychometric services.

Colorado Technical Advisory Committee (TAC)

The Colorado TAC was comprised of psychometric, assessment, and special populations experts tasked with providing high-level consulting and expert advice regarding validity and reliability issues. Topics for which the TAC has provided input included blueprint design, scoring, scaling and equating, score reporting, and standard setting. The TAC included the following members:

- 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. Suzanne Lane, Professor, University of Pittsburgh
- 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 follows the same test development process as CMAS Science. However, the CoAlt: Science test development process reflects the unique characteristics of the assessment program, specifically the item types included on 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 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 item and test development:

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

Pearson's proprietary software, ABBI (Assessment Banking and Building solutions for Interoperable assessments), is used to support the item and test development processes. As described in the following sections, items can be moved into different statuses in ABBI, each status represents a step in the item development process.

Item Development Process

The item development process was a tiered, inter-related process that began with the development of the test blueprints for each grade level within each subject. The item development process continued with designing IDPs to forecast the targeted number of items needed to create a robust item bank that would be refreshed over time. Once written, the items went through multiple rounds of review, including contractor, department and Colorado educator content, bias and data reviews.

Specifications Development

CDE and Pearson collaboratively developed the CoAlt: Science test blueprints. The test blueprints contain the number of test items by content standard/reporting category, Grade Level Expectation (GLE), EEO, and item type. The CoAlt: Science test blueprints can be found in Appendix B. During this stage, Pearson also created IDPs which were designed to determine the number of items needed to construct the assessment based on the test blueprint requirements. The grade-level IDPs in each content area delineate the target number of items per content standard/reporting category, GLE, and EEO and helped to forecast the number of items that were

needed to create a robust operational item bank that would be refreshed over time. To construct the IDPs, the item bank was analyzed and gaps were identified. Each IDP was updated at the beginning of the item development cycle to inform item development targets that address item shortages.

Item Development

After the test blueprints and IDPs were developed, item writers began writing items. CoAlt: Science items were written by Colorado educators, content specialists, and professional item writers with guidance and input from CDE. SR and SPT items for each assessment were written to measure concepts and skills found in the EEOs. 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 CoAlt: Science items were written and entered into ABBI, 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 on 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 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 regarding 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 ABBI 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 was accurate, the EEO alignment was appropriate, the language was appropriate for the grade level and student population, and the graphics were clear and relevant to the item. Items that were accepted based on the CDE review were re-classified in ABBI as ready for bias and sensitivity review.

Accepted items were then reviewed by Colorado educators to evaluate whether the items were 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 ABBI as ready for field testing.

Data Review

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

- Classical statistics, such as the item sample size, item mean score, item-total correlation, and response distribution.
- IRT statistics, such as item difficulty and item fit values.
- DIF statistics by gender and ethnicity when group sample sizes are sufficient.

Student performance data were 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. Field test items that are accepted based on the evaluation of student performance data were re-classified in the item bank as available for use on future operational assessments. Items

that were rejected were re-classified to eliminate them from use on an operational test. These items may be modified and field tested again on future test forms.

Item Banking Process

Item banking is handled by Pearson's proprietary software, ABBI, which houses the items from creation through retirement in a secure environment. The secure item bank serves as the repository from which items for current and future test 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 on future assessments.

Item Bank Statistics

The metadata for each item are included in the item bank, which includes: the item image, test data, 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.

Custom reports can be generated out of ABBI. This feature allows CDE, Pearson assessment specialists, and psychometricians to generate Excel reports that capture metadata (e.g., unique item number, EEO, item type, depth of knowledge [DOK], associated stimulus, item status, item statistics and comments) useful for analyzing the item bank. ABBI is the source of reference for how and when changes to the item and the metadata have been implemented.

ABBI also supports the test construction process.

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 test 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 test blueprint and should include a variety of topics and contexts with specified psychometric targets.

The following guidelines are used during test 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 that the content within items is 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 test form meets the test blueprint and the test construction specifications (i.e., the required number of items, EEO coverage, and item types). The test form is then presented to psychometrics for analysis; the psychometrician verifies that the test form falls within the established psychometric and test blueprint parameters. The psychometric lead also identifies the anchor item set within each operational test form.

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

After the operational test form is approved, field test items are selected from the items in ABBI 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 test 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 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 8 Science	1	24	2	3	1	30	108
HS Science	1	23	3	3	1	30	110

CHAPTER 4: TEST ADMINISTRATION PROCEDURES

This chapter provides information related to the CoAlt: Science administration procedures. Prior to the administration, 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 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 to the appropriate individuals via manuals and in-person and recorded trainings as described below.

Manuals

Several manuals were created to support the CoAlt: Science administration, described in the following sections.

CoAlt: Science Test Examiner's Manual

This manual describes the procedures Test Examiners were to follow when administering the CoAlt: Science assessments. Prior to administering any CoAlt: Science assessment, Test Examiners were to carefully read this manual. Test administration policies and procedures, including scoring information, were to be followed as written so that all testing conditions were uniform statewide, ensuring every student in Colorado received the same standard directions and scoring during the administration of the test.

CMAS and CoAlt Procedures Manual

This manual provides instructions for the coordination of the CoAlt: Science assessments. Instructions include the protocols that all school staff were to follow related to test security and test administration and completing tasks like score entry. The manual also includes the tasks that were to be completed by District Assessment Coordinators (DACs), School Assessment Coordinators (SACs), District Technology Coordinators (DTCs), and data specialists before, during, and after test administration.

PearsonAccess^{next} Online User Guide

This guide provides guidance for DACs, SACs, DTCs, Test Examiners, and Student Enrollment personnel who utilize PearsonAccess^{next}.

Training

Administration training is intended to make sure all individuals involved in CoAlt: Science assessment activities at the school and district levels are prepared to follow administration processes and procedures with fidelity, as well as support adherence to security procedures. Fidelity to standardized test administration processes and procedures helps to ensure the comparability of resulting scores and accurate interpretation of results. Thorough trainings were conducted by CDE for DACs and district-based special education staff across Colorado. The in-person regional trainings contained information regarding proper procedures for administration.

Live online training sessions covered CoAlt: Science 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. Additionally, recorded versions of the live trainings were posted on the CDE Assessment Unit website. Administration training materials, including slide decks, manuals, and how-to guides were also available on the CDE Assessment Unit website for training SACs and Test Examiners. After CDE trained DACs and special education staff, these individuals trained SACs, Test Examiners, and any other individuals within the district who planned to participate in the 2021 administration.

Pearson customer service center staff were also trained to answer questions thoroughly and knowledgeably about the administration, and to escalate inquiries as necessary. A knowledge base of commonly asked questions was created to ensure accurate and consistent responses to school and district personnel. The knowledge base was created by the CDE Assessment Unit and Pearson Program Team based on information covered in the training materials and manuals. Revisions and additions were made to the knowledge base as needed. CDE met with Pearson daily during the administration window to review questions from districts and to ensure that appropriate answers were provided. Policy questions received by the Pearson customer service center were referred to the Department.

Accessibility and Accommodations

The CoAlt: Science 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: Science assessments and administration procedures. For example, CoAlt: Science assessments are read aloud to students and all students who take CoAlt: Science assessments are assessed individually. In addition, the assessments can be administered over several days for 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 assessments 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 assessments 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

Procedures described in this section were put in place to enhance the likelihood that security was maintained before, during, and after assessment administration. Materials used during the administration of the assessment were to be kept in locked storage locations when not under the direct supervision of Pearson or approved assessment coordinators and Test Examiners. All district and school personnel involved in the assessment administration were required to participate in annual local training on the CoAlt: Science assessments. DACs and district special education staff were responsible for overseeing training for the district, including verifying that the SACs were trained. SACs were responsible for ensuring that Test Examiners and all individuals involved in handling test materials at the school level were trained and subsequently acted in accordance with all security requirements. A chain of custody plan for materials was required to be written and implemented to ensure materials were securely distributed from DACs to SACs to Test Examiners and securely returned from Test Examiners to SACs and then to DACs. SACs were required to distribute materials to and collect materials from Test Examiners each day of testing, and securely store and deliver materials to DACs after testing was completed in accordance with the instructions in the *2021 CMAS and CoAlt Procedures Manual*.

All individuals involved in the administration of the assessments were required to sign a security agreement prior to handling test materials, which required them to follow all procedures set forth in the aforementioned manuals and prevented them from divulging the contents of the assessment, copying any part of the assessment, reviewing test questions with the students, allowing students to remove test materials from the room where testing was to take place, or interfering with the independent work of any student taking the assessment.

The PearsonAccess^{next} online administration platform used during the administration included permissions-based user role access to all information within the system including accessing student information, setting up student tests, entering scores, and accessing reports. Access to this information was tightly controlled before, during and after test administration, requiring a login ID and password to enter the system.

After all testing was completed at a school, used and unused materials were required to be securely stored and returned to the DAC by the district deadline for shipment to Pearson. DACs were required to report any missing test materials or test irregularities and to complete the appropriate documentation.

Score Monitoring

Pearson typically sends score monitors out to a small sample of schools to observe the administration of CoAlt: Science assessments. Monitors record a number of metrics including adherence to administration procedures and security measures. Due to COVID-19 conditions in Colorado, score monitoring was suspended for the spring 2021 administration.

CHAPTER 5: SCORING THE ASSESSMENTS

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

Selected Response Scoring Rubric

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

Table 2. Selected Response Scoring Rubric

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

Supported Performance Task Scoring Rubric

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

Table 3. Supported Performance Task Scoring Rubric

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

Additional Scoring Information

Test Examiners record all student scores within the test book or on the score recording form that is included with the task manipulatives set provided for each test. Recorded responses are then entered into PearsonAccess^{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 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 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). Only the performance level labels were updated, the Performance Level Descriptors (PLDs) and cut scores were not changed. Performance standards were set for grade 8 Science after the first operational administration in spring 2014. The full standard setting report for the Elementary/Middle School standard setting can be found in the *Spring 2014 CoAlt: Science and Social Studies Technical Report*. Performance standards were set for High School Science after the first operational administration in fall 2014. The full standard setting report for the High School Science standard setting 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 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 *CMAS and CoAlt Interpretive Guide 2021* (Colorado Department of Education, 2021), available at https://www.cde.state.co.us/assessment/cmas_coalt_interpretiveguide_2021.

Description of Scores

CoAlt: Science 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 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 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 assessments have four performance levels:

- Advanced
- At Target
- Approaching Target
- Emerging

These labels were updated in 2016 to match the levels used in the CoAlt: English Language Arts and Mathematics (DLM) assessments. The PLDs and cut scores were not changed. For those students who did not respond to any of the CoAlt: Science 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 at the content standard level (i.e. Physical Science, Life Science, and Earth Science). 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 for CoAlt: Science: student level and aggregate. A sample student performance report can be found in Appendix C.

Student Performance Reports

The Student Performance Report provides information about the performance of a student on the CoAlt: Science assessment. The student's scale score, associated performance level, and percent of points earned for each content standard are displayed on a one-page report along with comparative information related to state performance. In addition, performance level descriptions 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: Science:

- 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 Colorado school districts.

CHAPTER 8: CALIBRATION, EQUATING, AND SCALING

IRT was used to develop, calibrate, equate, and scale the CoAlt: Science 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 Elementary/Middle School assessments in spring 2014 and the High School 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 in the content areas, the Rasch Partial-Credit Model (RPCM) was used to place the CoAlt: Science 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

A pre-equating model was used for Science grades 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. All operational items on these forms had been previously calibrated and equated to the base scale. 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 equating model was chosen for the Science assessments due to anticipated lower than normal student participation due to COVID-19. It was determined that using pre-equating parameters based on a full student population would be more reliable than post-equating with the smaller sample (de Ayala, 2009).

Ability Estimates

After the pre-equated item parameter estimates were obtained for the Science 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 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 completed for each CoAlt: Science 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, Equating, and Scaling

- Operational Items
    - All operational items were already equated to the base scale.
    - Used Winsteps to estimate student abilities

- 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 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: Science, coefficient alpha estimates are provided for the overall test as well as for subgroups. The coefficient alpha estimates can be found in Tables 7–13.

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 7–13.

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 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 22–23.

Decision Consistency and Accuracy

The CoAlt: Science 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 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 24.

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: Science, 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 the CoAlt: Science administration. 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: Science 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. Due to COVID-19 conditions in Colorado, the score monitoring activity, which provides the second set of scores for inter-rater agreement, was not conducted for the spring 2021 administration.

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 assessments that are validated but rather the interpretations of the CoAlt scores. The purpose of the CoAlt: Science 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 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 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 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, an item goes through numerous reviews to confirm that it adequately aligns to the EEO that it is intended to measure. In addition, statistical bias analyses (i.e., DIF analyses) are conducted on field test items 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: Science in 2016. The alignment study was conducted to demonstrate that CoAlt: Science 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 items and the EEOs for science. 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 DOK ratings for the each of the EEOs for science, and (b) evaluating the science 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 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: Science 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. The results of the alignment study have been taken into account during the item development process for subsequent administrations.

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 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 are administering and scoring the assessment as expected, an inter-rater agreement study is conducted annually where external observers, called score monitors, visited schools to observe Test Examiners administering the assessment. The score monitors collect information such as how teachers administer the assessment and provide additional student-level score information that is used to evaluate the consistency of scoring. Though score monitoring was not conducted in spring 2021 due to COVID-19, these results will continue to be gathered when conditions are favorable for score monitors to enter schools.

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 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 scores and variables related to the student. Student scale scores were correlated with Test Examiners' 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 correlations between the student scores and these variables can be found in Tables 26–28.

Evidence for Validity and Consequences of Testing

Some of the intended consequences of the CoAlt: Science 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. Because the Elementary/Middle School assessments have been in use for 6 years (5 years for High School Science) there is now enough longitudinal data that could be used to investigate some of these consequences.

In 2021, participation in CoAlt Science was very low, 53% in 8th grade and 41% in High School. Combined with the potential disruptions in learning due to COVID-19 it is not appropriate to use the longitudinal trend data to evaluate the impact of the assessments.

The teacher survey included with assessments can also provide information on whether instruction trends have changed over time as a result of alternate testers being included in the assessment system. Two questions that have shown consistent correlations with test performance are the hours of instruction in the content area and the percent of the time spent in the general classroom.

Tables 4 and 5 show the percent of students receiving at least 4 hours of instruction per week in science and the percentage of students receiving at least 75% of their instruction in the general classroom. These measures had been increasing until 2019 but in 2021 decreased from the levels at the initial testing year (2014/2016). This is likely due to changes in the school environment caused by COVID-19 protocols. Students who were in school in person were generally kept in cohorts which would remove some flexibility for this population to move in and out of a general education classroom.

Table 4. Change in Hours of Classroom Instruction Per Week Survey Results Across Years

Subject	Grade	2014/2016 % of Students Receiving At Least 4 Hours of Instruction	2019 % of Students Receiving At Least 4 Hours of Instruction	2021% of Students Receiving At Least 4 Hours of Instruction	2019 % Change Students Receiving at Least 4 Hours of Instruction	2021% Change Students Receiving at Least 4 Hours of Instruction
Science	8	42.74	48.29	39.83	5.55	-2.91
	HS*	29.84	36.98	28.20	7.14	-1.64

*Administered in spring 2016

Table 5. Change in Classroom Instruction in the General Education Classroom Survey Results Across Years

Subject	Grade	2014/2016 % of Students Receiving 75% or More of Their Instruction in the General Education Classroom	2019 % of Students Receiving 75% or More of Their Instruction in the General Education Classroom	2021% of Students Receiving 75% or More of Their Instruction in the General Education Classroom	2019 % Change Students Receiving 75% or More of Their Instruction in the General Education Classroom	2021% Change Students Receiving 75% or More of Their Instruction in the General Education Classroom
Science	8	42.73	48.46	40.68	5.73	-2.05
	HS*	23.46	23.56	22.18	0.10	-1.28

*Administered in spring 2016

To further investigate the impact of testing on the appropriate use of the CoAlt: Science assessments, the percentage of the overall population identified as eligible for CoAlt: Science since the start of the assessments was evaluated. The federal government requires that alternate assessments be administered only to students with the most significant disabilities comprising of at most 1% of the student population. In 2015, Colorado passed legislation (C.R.S. §22-7-1013 (8) (a-c)) that allows for parents to excuse their child(ren) from testing, so this legislation can potentially impact subsequent student counts.

Looking at overall students counts in Table 6 (not valid scores because those counts are impacted by Parental Excuse after 2016), in 2014 in 4th grade, 1.12% of the student population was identified as CoAlt: Science eligible, 1.10% for 5th grade, 1.02% for 7th grade, 1.01% for 8th grade and 0.83% for High School Science. In all cases, students were being under or over identified. In 2021, only grade 8 and high school science were administered. In grade 8 the percent of the population identified as eligible for CoAlt is just under 1% which is in line with federal requirements. In High School, the percent of students identified is actually lower than 1% and similar to the percent in the first year of the assessment.

Table 6. Change in Student Population Identified as CoAlt: Science Eligible Across Years

		2014/2016			2021		
Subject	Grade	CoAlt Population	Total Population	CoAlt % of Population	CoAlt Population	Total Population	CoAlt % of Population
Science	8	644	63,730	1.01	655	68,327	0.96
	HS*	498	59,645	0.83	648	64,357	0.85

*Administered in spring 2016

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 assessments.

Universal Design

The CoAlt: Science 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 potential construct-irrelevant explanations for the flags.

Accessibility and Accommodations

As described in Chapters 3 and 4, the CoAlt: Science 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 the CoAlt: Science assessments and before each subsequent test administration.

Participation Considerations

Participation information must be reviewed and taken into consideration thoughtfully when interpreting 2021 district and school results. Participation rates for districts, schools and student groups were significantly lower overall than in past years. As participation rates decrease and vary across student, school and district groups, challenges with interpreting results increase. Depending on the specific school or district, some student groups were overrepresented in the spring 2021 results and others may have been underrepresented. Participation rates and how well the students who tested reflected the district/school as a whole varied greatly across the state in spring 2021. Participation information indicated that in some cases, conclusions should be drawn with caution or completely avoided. Due to these factors and many more challenges experienced during the pandemic, for some districts, schools and student groups, the spring 2021 data did not support all of the cross-state comparisons and uses made with prior years' data. However, where appropriate, the spring 2021 results were available to be used as a temperature check to better address and track future COVID-19 school recovery efforts.

The *2021 CMAS and CoAlt Interpretive Guide* includes additional considerations for parents and teachers in using the 2021 scores given the impact of COVID-19. In addition, a Participation Report was added for spring 2021 which provides a comparison of the demographic characteristics of their tested students compared to all students eligible for testing. This information can assist district in determining how to interpret their aggregated results.

PART II: STATISTICAL SUMMARIES

This section contains an overview of the statistical summaries for the spring 2021 administration. Administration summaries, calibration results, performance results, reliability evidence, and validity evidence are included for the operational items.

CHAPTER 1: OPERATIONAL ITEMS

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

Administration Summary

Approximately 620 students took the CoAlt: Science assessments. Tables 7–13 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 assessments ranged from 0.96 to 0.97. The SEM values for the total group ranged from 4.33 to 4.53.

Calibration Results

Item Statistics

Tables 14–15 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 16–17 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 and equating process.

Performance Results

The cuts scores, percent of students in each performance level, and the scale score ranges are provided in Tables 18–19. The scale score distributions for each assessment are shown in Tables 20–21. Tables 22–23 are provided and include the raw score, scale score, and CSEM values.

Decision Consistency and Accuracy

Table 24 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 25 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 26–28 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.

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 TABLES 7–28

Table 7. 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	8	Total	349	100.00%	146.35	37.56	0	236	79.12	24.26	0	107	0.97	4.53
		Female	134	38.40%	149.88	29.14	0	192	81.82	20.18	0	103	0.95	4.44
		Male	215	61.60%	144.14	41.87	0	236	77.43	26.39	0	107	0.97	4.57
		American Indian	2	0.57%	-	-	-	-	-	-	-	-	-	-
		Asian	9	2.58%	-	-	-	-	-	-	-	-	-	-
		Black or African American	19	5.44%	155.47	21.51	110	198	83.79	16.45	40	104	0.94	4.14
		Hispanic or Latino	159	45.56%	146.23	35.20	0	236	78.48	23.47	0	107	0.96	4.66
		White	143	40.97%	146.43	39.76	0	206	79.69	25.18	0	105	0.97	4.45
		Native Hawaiian or other Pacific Islander	2	0.57%	-	-	-	-	-	-	-	-	-	-
		Two or More Races	13	3.72%	-	-	-	-	-	-	-	-	-	-
		Missing	2	0.57%	-	-	-	-	-	-	-	-	-	-
	HS	Total	266	100.00%	154.77	31.40	0	250	85.44	22.52	0	110	0.96	4.34
		Female	92	34.59%	148.95	34.59	0	198	80.96	25.94	0	107	0.97	4.46
		Male	174	65.41%	157.86	29.21	0	250	87.82	20.17	0	110	0.96	4.27
		American Indian	1	0.38%	-	-	-	-	-	-	-	-	-	-
		Asian	3	1.13%	-	-	-	-	-	-	-	-	-	-
		Black or African American	20	7.52%	152.95	25.75	77	193	82.90	22.15	14	106	0.96	4.54
		Hispanic or Latino	115	43.23%	154.21	33.98	0	208	85.36	23.15	0	108	0.97	4.25
		White	115	43.23%	157.78	25.89	22	250	87.30	19.64	2	110	0.95	4.40
		Native Hawaiian or other Pacific Islander	0	0.00%	-	-	-	-	-	-	-	-	-	-
		Two or More Races	12	4.51%	-	-	-	-	-	-	-	-	-	-
		Missing	0	0.00%	-	-	-	-	-	-	-	-	-	-

Table 8. Descriptive Statistics by Economically Disadvantaged

	Grade	Subgroup	N	%	Scale Score				Raw Score				Alpha	SEM
					Mean	SD	Min	Max	Mean	SD	Min	Max		
Content	8	No	172	49.28%	143.34	39.49	0	206	76.87	25.95	0	105	0.97	4.61
		Yes	177	50.72%	149.27	35.45	0	236	81.29	22.35	0	107	0.96	4.45
	HS	No	147	55.26%	154.69	27.77	0	198	85.71	20.59	0	107	0.95	4.61
		Yes	119	44.74%	154.88	35.50	0	250	85.12	24.78	0	110	0.97	3.98

Table 9. 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	8	Language Proficiency	Not Applicable	255	50.72%	147.22	36.45	0	20	79.81	23.94	0	10	0.97	4.51
			NEP	39	73.07%	130.54	54.89	0	236	68.41	30.81	0	107	0.98	4.59
			LEP	9	11.17%	-	-	-	-	-	-	-	-	-	-
			PHLOTE	2	2.58%	-	-	-	-	-	-	-	-	-	-
			FELL	2	0.57%	-	-	-	-	-	-	-	-	-	-
			FEP Monitored Y1	14	0.57%	-	-	-	-	-	-	-	-	-	-
			FEP Monitored Y2	16	4.01%	153.75	24.75	75	183	84.00	20.30	15	101	0.95	4.34
			FEP Exited Y1	7	4.58%	-	-	-	-	-	-	-	-	-	-
			FEP Exited Y2	1	2.01%	-	-	-	-	-	-	-	-	-	-
			Missing	4	0.29%	-	-	-	-	-	-	-	-	-	-

Table 10. 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	207	77.82%	156.87	29.94	0	250	86.83	21.92	0	110	0.96	4.25
			NEP	22	8.27%	145.50	37.16	0	187	78.91	23.47	0	105	0.96	4.92
			LEP	5	1.88%	-	-	-	-	-	-	-	-	-	-
			PHLOTE	2	0.75%	-	-	-	-	-	-	-	-	-	-
			FELL	5	1.88%	-	-	-	-	-	-	-	-	-	-
			FEP Monitored Y1	11	4.14%	-	-	-	-	-	-	-	-	-	-
			FEP Monitored Y2	5	1.88%	-	-	-	-	-	-	-	-	-	-
			FEP Exited Y1	7	2.63%	-	-	-	-	-	-	-	-	-	-
			FEP Exited Y2	0	0.00%	-	-	-	-	-	-	-	-	-	-
			Missing	2	0.75%	-	-	-	-	-	-	-	-	-	-

Table 11. 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	8	Autism	51	14.61%	143.84	38.02	0	206	77.71	24.37	0	105	0.97	4.54
		Deaf-Blindness	0	0.00%	-	-	-	-	-	-	-	-	-	-
		Developmental Delay	0	0.00%	-	-	-	-	-	-	-	-	-	-
		Emotional Disturbance	0	0.00%	-	-	-	-	-	-	-	-	-	-
		Hearing Impairment	0	0.00%	-	-	-	-	-	-	-	-	-	-
		Intellectual Disability	107	30.66%	155.48	28.42	0	236	85.22	17.31	0	107	0.94	4.36
		Multiple Disabilities	170	48.71%	139.79	42.17	0	198	4.67	27.45	0	104	0.97	4.67
		Not Collected	0	0.00%	-	-	-	-	-	-	-	-	-	-
		Orthopedic Impairment	1	0.29%	-	-	-	-	-	-	-	-	-	-
		Other Health Impairment	13	3.72%	-	-	-	-	-	-	-	-	-	-
		Specific Learning Disability	4	1.15%	-	-	-	-	-	-	-	-	-	-
		Speech or Language Impairment	0	0.00%	-	-	-	-	-	-	-	-	-	-
		Traumatic Brain Injury	1	0.29%	-	-	-	-	-	-	-	-	-	-
		Visual Impairment	0	0.00%	-	-	-	-	-	-	-	-	-	-
		Missing	0	0.00%	-	-	-	-	-	-	-	-	-	-

Table 12. 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	36	13.53%	142.31	47.87	0	198	86.45	17.50	0	108	0.93	4.63
		Deaf-Blindness	0	0.00%	-	-	-	-	-	-	-	-	-	-
		Developmental Delay	0	0.00%	-	-	-	-	-	-	-	-	-	-
		Emotional Disturbance	0	0.00%	-	-	-	-	-	-	-	-	-	-
		Hearing Impairment	0	0.00%	-	-	-	-	-	-	-	-	-	-
		Intellectual Disability	81	30.45%	165.03	24.05	22	250	92.51	15.81	2	110	0.94	3.84
		Multiple Disabilities	141	53.01%	151.78	28.71	0	198	79.10	26.01	0	108	0.97	4.62
		Not Collected	0	0.00%	-	-	-	-	-	-	-	-	-	-
		Orthopedic Impairment	1	0.38%	-	-	-	-	-	-	-	-	-	-
		Other Health Impairment	3	1.13%	-	-	-	-	-	-	-	-	-	-
		Specific Learning Disability	1	0.38%	-	-	-	-	-	-	-	-	-	-
		Speech or Language Impairment	1	0.38%	-	-	-	-	-	-	-	-	-	-
		Traumatic Brain Injury	1	0.38%	-	-	-	-	-	-	-	-	-	-
		Visual Impairment	0	0.00%	-	-	-	-	-	-	-	-	-	-
		Missing	1	0.38%	-	-	-	-	-	-	-	-	-	-

Table 13. 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	8	Assistive Technology	No and Missing	328	93.98%	147.27	36.91	0	236	79.86	23.68	0	107	0.96	4.48
			Yes	21	6.02%	131.86	45.18	0	183	67.43	30.30	0	101	0.97	5.12
		Eye Gaze	No and Missing	343	98.28%	148.08	34.60	0	236	80.15	22.75	0	107	0.96	4.54
			Yes	6	1.72%	-	-	-	-	-	-	-	-	-	-
		Modified Picture Symbols	No and Missing	331	94.84%	147.53	36.82	0	236	80.18	23.55	0	107	0.96	4.48
			Yes	18	5.16%	124.56	45.06	0	206	59.50	29.20	0	105	0.97	5.25
		Objects	No and Missing	332	95.13%	147.97	36.30	0	236	80.38	23.35	0	107	0.96	4.48
			Yes	17	4.87%	114.65	47.89	0	167	54.35	28.91	0	95	0.97	5.22
		Sign Language	No and Missing	345	98.85%	146.77	37.14	0	236	79.43	23.86	0	107	0.96	4.53
			Yes	4	1.15%	-	-	-	-	-	-	-	-	-	-
		Translation into Native Language	No and Missing	347	99.43%	146.71	37.15	0	236	79.37	23.98	0	107	0.96	4.53
			Yes	2	0.57%	-	-	-	-	-	-	-	-	-	-
		Other	No and Missing	327	93.70%	146.96	36.94	0	236	79.60	23.77	0	107	0.96	4.49
			Yes	22	6.30%	137.27	45.78	0	198	71.86	30.34	0	104	0.97	5.01
	HS	Assistive Technology	No and Missing	246	92.48%	157.41	29.95	0	250	87.86	20.20	0	110	0.96	4.20
			Yes	20	7.52%	122.35	31.36	60	168	55.75	28.47	8	98	0.96	5.77
		Eye Gaze	No and Missing	261	98.12%	155.46	31.20	0	250	86.15	22.00	0	110	0.96	4.29
			Yes	5	1.88%	-	-	-	-	-	-	-	-	-	-
		Modified Picture Symbols	No and Missing	252	94.74%	157.21	27.70	0	250	87.31	19.73	0	110	0.95	4.29
			Yes	14	5.26%	-	-	-	-	-	-	-	-	-	-
		Objects	No and Missing	250	93.98%	157.04	29.33	0	250	87.37	20.20	0	110	0.95	4.32
			Yes	16	6.02%	119.44	134.50	22	166	55.31	34.04	2	97	0.98	4.66
		Sign Language	No and Missing	264	99.25%	154.84	31.49	0	250	85.50	22.57	0	110	0.96	4.33
			Yes	2	0.75%	-	-	-	-	-	-	-	-	-	-

		Translation into Native Language	No and Missing	265	99.62%	154.83	31.44	0	250	85.50	22.55	0	110	0.96	4.33
			Yes	1	0.38%	-	-	-	-	-	-	-	-	-	-
		Other	No and Missing	234	87.97%	155.43	30.91	0	250	85.89	22.27	0	110	0.96	4.34
			Yes	32	12.03%	149.97	34.92	0	193	82.19	24.41	0	106	0.97	4.33

Table 14. Grade 8 Science Classical Statistics

ITEM	TYPE	% 0	% 1	% 2	% 3	% 4	% 5	% 6	MEAN SCORE	ITEM-TOTAL CORR
1	SR	6.6	9.5	14.6	25.2	44.1			2.908	0.647
2	SR	6.3	8.9	13.5	13.8	57.6			3.074	0.791
3	SR	6.9	4.6	11.2	10.9	66.5			3.255	0.774
4	SR	5.7	9.2	18.4	21.3	45.4			2.914	0.715
5	SPT	8	1.4	1.4	24.4	21.8	19.2	23.8	4.032	0.721
6	SR	6.6	7.5	10.6	15.8	59.5			3.141	0.802
7	SR	6.6	11.2	21.8	24.1	36.2			2.721	0.697
8	SR	6	8.3	14.9	32.2	38.5			2.888	0.636
9	SR	7.4	11.7	18.1	17.8	45			2.811	0.728
10	SR	7.2	10.6	22.7	20.4	39.1			2.736	0.642
11	SR	6	10.9	22.6	22.6	37.8			2.754	0.687
12	SR	6.9	11.7	18.6	18.6	44.1			2.814	0.704
13	SR	6	7.2	8.6	10	68.2			3.272	0.838
14	SR	6.6	8.3	19.5	24.9	40.7			2.848	0.708
15	SR	6.9	5.2	8.6	18.3	61			3.215	0.765
16	SR	7.2	9.2	22.9	36.4	24.4			2.616	0.670
17	SR	6.6	9.5	12.9	15.5	55.6			3.04	0.802
18	SR	6.6	8.6	20.6	24.1	40.1			2.825	0.688
19	SR	6.9	5.4	10.3	11.7	65.6			3.238	0.808
20	SR	7.2	4	8.9	19.3	60.5			3.219	0.736
21	SR	8	15.8	22.6	26.4	27.2			2.49	0.672
22	SPT	6.3	1.7	2.6	19.2	37.2	22.1	10.9	3.891	0.695
23	SR	6	8.9	13.8	14	57.3			3.077	0.821
24	SR	6.3	7.2	10	14.3	62.2			3.189	0.850
25	SR	6.9	7.4	12	27.8	45.8			2.983	0.706
26	SR	6.9	6	11.7	8.6	66.8			3.223	0.814

Table 15. HS Science Classical Statistics

ITEM	TYPE	% 0	% 1	% 2	% 3	% 4	% 5	% 6	MEAN SCORE	ITEM-TOTAL CORR
1	SR	3	3.8	3.4	13.5	76.3			3.564	0.801
2	SR	4.2	6.8	9.8	19.2	60			3.242	0.682
3	SR	4.5	10.2	11.7	19.6	54			3.083	0.777
4	SPT	4.1	0.4	3.4	7.9	16.9	32	35.3	4.703	0.751
5	SR	3.8	12	15.8	21.4	47			2.959	0.720
6	SR	4.5	9.4	13.5	33.5	39.1			2.932	0.701
7	SR	3	9.4	15.4	30.5	41.7			2.985	0.658
8	SR	3.8	11.3	15.1	27.2	42.6			2.936	0.728
9	SR	4.1	7.5	19.2	23.3	45.9			2.992	0.666
10	SR	5.3	4.9	8.3	21.4	60.2			3.263	0.725
11	SPT	5.3	1.5	1.1	13.5	17.3	25.2	36.1	4.56	0.735
12	SR	4.9	10.9	11.7	20.3	52.3			3.041	0.743
13	SR	4.9	6	11.7	27.4	50			3.117	0.729
14	SR	4.1	8.3	7.5	16.2	63.9			3.274	0.822
15	SR	4.5	8.3	9.4	16.9	60.9			3.214	0.812
16	SR	3.4	10.2	25.9	25.6	35			2.786	0.555
17	SR	4.9	3.4	14.7	21.1	56			3.199	0.788
18	SR	4.1	7.9	9	27.1	51.9			3.147	0.720
19	SR	5.3	5.6	16.2	24.4	48.5			3.053	0.664
20	SR	5.6	7.1	10.2	27.4	49.6			3.083	0.762
21	SPT	3.4	1.1	1.1	7.1	8.6	16.9	61.7	5.139	0.773
22	SR	4.9	3	12	18.8	61.3			3.286	0.706
23	SR	5.3	10.2	20.3	30.1	34.2			2.778	0.671
24	SR	4.9	6.8	6.8	9.4	72.2			3.372	0.833
25	SR	5.3	11.7	24.5	20.8	37.7			2.74	0.657
26	SR	5.6	6.8	17.7	17.7	52.3			3.041	0.663

Table 16. Grade 8 Science Item Parameter Estimates

ITEM	TYPE	B	D1	D2	D3	D4	D5	D6	D7	INFIT	OUTFIT
1	SR	0.2254	0	-1.3491	0.2821	0.3325	0.7346			1.07	1.18
2	SR	0.2731	0	-1.2782	0.4724	0.9002	-0.0944			0.79	0.70
3	SR	-0.0583	0	-0.741	0.1375	0.6842	-0.0807			1.08	0.91
4	SR	0.384	0	-1.3504	0.2585	0.3019	0.79			0.95	0.90
5	SPT	0.1292	0	-0.5492	-0.6137	-2.5213	1.1065	1.616	0.9617	1.01	1.10
6	SR	-0.0936	0	-2.0243	1.1639	0.3161	0.5443			0.79	0.67
7	SR	0.2894	0	-1.9662	0.2239	0.6299	1.1123			0.98	1.01
8	SR	0.0802	0	-1.7357	0.1018	0.1521	1.4818			1.25	1.16
9	SR	0.425	0	-1.6801	0.1421	0.9772	0.5609			0.96	0.94
10	SR	0.3076	0	-1.4943	-0.1204	1.0207	0.594			1.24	1.24
11	SR	0.3901	0	-1.8583	0.5554	0.3618	0.9412			0.85	0.81
12	SR	0.3179	0	-1.4048	-0.197	1.136	0.4658			1.07	1.05
13	SR	-0.0116	0	-1.3537	0.7331	1.0361	-0.4156			0.84	0.71
14	SR	0.2072	0	-1.4528	0.2151	0.504	0.7337			0.99	0.94
15	SR	-0.0317	0	-1.2016	0.0507	0.9201	0.2308			1.11	0.99
16	SR	0.5111	0	-1.7263	-0.5202	0.4961	1.7503			1.06	1.03
17	SR	0.1989	0	-1.2971	0.4798	0.6491	0.1682			0.81	0.76
18	SR	0.3207	0	-1.5338	-0.2799	0.7203	1.0935			1.24	1.21
19	SR	-0.0471	0	-1.1555	0.621	0.5857	-0.0512			0.88	0.73
20	SR	-0.1624	0	-0.921	-0.2127	0.7289	0.4048			1.21	1.19
21	SR	0.4242	0	-1.6286	-0.0777	0.7529	0.9535			0.96	0.96
22	SPT	0.5633	0	0.2253	-1.2384	-2.8452	0.2121	1.2474	2.3988	1.23	1.24
23	SR	0.0587	0	-1.4117	0.4486	0.8646	0.0985			0.84	0.69
24	SR	0.1766	0	-0.8329	0.516	0.5382	-0.2212			0.75	0.80
25	SR	0.2862	0	-1.4079	0.1094	0.3236	0.9749			1.12	1.03
26	SR	-0.0515	0	-1.5727	0.761	1.57	-0.7584			0.74	0.61

Table 17. 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			1.15	0.93
2	SR	-0.5184	0	-2.1061	0.2195	0.9434	0.9433			1.08	0.99
3	SR	-0.0847	0	-1.496	-0.0074	0.6511	0.8522			0.99	0.93
4	SPT	-0.1258	0	-1.244	-0.1547	-1.5832	0.2891	1.1881	1.5048	1.07	1.16
5	SR	0.1458	0	-1.9366	0.2219	1.1648	0.5499			0.92	0.90
6	SR	0.2797	0	-1.8174	0.3853	0.147	1.2851			0.95	1.01
7	SR	0.2932	0	-1.4265	-0.2011	0.719	0.9086			1.04	1.04
8	SR	0.2204	0	-1.8758	0.179	0.8647	0.8321			0.94	1.02
9	SR	-0.1146	0	-2.2841	0.5031	0.6997	1.0813			0.98	1.01
10	SR	-0.1691	0	-2.1532	0.5199	1.0258	0.6076			1.05	1.08
11	SPT	0.0201	0	-0.3036	-0.8041	-2.1901	0.8515	0.9046	1.5417	1.02	1.09
12	SR	0.0315	0	-1.8582	0.2223	0.9184	0.7175			0.96	0.99
13	SR	0.0491	0	-1.5344	-0.1217	0.6584	0.9977			0.98	1.03
14	SR	-0.2761	0	-1.1544	0.0227	0.9848	0.1469			1.02	0.77
15	SR	-0.0606	0	-1.6195	0.6173	0.8159	0.1864			0.82	0.68
16	SR	0.3846	0	-1.6778	-0.72	1.2687	1.1292			1.25	1.20
17	SR	-0.1199	0	-1.3246	-0.2161	0.5995	0.9412			1.22	1.05
18	SR	0.0565	0	-1.0219	-0.1391	0.6189	0.5421			0.88	0.74
19	SR	-0.0997	0	-1.6232	-0.2646	1.1381	0.7497			1.31	1.25
20	SR	0.1051	0	-1.6232	0.2092	0.6236	0.7903			0.92	0.93
21	SPT	-0.3763	0	0.1187	-0.648	-1.9409	0.6837	1.2735	0.5132	0.99	0.94
22	SR	-0.2383	0	-1.5895	0.586	0.4658	0.5377			1.09	0.92
23	SR	0.1478	0	-2.3932	0.046	0.9761	1.3711			1.09	1.12
24	SR	-0.0973	0	-1.1545	0.4857	0.9097	-0.2409			0.74	0.52
25	SR	0.3537	0	-1.7139	0.1344	0.7676	0.8119			1.01	1.05
26	SR	0.005	0	-1.9186	-0.0368	1.112	0.8434			1.35	1.35

Table 18. 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	%
	8	59	93	102	43	12	202	58	89	26	15	4	104	30
	HS	72	95	106	41	15	113	43	100	38	12	5	112	42

Table 19. Scale Score Ranges for Each Performance Level

	Emerging Level	Approaching Target Level	At Target Level	Advanced Level
Grade 8 Science	0–127	128–163	164–189	190–250
HS Science	0–139	140–163	164–192	193–250

Table 20. Grade 8 Science Scale Score Frequency Distributions

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	10	2.87	10	2.87
1	1	0.29	11	3.15
12	1	0.29	12	3.44
24	1	0.29	13	3.72
32	2	0.57	15	4.3
39	3	0.86	18	5.16
50	1	0.29	19	5.44
75	1	0.29	20	5.73
82	1	0.29	21	6.02
87	1	0.29	22	6.3
89	1	0.29	23	6.59
91	1	0.29	24	6.88
92	1	0.29	25	7.16
93	1	0.29	26	7.45
95	1	0.29	27	7.74
104	1	0.29	28	8.02
110	1	0.29	29	8.31
111	1	0.29	30	8.6
113	1	0.29	31	8.88
114	2	0.57	33	9.46
117	1	0.29	34	9.74
119	1	0.29	35	10.03
120	2	0.57	37	10.6
121	1	0.29	38	10.89
123	2	0.57	40	11.46
125	1	0.29	41	11.75
127	2	0.57	43	12.32
128	2	0.57	45	12.89
129	1	0.29	46	13.18

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
132	2	0.57	48	13.75
133	4	1.15	52	14.9
134	3	0.86	55	15.76
135	2	0.57	57	16.33
136	3	0.86	60	17.19
137	4	1.15	64	18.34
138	5	1.43	69	19.77
139	5	1.43	74	21.2
140	14	4.01	88	25.21
141	5	1.43	93	26.65
142	4	1.15	97	27.79
143	7	2.01	104	29.8
144	5	1.43	109	31.23
145	8	2.29	117	33.52
146	7	2.01	124	35.53
147	4	1.15	128	36.68
148	5	1.43	133	38.11
149	9	2.58	142	40.69
151	8	2.29	150	42.98
152	12	3.44	162	46.42
153	11	3.15	173	49.57
154	11	3.15	184	52.72
156	13	3.72	197	56.45
157	13	3.72	210	60.17
158	12	3.44	222	63.61
160	7	2.01	229	65.62
162	16	4.58	245	70.2

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
164	20	5.73	265	75.93
165	11	3.15	276	79.08
167	9	2.58	285	81.66
169	9	2.58	294	84.24
171	7	2.01	301	86.25
174	9	2.58	310	88.83
177	13	3.72	323	92.55
180	2	0.57	325	93.12
183	9	2.58	334	95.7
190	5	1.43	339	97.13
192	6	1.72	345	98.85
198	2	0.57	347	99.43
206	1	0.29	348	99.71
236	1	0.29	349	100

Table 21. HS Science Scale Score Frequency Distributions

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	4	1.5	4	1.5
22	1	0.38	5	1.88
60	1	0.38	6	2.26
70	1	0.38	7	2.63
72	1	0.38	8	3.01
75	1	0.38	9	3.38
77	3	1.13	12	4.51
88	1	0.38	13	4.89
94	1	0.38	14	5.26
107	1	0.38	15	5.64
109	1	0.38	16	6.02
110	1	0.38	17	6.39
118	2	0.75	19	7.14
123	1	0.38	20	7.52
124	1	0.38	21	7.89
127	2	0.75	23	8.65
128	2	0.75	25	9.4
132	1	0.38	26	9.77
133	1	0.38	27	10.15
134	2	0.75	29	10.9
135	2	0.75	31	11.65
136	1	0.38	32	12.03
137	4	1.5	36	13.53
138	2	0.75	38	14.29
139	3	1.13	41	15.41
140	1	0.38	42	15.79
141	6	2.26	48	18.05
142	3	1.13	51	19.17
143	1	0.38	52	19.55

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
144	5	1.88	57	21.43
145	3	1.13	60	22.56
146	7	2.63	67	25.19
147	3	1.13	70	26.32
148	1	0.38	71	26.69
149	3	1.13	74	27.82
150	4	1.5	78	29.32
151	8	3.01	86	32.33
152	7	2.63	93	34.96
153	5	1.88	98	36.84
154	10	3.76	108	40.6
155	7	2.63	115	43.23
157	8	3.01	123	46.24
158	6	2.26	129	48.5
159	12	4.51	141	53.01
160	5	1.88	146	54.89
162	8	3.01	154	57.89
164	7	2.63	161	60.53
165	8	3.01	169	63.53
166	13	4.89	182	68.42
168	15	5.64	197	74.06
170	8	3.01	205	77.07
172	9	3.38	214	80.45
174	4	1.5	218	81.95
177	5	1.88	223	83.83
180	12	4.51	235	88.35
183	10	3.76	245	92.11

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
187	9	3.38	254	95.49
193	5	1.88	259	97.37
198	5	1.88	264	99.25
208	1	0.38	265	99.62
250	1	0.38	266	100

Table 22. Grade 8 Science Scale Scores and Conditional Standard Error of Measurement (CSEM)

Raw Score	Scale Score	CSEM
0	0	50
1	1	28
2	12	20
3	24	17
4	32	14
5	39	13
6	45	12
7	50	11
8	54	11
9	58	10
10	61	9
11	64	9
12	67	9
13	70	8
14	72	8
15	75	8
16	77	8
17	79	7
18	81	7
19	82	7
20	84	7
21	86	7
22	87	7
23	89	6
24	91	6
25	92	6

26	93	6
27	95	6
28	96	6
29	98	6
30	99	6
31	100	6
32	101	6
33	103	6
34	104	6
35	105	6
36	106	6
37	107	6
38	108	5
39	109	5
40	110	5
41	111	5
42	112	5
43	113	5
44	114	5
45	115	5
46	116	5
47	117	5
48	118	5
49	119	5
50	120	5
51	121	5
52	122	5
53	123	5
54	123	5
55	124	5

56	125	5
57	126	5
58	127	5
59	128	5
60	128	5
61	129	5
62	130	5
63	131	5
64	132	5
65	133	5
66	133	5
67	134	5
68	135	5
69	136	5
70	137	5
71	138	5
72	139	5
73	140	5
74	140	5
75	141	5
76	142	5
77	143	5
78	144	5
79	145	5
80	146	5
81	147	5
82	148	5
83	149	6
84	151	6
85	152	6

86	153	6
87	154	6
88	156	6
89	157	6
90	158	6
91	160	7
92	162	7
93	164	7
94	165	7
95	167	7
96	169	8
97	171	8
98	174	8
99	177	9
100	180	10
101	183	10
102	190	11
103	192	12
104	198	14
105	206	16
106	217	19
107	236	28
108	250	50

Table 23. HS Science Scale Scores and Conditional Standard Error of Measurement (CSEM)

Raw Score	Scale Score	CSEM
0	0	44
1	1	24
2	22	17
3	33	14
4	41	13
5	47	12
6	52	11
7	56	10
8	60	9
9	64	9
10	67	8
11	70	8
12	72	8
13	75	8
14	77	7
15	79	7
16	81	7
17	83	7
18	85	6
19	87	6
20	88	6
21	90	6
22	91	6
23	93	6
24	94	6
25	95	6

26	97	6
27	98	5
28	99	5
29	100	5
30	101	5
31	103	5
32	104	5
33	105	5
34	106	5
35	107	5
36	108	5
37	109	5
38	110	5
39	111	5
40	112	5
41	113	5
42	114	5
43	115	5
44	116	5
45	117	5
46	118	5
47	119	5
48	120	5
49	121	5
50	122	5
51	123	5
52	124	5
53	124	4
54	125	4
55	126	4

56	127	4
57	128	4
58	129	4
59	129	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	141	4
74	141	4
75	142	4
76	143	4
77	144	5
78	145	5
79	146	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	166	6
98	168	7
99	170	7
100	172	7
101	174	8
102	177	8
103	180	9
104	183	9
105	187	10
106	193	11
107	198	13
108	208	16
109	223	24
110	250	43

Table 24. 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)
	8	0.63	0.45	0.33	0.40	0.69	0.30	0.01
	HS	0.60	0.36	0.38	0.40	0.66	0.09	0.25

Table 25. Test Validity Questions Summary

Question	Subject	Grade	Very Familiar	Somewhat Familiar	Familiar	Somewhat Unfamiliar	Unfamiliar	Missing			
How familiar are you with this student?	SC	8	87.39%	6.3%	4.01%	0.57%	0.29%	1.43%			
		HS	84.21%	9.4%	3.38%	0.75%	0.00%	2.26%			
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?	SC	8	12.32%	12.89%	15.76%	17.19%	28.08%	11.75%	0.57%	1.43%	
		HS	11.28%	13.53%	15.79%	27.82%	17.29%	10.9%	0.75%	2.63%	
Question		Grade	25%	50%	75%	100%	None	Missing			
Approximately how much instructional time for this content area is in the general education classroom?	SC	8	13.75%	9.17%	13.75%	26.93%	34.96%	1.43%			
		HS	16.17%	5.64%	5.64%	16.54%	53.38%	2.63%			
Question		Grade	Oral Language	Sign Language	Reading	Picture Communication	Tactile	Other	Do Not Know	Missing	
This student's primary receptive communication is:	SC	8	93.70%	0.57%	0.00%	2.29%	0.29%	1.43%	0.00%	1.72%	
		HS	93.23%	0.38%	1.13%	1.5%	0.00%	0.38%	1.13%	2.26%	
Question		Grade	Oral Language	Sign Language	Writing	Picture Communication	Augmentative Communication Device	Tactile	Other	Do Not Know	Missing
This student's primary expressive communication is:	SC	8	83.67%	1.43%	0.57%	2.58%	6.3%	0.00%	4.01%	0.00%	1.43%
		HS	81.95%	0.38%	0.38%	3.01%	6.77%	0.00%	4.14%	1.13%	2.26%
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.	SC	8	43.84%	37.25%	10.89%	3.72%	2.29%	0.57%	1.43%		
		HS	46.24%	34.59%	9.02%	4.89%	1.5%	1.5%	2.26%		
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
	SC	8	1.15%	26.36%	58.17%	9.17%	1.72%	0.57%	0.00%	0.57%	2.29%

How much time did this student take on the assessment?		HS	0.38%	22.56%	64.29%	8.27%	1.50%	0.00%	0.38%	0.38%	2.26%

Table 26. Correlation Between Student Scores and Familiarity with the Student

Subject	Assessment	N	Correlation
Science	Grade 8	349	0.02
	HS	266	-0.04

Table 27. Correlation Between Student Scores and Hours Per Week in Instruction on the Content Area

Subject	Assessment	N	Correlation
Science	Grade 8	349	0.20
	HS	266	0.15

Table 28. Correlation Between Student Scores and How Much Instructional Time in the Content Area Is in the General Education Classroom

Subject	Assessment	N	Correlation
Science	Grade 8	349	0.32
	HS	266	0.17

APPENDICES

APPENDIX A: COALT: SCIENCE 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:
Gina Herrera Herrera_g@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:
Gina Herrera Herrera_g@cde.state.co.us Exceptional Student Services Unit Colorado Dept. of Education
Rev. 9/15

APPENDIX B: COALT: SCIENCE TEST BLUEPRINTS

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 SAMPLE SCORE REPORT

Confidential Student Performance Report

Colorado Alternate Assessment

Student: **STUDENT NAME**

SASID: **9999999999** Birthdate: **MM/DD/YYYY**
School: **SCHOOL NAME (9999)**
District: **DISTRICT NAME (9999)**

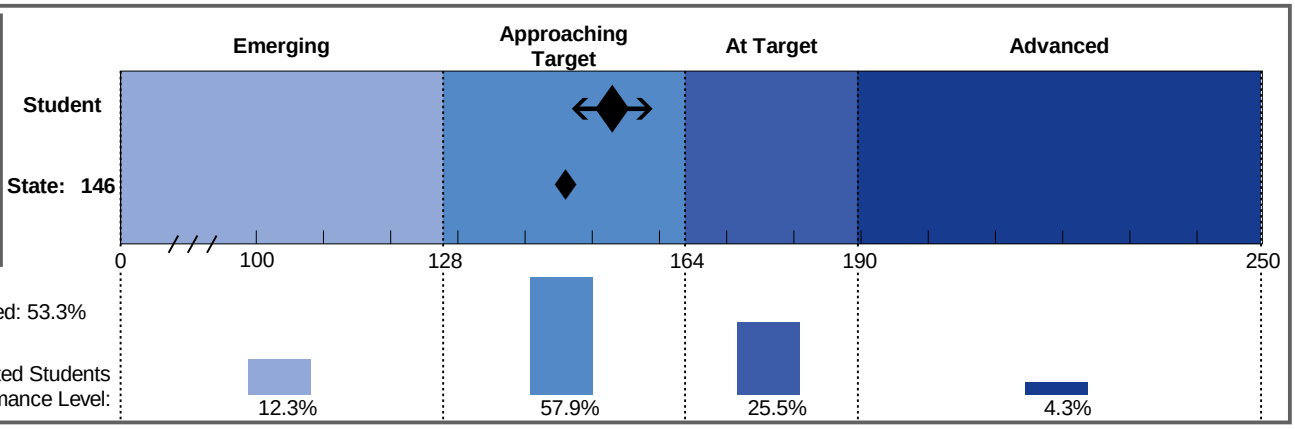
Spring 2021

Science — Grade 8

This Colorado Alternate (CoAlt) report provides information about your student's mastery of skills and concepts in the Extended Evidence Outcomes of the academic standards which are the basis for instruction in Colorado schools. Your student's performance on this test is represented by a scale score and a performance level.

- Scale scores are represented by diamonds on the graph. The arrows around your student's diamond show the range of scores your student would likely receive if the assessment were taken multiple times. Scale scores can be compared across years.
- State information allows you to 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.
- The performance levels are separated by dotted lines.
- You are encouraged to discuss this report with your student's teacher.

Your Student's Score
153
Approaching Target



Content Standard Performance

Reporting Category Description	Points Earned	Points Possible	Percent of Points Earned*
Physical Science			
Students know and understand common properties, forms, and changes in matter and energy.	24	30	80% / 72%
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.	23	28	82% / 74%
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.	39	50	78% / 73%

*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

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

Science Performance Level Descriptions

Students demonstrate science concepts and skills aligned to the Grade Level Expectations and Extended Evidence Outcomes contained in the Colorado Academic Standards. The performance level descriptors are organized in a manner that assumes students demonstrating higher levels of command have mastered the concepts and skills within the lower levels. For example, a student who is At Target has also mastered the concepts and skills included in the Approaching Target performance level.

With appropriate support, Advanced students can typically:

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

With appropriate support, At Target students can typically:

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

With appropriate support, Approaching Target students can typically:

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

With appropriate support, Emerging students can typically:

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

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

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