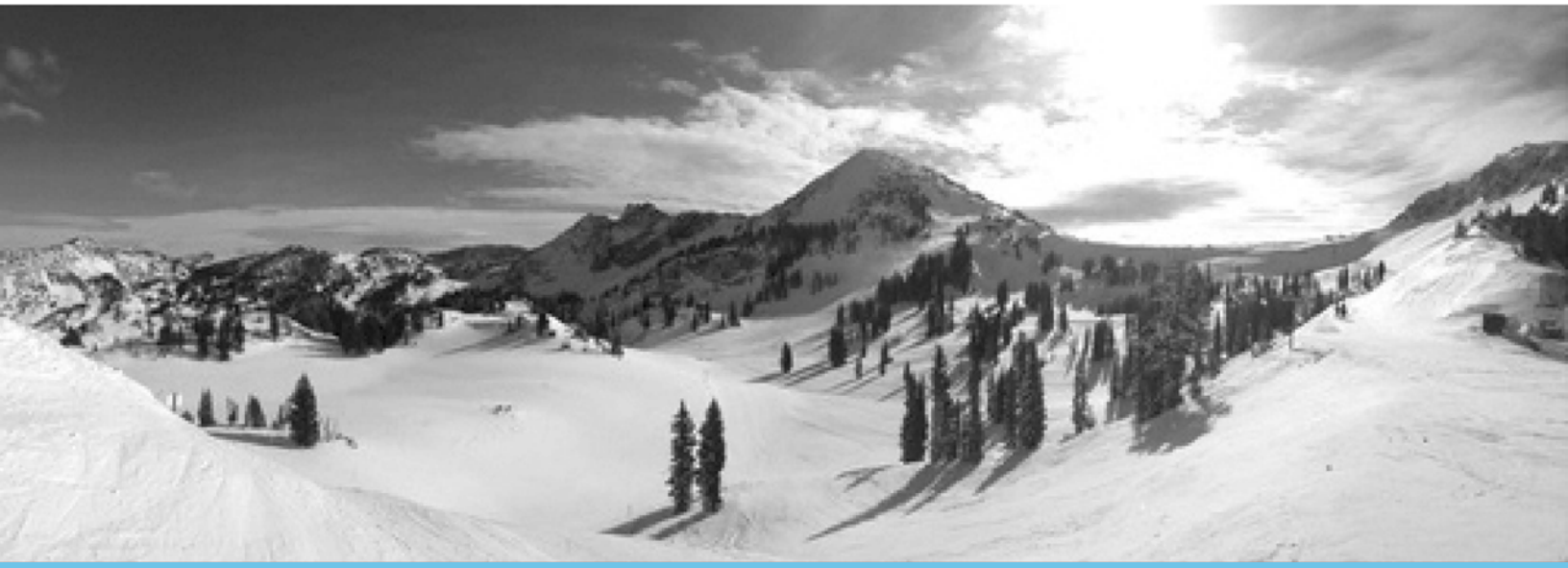
Colorado Alternate Assessment Program



Technical Report

Science

2022

Colorado Alternate Assessment (CoAlt) Science Technical Report

2021–2022

Table of Contents

List of Tables

Chapter 1: Introduction

The purpose of this technical report is to inform users and other interested parties about the development, content, administration, and technical characteristics of the Spring 2022 Colorado Alternate Assessment (CoAlt) Science assessment in Grades 5, 8, and 11 for students with the most significant cognitive disabilities. The report includes an overview and summary of the components of the program, including information regarding the planning and administration of the assessments and details regarding item development, test construction, administration procedures, scoring, reporting, reliability, and validity, as well as a statistical summary of the Spring 2022 items.

The CoAlt assessments are administered in English language arts (ELA), mathematics, and science. The CoAlt ELA and Mathematics assessments are administered by the Dynamic Learning Maps (DLM) consortium and are documented in a separate technical report. Social studies was not assessed in Spring 2022.

1.1. Testing Requirements

All public schools in Colorado are required by state law to administer a standards-based summative assessment each year in specified content areas and grade levels. Every student, regardless of ability or language background, must be provided with the opportunity to demonstrate their content knowledge through the state assessments. The Colorado Measures of Academic Success (CMAS) assessments in mathematics, ELA, and science are Colorado's end-of-year standards-based assessments designed to measure students' achievement of the grade-level Colorado Academic Standards (CAS).

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 Every Student Succeeds Act of 2015 (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, Colorado developed the CoAlt science assessment. Colorado also administers the DLM consortium assessments for the CoAlt ELA and Mathematics assessments, the technical report for the DLM assessments can be found here: https://2022-ye-techmanual.dynamiclearningmaps.org/.

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

1.2. Intended Population

The CoAlt assessments are designed for students with the most significant cognitive disabilities who have significant limitations in cognitive functioning and deficits in adaptive behavior. These students may also 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 CMAS assessment, even with accommodations, may take the CoAlt assessment. These students must be identified as having a significant cognitive disability, although Intellectual Disability does not have to be the student's primary disability label for IDEA eligibility.

CoAlt participation is determined by a student's Individualized Education Program (IEP) team that decides whether the student meets the criteria in the alternate academic achievement standards and the Alternate Assessment Participation Guidelines Worksheet provided in Appendix A.[1] The IEP team can decide that the CoAlt assessment is most appropriate 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 significant cognitive disability.
- The student has a significant cognitive disability.
- The student is receiving daily instruction based on the alternate academic achievement standards.

1.3. CoAlt Background

The CoAlt assessments follow the direction of the Office of Standards and Instruction (SIS) and Exceptional Student Services Unit (ESSU) at the Colorado Department of Education (CDE). 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. The CAS for science were originally adopted in December 2009. On August 3, 2011, the State Board of Education adopted the Extended Evidence Outcomes (EEOs) for students with the most significant cognitive disabilities who qualify for an alternate assessment. In partnership with Colorado educators and Pearson, CDE developed the CoAlt Science assessments to evaluate student mastery of the CAS in science for students with the most significant cognitive disabilities. For eligible students, these end-of-year assessments provide an indicator of student progress toward the EEOs of the CAS, known as the alternate academic achievement standards. The first operational administration of the CoAlt Science assessments occurred in Spring 2014 for Grades 5 and 8 and in Fall 2014 for high school.

The Spring 2020 CoAlt administration was cancelled due to the COVID-19 pandemic. In 2021, Colorado received a partial waiver of the federal assessment requirements from the U.S. Department of Education (USED) due to COVID-19 conditions in Colorado. With the exception of students with a parent/guardian excusal, students in Grade 8 and high school took the CoAlt Science assessment. The Grade 5 science assessment was not administered. The Grades 4 and 7 social studies assessments were also not administered.

In 2008, Colorado passed Senate Bill 212 (also known as CAP4K) that required the State Board of Education to adopt content standards that prepare students for the 21st century workforce and for active citizenship upon receiving a high school diploma. It also required a revision to the CAS by July 1, 2018, and every six years thereafter. As such, the 2009/2010 CAS were reviewed and revised, resulting in the 2020 CAS. While minimal changes were made to the mathematics and ELA CAS, the science CAS underwent a substantial update to keep up with the shift to the Next Generation Science Standards (NGSS; NGSS Lead States, 2013)[2]. After the CAS were adopted, a joint effort between SIS and ESSU commenced to develop the EEOs to which the

[1] The participation guideline worksheet is also available online at http://www.cde.state.co.us/cdesped/accommodationsmanual_participationguidelinesworksheet.

[2] Next Generation Science Standards is a registered trademark of WestEd. Neither WestEd nor the lead states and partners that developed the NGSS were involved in the production of this product and do not endorse it.

CoAlt is aligned. An educator committee composed of both special educators and content educators convened to adapt the Evidence Outcomes (EOs) from the 2020 CAS to EEOs. Item development for the new CoAlt Science assessment then began in Spring 2021.

Colorado students saw items aligned to the 2020 CAS for the first time in spring 2022. The new science assessments based on the new three-dimensional science standards were administered to all tested students which made it possible to test enough new content to allow for a robust item bank and to obtain a sufficient number of students to conduct field test analyses.

The goals of the Colorado Assessment System, including the CoAlt assessments, are to measure and support student progress toward the content standards; provide students, parents/guardians, and other stakeholders with information regarding student achievement; and gauge the quality and efficiency of educational programs in public schools. For CoAlt in particular, 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. CoAlt also promotes improved instruction toward grade-level expectations, growth over time toward independent performance, and high expectations toward achievement in the content areas. CoAlt results may be used in many ways, including to:

- inform instruction in the classroom;
- inform district and school leaders about potential programming and instruction priorities;
- provide the community with information on how well the state's education system is meeting the goals of helping every student attain academic proficiency in accordance with Colorado's alternate standards;
- provide aggregated data for the state's accountability system; and
- allow students to demonstrate their mastery of skills and concepts in the EEOs.

1.4. Assessment Development Partners

Activities specific to the CoAlt Science assessments were conducted collaboratively by CDE, the Colorado educator community, and Pearson, the assessment contractor. Input and advice were also provided by the Colorado Technical Advisory Committee (TAC).

1.4.1. Colorado Department of Education

As the administrative arm of the State Board 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, with CDE serving as the ultimate approver of services and products provided.

1.4.2. Colorado Educator Community

Educator participation in the CoAlt development process is critical to ensuring that the assessments are aligned to the EEOs of the CAS, are appropriate for Colorado students with the most significant cognitive disabilities at the assessed grade level, and are free from potential bias and sensitivity issues. Throughout the test development process, educators provide input through

participation in content and bias review, data 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. Table 1.1 presents a schedule of major events from the 2021–2022 testing cycle that includes meetings with educator participation.

Table 1.1. Schedule of Major Events

Event	Date(s)
DAC Administration Training	December 2021
Spring 2022 Administration Window	April 11–29, 2022
CoAlt Educator Feedback Session	June 7–9, 2022
Reports Released	August 22, 2022
Data Review	August 23–24, 2022
Standard Setting	October 25–26, 2022

1.4.3. Pearson

As the primary contractor responsible for the end-to-end assessment cycle services and products, Pearson works closely with CDE throughout the CMAS (all content areas) and CoAlt (science) assessment development and administration processes. This includes item and test development, forms creation, enrollment, packaging and distribution, test delivery, scoring, customer service, standard setting, score reporting, and psychometric services.

1.4.4. Colorado Technical Advisory Committee

The Colorado TAC is 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 include blueprint design, scaling and equating, scoring, reporting, and standard setting. The TAC included the following members during the 2022 assessment cycle:

- 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, Psychometric Consultant
- Dr. Suzanne Lane, Professor, University of Pittsburgh
- Dr. Martha Thurlow, Director, National Center on Educational Outcomes

Chapter 2: Test Design

The CoAlt Science assessment was designed to provide this unique population of students with an opportunity to demonstrate their knowledge of the EEOs. The assessments include paper-based test books used by the Test Administrator to administer test items to the students. Each assessment is administered one-on-one and can be split over as many sessions/days as appropriate for the student. The test books are designed to sit on the table, allowing the Test Administrator to read the item and answer options to the student while allowing the student to view the answer options. The test books include scripted text for the Test Administrator to read that include both the test items and answer options to the student. There is flexibility for presentation and response based on the student's mode of communication, but the script and order in which the answer choices are presented to the student must remain the same. During the administration, the Test Administrator records the student's responses on an answer document included with the task manipulatives set provided for each test. The answer document is then returned for scoring.

2.1. Alternate Academic Achievement Standards

The EEOs are alternate academic standards aligned to the grade-level 2020 CAS in science but reduced in depth, breadth, and complexity. They can be found online at http://www.cde.state.co.us/CoExtendedEO/StateStandards. The standards are considered three-dimensional in that they incorporate Disciplinary Core Ideas (DCIs), Science and Engineering Practices (SEPs), and Crosscutting Concepts (CCCs). The DCIs encompass the content that occurs at each grade and provides the background knowledge for students to develop sense-making around phenomena in the three standards of Physical Science, Life Science, and Earth and Space Science:

- Physical Science: Students know and understand common properties, forms, and changes in matter and energy.
    - PS1: Matter and its interactions
    - PS2: Motion and stability: Forces and interactions
    - PS3: Energy
    - PS4: Waves and their applications in technologies for information transfer

- 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.
    - LS1: From molecules to organisms: Structures and processes
    - LS2: Ecosystems: Interactions, energy, and dynamics
    - LS3: Heredity: Inheritance and variation of traits
    - LS4: Biological evolution: Unity and diversity

- Earth and Space 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.
    - ESS1: Earth's place in the universe
    - ESS2: Earth's systems
    - ESS3: Earth and human activity

The SEPs describe how scientists investigate and build models and theories of the natural world or how engineers design and build systems. They reflect science and engineering as they are practiced and experienced. There are eight practices:

1. Asking questions (for science) and defining problems (for engineering)
2. Developing and using models
3. Planning and carrying out investigations
4. Analyzing and interpreting data
5. Using mathematics and computational thinking
6. Constructing explanations (for science) and designing solutions (for engineering)
7. Engaging in argument from evidence
8. Obtaining, evaluating, and communicating information

CCCs cross boundaries between science disciplines and provide an organizational framework to connect knowledge from various disciplines into a coherent and scientifically based view of the world. They build bridges between science and other disciplines and connect the DCIs and SEPs throughout the fields of science and engineering. There are seven CCCs:

1. Patterns
2. Cause and Effect
3. Scale, Proportion, and Quantity
4. Systems and System Models
5. Energy and Matter
6. Structure and Function
7. Stability and Change

The most substantial revision from the 2009 EEOs is the addition of a one-to one correspondence to each EO, thereby increasing the rigor for students with the most significant cognitive disabilities. Prior iterations of the EEOs had only 1–4 outcomes for each standard. SEPs and CCCs are incorporated into the EEOs, though not all EEOs are three-dimensional. SEPs and CCCs are also tested within the items.

The CoAlt Science assessment is given in Grades 5, 8, and 11. Consistent with the standards, the Grade 5 assessment assesses the grade-level standards. Because the science standards are articulated by grade band at the middle school and high school levels rather than grade levels, the Grade 8 CoAlt Science assessment assesses all middle school science standards, and the Grade 11 assessment assesses all high school science standards.

2.2. Cognitive Complexity

All CoAlt Science items are assigned a Depth of Knowledge (DOK) level that indicates the cognitive complexity of the item. DOK refers to the level of rigor or sophistication of the task in an item designed to reflect the complexity of the CAS. To ensure that the assessments include a deep pool of items that span a full range of cognitive levels and skills, each item was evaluated and tagged with one of the following DOK levels: Level 1: Recall, Level 2: Skill & Concepts, and Level 3: Strategic Thinking. DOK Level 4: Extended Thinking items are not included because the tests do not contain any extended-response items.

2.3. Item Types

CoAlt Science assessments include selected response (SR) and supported performance task (SPT) item types. SR items are scaffolded items presented in a three-item set with an initial stimulus containing a scientific phenomenon. Each item has 3–4 answer options and corresponds to the same stimulus but the items are unrelated to each other. The stimulus provides background information, and the art is repeated on the student-facing page with each item. The first item in the set has three picture answer options and is one-dimensional, testing only the DCI from the EEO. These items do not require sensemaking (i.e., the items are DOK Level 1, meaning they are just recall and do not require the student to figure something out). The second item in the set has three picture answer options and is two-dimensional, requiring sensemaking. It tests the DCI and either the SEP or CCC. The third item in the set has four answer options that are primarily picture-based. It is three-dimensional and requires sensemaking. Each SR item is scored independently and is worth 1 point.

SPT items consist of three related items, or tasks. Test Administrators are provided with three prompts, and the students respond to each prompt using a set of option cards. Students manipulate the option cards by placing them in designated boxes within a chart or diagram. Students may manipulate the option cards independently or indicate the desired placement through their preferred mode of expressive communication, such as verbal directions or eye gaze. Test Administrators score the student's performance on each prompt using a 1-point scoring rubric that is built into the item. The points for the three prompts are added together to provide one score for the SPT item. This item type reveals a different level of understanding of specific concepts and skills than those demonstrated through SR items alone. These items are all three-dimensional, are phenomenon based, and require sensemaking.

SPT tasks are classified as "Give a card" or "Find a card." For "Give a card" tasks, the Test Administrator gives the student a card to place in a table or other graphic organizer. The tasks have three answer cards, all of which are used. For "Find a card" tasks, the Test Administrator asks the student to search for a card of four provided cards in response to an item and place that card in a table or other graphic organizer. These tasks have four answer cards, one of which is not used.

2.4. Test Frameworks and Blueprints

The CoAlt 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, policymakers, and the public by clearly defining the elements of the EEOs that are suitable for state testing. The CoAlt frameworks can be found online at http://www.cde.state.co.us/assessment/newassess-coaltsss.[3] The test blueprints take the frameworks a step further by specifying the number of test items by content standard, grade-level expectation (GLE), EEO, and item type. The specificity of the test blueprints ensures that the assessments cover the breadth of the content indicated by the CAS within the associated grade or grade band. CDE and Pearson collaboratively developed the CoAlt Science test blueprints based on the CMAS blueprints.

[3] The frameworks located online will be applicable starting in 2023. The 2022 tests were built based on a different but similar set of frameworks.

While the complete blueprints are used internally, Table 2.1, Table 2.2, and Table 2.3 present the high-level CoAlt Science blueprints that summarize the number of items and percentage of score points on each test.

Table 2.1. 2022 CoAlt Science Test Blueprint—Grade 5

Subclaim	#Item Sets	Total #Items	#1-Point SR Items	#3-Point SPT Items	Total #Points	% of Total Points
Physical Science	2	7	6	1	9	30%
Physical Science/Life Science	3	9	9	0	9	30%
Earth and Space Science	3	10	9	1	12	40%
Total	8	26	24	2	30	100%

Note. SR = selected-response, SPT = supported performance task. One Physical Science EEO will always be clustered with the Life Science EEOs.

Table 2.2. 2022 CoAlt Science Test Blueprint—Grade 8

Subclaim	#Item Sets	Total #Items	#1-Point SR Items	#3-Point SPT Items	Total #Points	% of Total Points
Physical Science	4	13	12	1	15	38%
Life Science	3	10	9	1	12	31%
Earth and Space Science	3	10	9	1	12	31%
Total	10	33	30	3	39	100%

Note. SR = selected-response, SPT = supported performance task

Table 2.3. 2022 CoAlt Science Test Blueprint—Grade 11

Subclaim	#Item Sets	Total #Items	#1-Point SR Items	#3-Point SPT Items	Total #Points	% of Total Points
Physical Science	4	13	12	1	15	38%
Life Science	3	10	9	1	12	31%
Earth and Space Science	3	10	9	1	12	31%
Total	10	33	30	3	39	100%

Note. SR = selected-response, SPT = supported performance task

2.5. Test Composition

The Spring 2022 test forms included a set of core items held constant across all forms and a set of embedded field test items differing from form to form. The set of core items were used as pseudo-operational items for scoring purposes; only the pseudo-operational items were included in students' final scores. Table 2.4 presents the number of items on each test form, including the number of operational vs. embedded field test items and the total number of score points possible.

Table 2.4. 2022 CoAlt Science Test Designs

Grade	#Test Forms	Total #OP + FT Items	#1-Point OP SR Items	#3-Point OP SPT Items	#1-Point FT SR Items	#3-Point FT SPT Items	Total #OP Points
5	2	49	24	2	18	5	30
8	2	54	30	3	18	3	39
11	2	57	30	3	21	3	39

Note. OP = pseudo-operational, FT = field test, SR = selected-response, SPT = supported performance task

Chapter 3: Item Development

CoAlt Science follows the same development process as the CMAS Science assessment to the extent possible, although it is modified to reflect the unique characteristics of the assessment program such as the item types and 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 to ensure that the CoAlt Science assessments are equitable for students and that they accurately measure the content standards.

The item development process is a tiered, inter-related process that begins with the development of the test blueprints for each grade level, followed by developing the item development plans (IDPs) to forecast the targeted number of items needed to create a robust item bank that is refreshed over time. Once written, the items go through multiple rounds of review, including contractor, department, and Colorado educator content, bias, and data reviews. While the Spring 2022 CoAlt Science item writing was conducted internally at Pearson, all items were reviewed by Colorado educators.

3.1. Item Banking System

Pearson's proprietary software, ABBI (Assessment Banking and Building solutions for Interoperable assessments), is used to support the test development processes from initial content authoring through the review cycles. ABBI is the authoritative source for all content, data, and functionality for all CoAlt system components. It serves as the repository where the item bank is housed, item revisions are catalogued, and items and item metadata are uploaded and revised by assessment specialists. Items can be moved into various statuses, each representing a step in the item development process. The items and associated stimuli are tracked, and revisions are recorded from creation through retirement in a secure environment.

Custom development reports can be generated out of ABBI, which allows users to generate Excel reports that capture metadata (e.g., unique item number, task type, cognitive complexity, 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.

3.2. Item Development Plan

An IDP is created at the beginning of each item development cycle to determine the number of items needed to construct the assessment based on the test blueprint requirements, informing item development targets that address item shortages. The grade-level IDPs delineate the target number of items per content standard/reporting category, GLE, and EEO and help to forecast the number of items needed to create a robust operational item bank that will be refreshed over time. To accomplish this, the item bank is analyzed and gaps are identified.

3.3. Item Writing

After the test blueprints and IDPs were developed, the internal item writing process began at Pearson. SR and SPT items for each assessment were written to measure concepts and skills found in the EEOs. Item writers used various guides and resources developed during specifications development, including the content standards, item specifications, and item writing guidelines.

3.4. Item Review

3.4.1. Internal Review

After the CoAlt items were written and entered into ABBI, they underwent a content review at Pearson to evaluate the standard and knowledge-and-skill match, quality of the items, adherence to the universal design principles, cognitive demand, item relevance to the purpose of the test, readability, and appropriateness of graphics. Additional fact-checking was also conducted to ensure the accuracy of item content.

Pearson's editorial team 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. Editors with content expertise in science also reviewed the items, adding a valuable layer of content validation and fact-checking. Alternate assessment specialists with expertise in the areas of special education and students with disabilities reviewed all items to ensure that they were appropriate for students with significant cognitive disabilities. Pearson also performed a universal design 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; and
- 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. CDE then 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 accepted based on the CDE review were re-classified in ABBI as accepted.

3.4.2. External Content and Bias Review

Items that passed the internal review were included in the item pool from which items were selected for the Spring 2022 CoAlt administration. After the administration was complete, educators were convened in an educator feedback session during which participants provided feedback about the 2022 administration, including the new item types, layout, and answer documents. Educators also reviewed the items for content and bias concerns, evaluating whether they were properly aligned to the content standards and identifying any potential bias in the items while considering the unique needs of students with significant cognitive disabilities. 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 operational testing.

3.5. Data Review

Psychometricians perform statistical analyses on the field-tested items that are used to evaluate their quality. In the spring 2022 administration, all items were being presented to students for the first time, so all items were analyzed and reviewed. Table 3.1 presents the statistical flags applied to the CoAlt items. Flagged items are then reviewed by a data review committee that decides whether to accept or reject the item. Items that are accepted based on the evaluation of student performance data are re-classified in the item bank as available for use on future operational assessments. Rejected items are reclassified as "do not use" or "revise and re-field test" to eliminate them from use on an operational test. These items may be modified and field tested again on future test forms. Table 3.2 presents the results of the data review based on Spring 2022 data (i.e., the number of field-tested items that were either accepted, accepted for revision and re-field test, or rejected as a result of the data review).

Table 3.1. Item Statistical Flagging Criteria

Statistic	Criterion	Possible Indication
P-value	< 0.1 or > 0.9	Very difficult or easy item
Item-total correlation	< 0.15	Poorly discriminating item
Distractor item-total correlation (SR only)	> 0.0	Possible miskey*
Score point percentage (multi-point items only)**	<1% or >50%	Very few students or many students got a certain score
Differential item functioning (DIF)***	B, C	Item could be biased toward a certain student demographic group

*Possible miskey because the key should have a positive item-total correlation
**If a multi-point item has less than 1% for a score point or more than 50% zeros, the item is flagged.
***B DIF indicates moderate DIF, whereas C DIF indicates significant DIF.

Table 3.2. Data Review Results

Grade	#Accepted	#Accepted for Revision and Re-field test	#Rejected
5	26	0	7
8	18	0	5
11	23	0	6

Chapter 4: Form Construction

When building operational test forms, Pearson 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
- Balance of gender, ethnicity, geographic regions, and relevant demographic factors
- Thorough review of each item to verify that the content is up-to-date and relevant
- Review of the full form, including embedded 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 who each verify that the test form meets the test blueprint and test construction specifications (i.e., the required number of items, EEO coverage, and item types). The psychometrician then verifies that the test form falls within the test blueprint parameters and identifies the anchor item set within each operational test form to use for linking. (For the Spring 2022 administration, the core set of operational items on each test form was used to link the test forms within a grade.) 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.

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 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 test map that provides content and psychometric information for each item
  - manage the CDE review process
  - provide the CDE with copies of proposed items and the associated test map
  - revise the proposed item set based on CDE comments
  - document edits/comments provided by CDE

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

Chapter 5: Test Administration

Beginning with the Spring 2022 administration, a scannable answer document was used for CoAlt. Online score entry through PearsonAccess^{next} is no longer used, and the Test Examiner role no longer exists in the system. There is no minimum or maximum testing time, and testing may extend over multiple days for a student. The assessment may be stopped and restarted at any time, but once an item is presented, it should be completed before stopping. CoAlt Science administration and training procedures were standardized to ensure that students would receive comparable test results while allowing flexibility to accommodate the unique needs of students in this population. Test administration procedures were communicated to the appropriate individuals via manuals and virtual and recorded trainings.

The District Assessment Coordinator (DAC) is responsible for establishing the administration schedule and ensuring that every student taking a CoAlt Science assessment is assessed within the state assessment window. Districts may use the entire state testing window for administration of this assessment, but it is expected that students taking the CoAlt Science assessment will test during the same testing window as their peers taking the CMAS assessments. It is important that scheduling of the assessment is based on the individual needs of the student.

5.1. Manuals

The following manuals are available online at https://coassessments.com/manuals/ to support the CoAlt Science administration:

- The *CoAlt Science Test Administrator Manual* provides instructions for administering the CoAlt Science assessments, including scoring procedures, as well as the before, during, and after testing tasks for the Test Administrator. Test administration policies and procedures, including scoring information, are to be followed as written so that all testing conditions are uniform statewide, ensuring that every student in Colorado receives the same standard directions and scoring during the test administration.
- The *CMAS and CoAlt Procedures Manual* provides instructions for the coordination of the CoAlt Science assessments. Instructions include the protocols that all school staff are to follow related to test security, test administration, and providing accommodations. The manual also includes the tasks to be completed by DACs, School Assessment Coordinators (SACs), and District Technology Coordinators (DTCs) before, during, and after the test administration.
- The *PearsonAccess^{next} Online User Guide* provides guidance for DACs, SACs, DTCs, and student enrollment personnel who use PearsonAccess^{next}, the website used for student registration, test setup, administration preparation, and assessment and data management.

5.2. Test Materials

Table 5.1 presents the paper-based test materials used by the Test Administrator during the administration of the CoAlt Science assessment, distributed by the SAC, as provided in the *CMAS and CoAlt Procedures Manual*. For the SR items, the student marks/points/indicates their response in the test book, and the Test Administrator marks the student answer on the answer document. SPT items have cutout cards that the student places/indicates placement of in the

correct box in the test book. The Test Administrator scores the student response and marks the student's score in the answer document.

Table 5.1. CoAlt Science Test Materials

Test Material	Description
Test Books	Test Administrator reads an administration script which faces the Test Administrator while student response pages face the student.
CoAlt Test Administrator Manual	Test Administrator uses the manual to support test administration and scoring. The CoAlt Test Administrator Manual contains the SPT Score Flow Chart for scoring SPT items.
Task Manipulatives	Students use task manipulatives to respond to the SPT items. Prior to testing, Test Administrators must prepare the task manipulatives by cutting them apart.
Answer Document	Test Administrators use the answer document to record student responses during testing. After testing, answer documents are returned to Pearson for scoring
Secure Return Envelope	Transport test materials between the testing environment and the central storage area in an unsealed secure return envelope. (Note: Test books will not fit in the envelopes.)

5.3. Administration Training

Administration training is intended to make sure all individuals involved in the 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 virtual trainings contained information regarding proper procedures for administration. 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. The trainings were posted on the CDE website at http://www.cde.state.co.us/assessment/trainings-archive. Administration training materials such as web-based modules, slide decks, and manuals were also available on the CDE website for training SACs. After CDE trained DACs and special education staff, these individuals trained SACs and any other individuals within the district who planned to participate in the CoAlt Science assessment 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 and Pearson 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 ensure that appropriate answers were provided. Policy questions received by the Pearson customer service center were referred to CDE.

5.4. Practice Resources

Colorado Practice Resources (CPRs) are available online at https://coassessments.com/practice-resources/ to help students become familiar with the SR and SPT item types on the CoAlt

Science assessments. Each grade has multiple SR clusters and SPT samples. As the assessment system progresses, the CPRs will be updated to reflect the current assessment.

5.5. Accessibility Features 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. The assessments can also 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, 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.

In addition to incorporating accessibility into the assessment, accommodations are also available to students who need additional changes to the test administration to access the assessment. 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

5.6. Test Security

Test security procedures are put in place to enhance the likelihood that security is maintained before, during, and after assessment administration. For example, materials used during the administration of the assessment are to be kept in locked storage locations when not under the direct supervision of Pearson or approved assessment coordinators and administrators. All district and school personnel involved in the CoAlt Science test administration are required to participate in annual local training. DACs and district special education staff are responsible for overseeing training for the district, including verifying that the SACs are trained. SACs are responsible for ensuring that all individuals involved in handling test materials at the school level are trained and subsequently act in accordance with all security requirements.

A chain of custody plan for materials is required to be written and implemented to ensure that materials are securely distributed from DACs to SACs to Test Administrators and securely returned from Test Administrators to SACs and then to DACs. SACs are required to distribute materials to and collect materials from the Test Administrators each day of testing and to securely store and deliver materials to DACs after testing is completed in accordance with the instructions in the *CMAS and CoAlt Procedures Manual*.

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

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

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

Chapter 6: Scoring

The SR items are machine-scored, while each of the three prompts in an SPT item is scored by the Test Administrator using a rubric to evaluate student performance.

6.1. Machine Scoring

The machine-scored items are key-based multiple-choice items. Initial scoring expectations are developed during item development and are included in the item review process. The scoring rules and correct responses are included in the items' XML coding. Prior to scoring, key checks are completed for all machine-scored items to verify that the machine is correctly identifying correct and incorrect responses. If there is a discrepancy in the scoring, content experts review the item and adjustments are made as needed. During testing, actual distribution of scores is compared to expected distribution. Further evaluation is completed if a discrepancy is identified.

6.2. Test Administrator Scoring

SPT items consist of three related items called prompts. Students are required to manipulate option cards by placing them in designated areas on a diagram or chart to respond to each of the three prompts. Student performance on each prompt is scored using a 1-point rubric, as shown in Table 6.1. To administer the item, the Test Administrator has the student response page and option cards ready for the student to engage with the item. The Test Administrator then presents the scripted text for the first prompt. Scores are assigned by the Test Administrator based on the following scenarios:

- If the student responds correctly, they receive 1 point.
- If the student responds incorrectly, they receive 0 points.
- If the student does not provide a response to the prompt, they receive an NR, or no response, which represents 0 points.

Table 6.1. SPT Scoring Rubric

Score Point	Requirement
1	Student responds correctly
0	Student responds incorrectly
NR	Student does not respond

Note. NR = no response, which represents 0 points. This rubric is used for each of the three prompts within each task.

If an incorrect response is given or the student does not respond, the Test Administrator places the correct option card in the response box and tells the student the correct answer. After the first prompt is completed, the Test Administrator completes the same steps for the remaining two prompts. For scoring and reporting purposes, the points for the three prompts are then added together to provide one score for the SPT item that can range from 0–3 points.

Chapter 7: Standard Setting

The purpose of a standard setting study is to determine the boundaries—or cut scores—along the score scale that differentiate student performance among performance levels (e.g., Cizek et al., 2004; Kane, 1994). Standard setting for the new science assessment took place in Fall 2022 using the Spring 2022 data; the standard setting report is still in progress, and details of the meeting and the final cuts will be included in the 2022–2023 technical report.

Chapter 8: Reporting

Because Colorado students saw CMAS and CoAlt Science items for the first time in spring 2022, standard setting was held after the spring 2022 reporting cycle. Science reporting for spring 2022 assessments provided normative achievement indicators in the form of percentile ranks but did not include criterion referenced reporting. Scale scores and performance levels will be reported starting with the Spring 2023 administration.

Two types of score reports were provided for the Spring 2022 CoAlt Science administration: (1) the student-level Individual Student Performance Report that provides information about the performance of a student on the CoAlt Science assessment and (2) the School Summary of Students Report that provides participation rates and median percentile ranks for the state, district, and school for the overall test. Percentile ranks and quartile indicators for the overall test are also provided for each student in the school. Appendix B presents a sample student performance report.

Individual Student Performance Reports were printed and shipped to districts for distribution to students and parents/guardians, whereas the School Summary of Students Report was provided electronically through PearsonAccess^{next}. Access to the reports was limited to users approved by Colorado school districts. For additional details on score reports, see the *CMAS and CoAlt Interpretive Guide to Assessment Reports*.

Chapter 9: Test Results and Analysis

This chapter presents the test results and statistical analyses for the Spring 2022 CoAlt Science administration.

9.1. Student Participation

Table 9.1 presents a breakdown of the number of students who took the assessment by various demographic characteristics. All forms were administered in paper format. Approximately 1,230 students took the CoAlt Science assessment in Spring 2022.

Table 9.1. Student Participation N-Count Demographic Distribution

Subgroup	Grade 5	Grade 8	Grade 11
Total	**395**	**429**	**404**
No IEP	0	0	0
IEP	395	429	404
No Accommodation	307	318	357
Accommodation	88	111	47
Am. Indian/Alaska Native	*	*	*
Asian	20	22	*
Black	33	29	32
Hispanic	163	161	168
White	153	186	166
Hawaiian/Pacific Islander	*	*	*
Two or More Races	22	23	21
Missing	*	*	*
No Economic Disadvantage	175	202	223
Economic Disadvantage	220	227	181
Female	140	154	149
Male	255	275	255
Language Proficiency NA	306	329	291
Language Proficiency NEP	59	46	28
Language Proficiency LEP	*	*	*
Language Proficiency FEP	*	39	52
Language Proficiency FELL	*	*	*
Language Proficiency PHLOTE	*	*	*
Missing	*	*	*
Not Migrant	388	419	397
Migrant	*	*	*
Missing	*	*	*

*n-count less than 16

9.2. Percentile Rank Performance Results

Table 9.2 presents the percentile rank performance summary. Each grade has a mean percentile rank of 50, which is to be expected given that most students were expected to fall in the middle of the distribution.

Table 9.2. Percentile Rank Performance Summary

Grade	N	Mean	SD	Median*
5	395	50	28.8	47
8	429	50	28.9	48
11	404	50	28.9	50

*The median is not 50 where there are an even number of obtained scores being ranked.

9.3. Subclaim Correlations

The CoAlt Science assessments have three subclaim scores: Physical Science, Life Science, and Earth and Space Science. One way to assess the internal structure of a test is through the evaluation of correlations among subclaim subscores, as presented in Table 9.3. The intercorrelations between the subclaims were between 0.45 and 0.68. The correlations between Physical Science and Earth and Space Science in Grades 5 and 8 tended to be higher than the correlations of those subclaims with Life Science.

Table 9.3. Correlations Between Subclaims

Grade	Subclaim	*Life Science	Earth and Space Science	Total Test
5	Physical Science	0.451	0.602	0.832
	Life Science*	–	0.479	0.760
	Earth and Space Science	–	–	0.867
8	Physical Science	0.674	0.684	0.908
	Life Science	–	0.612	0.861
	Earth and Space Science	–	–	0.865
11	Physical Science	0.613	0.594	0.863
	Life Science	–	0.634	0.866
	Earth and Space Science	–	–	0.855

*For Grade 5, the subclaim is Physical Science/Life Science.

Chapter 10: Calibration

The item response theory (IRT) Rasch Partial Credit model (RPCM) was used to calibrate the CoAlt Science assessments. All calibrations were accomplished within the IRT framework. Calibration for the operational and field test items were as follows for the Spring 2022 administration: Used Winsteps (Linacre, 2021) control files and incomplete data matrix (IDM) to obtain the operational and field test item parameter estimates. The entire process was completed for each CoAlt Science assessment. All steps were independently replicated by at least two members of the Pearson psychometrics team to ensure the accuracy of the processes.

10.1. IRT Model

RPCM 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 is a mathematical measurement model with a single item parameter relating a student's performance on a given item involving *m*+1 score categories. The probability of student *n* scoring *x* on *m* steps of item *i* is a function of the student's proficiency level, θ_n (also referred to as "ability"), and the step difficulties, δ_{ij}, of the *m* steps in item *i* as follows:

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

10.2. 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.
- IDMs were created.

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

- P-value <0.15
- Item-total score correlation < 0.10
- Incorrect option selected by more high-ability students (top 33%) than the keyed response
- Distractor *p*-value ≥ 40%
- Distractor-total score correlation > 0
- One or more score points earned by less than 1% of students

10.3. Calibration

Calibration refers to the estimation of item parameters in the IRT framework, which places items and students on a common scale. To obtain item parameter estimates for CoAlt Science, the RPCM was applied to the items. Winsteps was used for all calibrations, and all operational and field test item parameters were estimated in a single calibration (i.e., a concurrent calibration) for each assessment. The calibration supported the Fall 2022 standard setting and Spring 2023 test construction.

Chapter 11: Reliability

The *Standards for Educational and Psychological Testing* (AERA et al., 2014) refer to reliability as the "consistency of scores across replications of a testing procedure" (p. 33). A reliable test produces stable scores; very similar score distributions would result if the test were administered repeatedly under similar conditions to the same students without memory or fatigue affecting the scores. The level of reliability/precision of scores has implications for validity. In other words, scores must be consistent and precise enough to be useful for intended purposes. If scores are to be meaningful, tests should produce stable scores if the same group of students were to take the same test repeatedly without any fatigue or memory of the test. The range of certainty around the score should also be small enough to support educational decisions. Reliability for the Spring 2022 CoAlt Science administration was evaluated with the following analyses:

- Internal consistency (coefficient alpha)
- Standard error of measurement (SEM)

11.1. Internal Consistency (Coefficient 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. Internal consistency is typically measured via correlations among the items on an assessment and provides an indication of how much the items measure the same general construct. High reliability of test scores implies that the test items within a subclaim are measuring a single construct, which is a necessary condition for validity when the intention is to measure a single construct.

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). In the internal consistency method used to estimate reliability for the CoAlt Science assessments, a single form is administered to the same group of students to determine whether students 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 and polytomous test items and is computed as follows:

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

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.

Coefficient alpha ranges from 0.0 to 1.0, where higher values indicate a greater proportion of observed 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 students will respond consistently across items within the test.

Coefficient alpha estimates for CoAlt Science are provided for the overall test and by subclaim, as shown in Table 11.1. The coefficient alpha for the total group across the science assessments ranged from 0.78 to 0.84.

Table 11.1. Coefficient Alpha

Grade	Physical Science	*Life Science	Earth and Space Science	Total Test
5	0.53	0.57	0.60	0.78
8	0.66	0.62	0.61	0.84
11	0.60	0.64	0.60	0.82

*For Grade 5, the subclaim is Physical Science/Life Science.

11.2. Standard Error of Measurement (SEM)

The SEM is another measure of reliability. This statistic uses the standard deviation of test scores along with a reliability coefficient (e.g., 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 and SEM: the higher the reliability, the lower the SEM. Table 11.2 presents the SEM results by subclaim for the CoAlt Science assessment. The SEM values for the total group ranged from 2.64 to 3.07.

Table 11.2. SEM

Grade	Physical Science	*Life Science	Earth and Space Science	Total Test
5	1.52	1.35	1.62	2.64
8	1.86	1.67	1.68	3.02
11	1.87	1.65	1.70	3.07

*For Grade 5, the subclaim is Physical Science/Life Science.

Chapter 12: Validity

"Validity refers to the degree to which evidence and theory support the interpretations of test scores for proposed uses of tests" (AERA et al., 2014). As such, it is not the CoAlt Science assessments that are validated but rather the interpretations of the scores. The purpose of the CoAlt Science assessment is to provide information about a student's level of mastery of the EEOs of the CAS. In support of this, this technical report has described processes that were implemented throughout the CoAlt Science assessment cycle with validity and fairness considerations in mind. This chapter describes the various sources of validity evidence as outlined in the *Standards for Educational and Psychological Testing* (AERA et al., 2014), often referencing other chapters and sections of this report. As the CoAlt Science assessments mature, validity evidence supporting the assessments' interpretations will continue to be collected and documented.

12.1. 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 a focus on this goal, as outlined in Chapters 2 and 3 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. Statistical bias analyses (i.e., DIF analyses) were also conducted on the 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 to most optimally measure the EEOs. A formal alignment study will be conducted for CoAlt Science in 2023 to provide further evidence to support the claim that the content of the CoAlt Science test items match the intended content as specified in the EEOs.

12.2. 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 Administrators were asked a set of questions about students' instruction, their communication modes, and their item responses. These results, presented in Appendix C, help support the validity of the students' responses on the assessment.

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 most teachers believe that students are using their content knowledge to answer test items, although 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 most students use oral administration or picture communication to receive information, and they use these same methods when responding to others.

12.3. 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 as provided in Section 11.1.

12.4. Evidence Based on Relations to Other Variables

Evidence was collected showing the correlation between student scores and variables related to the student. Student test scores were correlated with Test Administrators' responses in Appendix C for several test validity questions to determine the strength of relationship between the variables. Table 12.1 presents the correlation coefficients between the student scores and these variables, providing validity evidence based on relations to other 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? Approximately how much instructional time for this content area is in the general education classroom?).

As shown in Table 12.1, the correlations between student scores and the familiarity of the Test Administrator with the student are small and indicate no meaningful relationship between the variables. The correlations between student scores and the instructional hours and instructional time variables are low positive correlations. The strength of these relationships will be reviewed for future administrations as Test Administrators and students have more opportunity to engage with the CAS in the classroom setting.

Table 12.1. Correlation Between Test Validity Questions and Student Scores

	Familiarity with the Student		Hours Per Week in Instruction on the Content Area		How Much Instructional Time in the Content Area Is in the General Education Classroom	
Grade	N	Correlation	N	Correlation	N	Correlation
5	375	-0.04	373	0.19	370	0.21
8	417	-0.06	417	0.26	419	0.25
11	374	-0.03	373	0.17	374	0.22

12.5. Evidence for Validity and Consequences of Testing

As the CAS become more fully integrated into the classroom, and with additional administrations of the CoAlt Science assessment, it is intended that information around the consequences of the assessment will be collected. Some of the intended consequences include the appropriate use of the assessment for students with the most significant cognitive disabilities, the inclusion of those students in the state assessment system, and the effective instruction of students with the most significant cognitive disabilities in the EEOs of the CAS. Longitudinal data will start being collected in Spring 2023, which will be the initial operational administration.

12.6. Fairness

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

- Sample items that 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 (Section 5.4)
- Universal design principles that are adhered to during the test development process with the goal of avoiding construct-irrelevant aspects of the assessment that could impact student performance (Chapter 3)
- DIF analyses to identify any items that appear to be unfairly favoring one subgroup over another. All items which show DIF are reviewed by educators for potential bias in the item. (Chapter 3)
- Accessibility tools and accommodations to allow students to fully demonstrate their content knowledge without being hindered by non-construct related elements in addition to being developed to be accessible for students with significant cognitive disabilities (Chapters 2 and 3, Section 5.5)

References

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

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.* Harcourt Brace Jovanovich College Publishers.

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

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

Kane, M. (1994). Validating the performance standards associated with passing scores. *Review of Educational Research*, *64*(3), 425–461.

Linacre, J. M. (2021). Winsteps® (version 4.8.1.0) [computer program]. Winsteps.com.

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). Science Research Associates.

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

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

Appendix A: CoAlt 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 significant 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** and ☐ 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 alternate 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 Grade-level 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 Alternate 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: Sample Student Performance Report

Confidential Student Performance Report

Colorado Alternate Assessment

Student: **FIRSTNAME003 O. (ROBBIE) LASTNAME003**

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

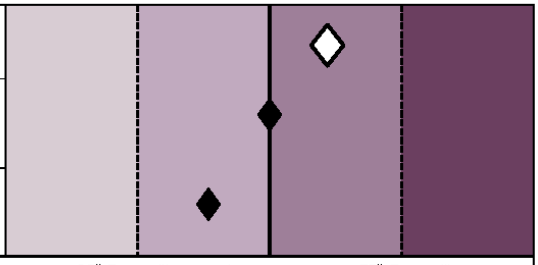
Spring 2022

Science **Grade 5**

The 2020 Extended Evidence Outcomes of the Colorado Academic Standards in science were tested for the first time through the spring 2022 Colorado Alternate (CoAlt) Assessment. This report provides information about how your student performed on that test in comparison to other Colorado students who took the test.

- Your student's percentile rank is represented by the white diamond on the graph below and indicates the percentage of students whom your student performed higher than in the state. For example, a student in the 37th percentile performed higher than 37% of the students who took CoAlt in the state.
- You may compare your student's performance to the performance of other groups of students. State and district average percentile ranks are represented by black diamonds on the graph. To protect student privacy, scores are not displayed for districts that do not meet the minimum number of students for reporting. The solid line on the graph indicates the 50th percentile and represents the state average.

Percentile ranks can change based on the group of students tested. Higher participation rates (the percentage of students who took the test) allow for stronger comparisons to all students in the state and district. Participation rates are included in the first column in the graph below.

Your student's percentile rank:	**XXth**	
Average state percentile rank: (State participation rate: xx%)	50th	
Average district percentile rank: (District participation rate: xx%)	XXth	

5

This report provides information on your student's performance in relation to other students' performance. Unlike other CoAlt reports, it does not provide information on your student's level of mastery of the concepts and skills in the Extended Evidence Outcomes of the Colorado Academic Standards. Performance levels will be determined in the future, in part based on student performance on this 2022 test.

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.

mmddyyy-Batch-9999-9999-1234567

Appendix C: Test Administrator Survey Responses

How familiar are you with this student?

Grade	Very Familiar	Somewhat Familiar	Familiar	Somewhat Unfamiliar	Unfamiliar	Missing
5	84.56%	5.82%	3.29%	0.76%	0.51%	5.06%
8	85.78%	5.59%	4.66%	0.23%	0.93%	2.80%
11	81.44%	7.92%	2.23%	0.25%	0.74%	7.43%

How many hours per week does this student spend in instruction on this content area?

Grade	<1 Hour	1 to <2 Hours	2 to <3 Hours	3 to <4 Hours	4 to<5 Hours	≥5 Hours	Do Not Know	Missing
5	20.76%	28.35%	22.03%	13.92%	5.82%	3.54%	0.00%	5.57%
8	14.45%	17.48%	11.89%	17.02%	31.24%	5.13%	0.00%	2.80%
11	13.37%	14.60%	15.35%	20.30%	21.04%	7.67%	0.00%	7.67%

Approximately how much instructional time for this content area is in the general education classroom?

Grade	25%	50%	75%	100%	None	Missing
5	21.01%	8.86%	12.91%	24.81%	26.08%	6.33%
8	12.12%	5.83%	11.19%	35.66%	32.87%	2.33%
11	11.14%	4.95%	4.70%	21.78%	50.00%	7.43%

This student's primary receptive communication is:

Grade	Oral Language	Sign Language	Reading	Picture Communication	Tactile	Other	Do Not Know	Missing
5	83.80%	0.51%	0.25%	5.32%	0.25%	1.77%	0.00%	8.10%
8	87.65%	0.47%	2.33%	3.50%	0.70%	0.23%	0.70%	4.43%
11	86.14%	0.74%	0.99%	2.23%	0.00%	0.74%	0.25%	8.91%

This student's primary expressive communication is:

Grade	Oral Language	Sign Language	Writing	Picture Communication	Augmentative Communication Device	Tactile	Other	Do Not Know	Missing
5	72.40%	0.76%	0.00%	3.54%	8.61%	0.00%	5.32%	0.00%	9.37%
8	75.99%	0.70%	2.56%	5.36%	7.23%	0.47%	2.33%	0.70%	4.66%
11	79.46%	1.49%	0.00%	1.73%	5.20%	0.25%	1.98%	0.25%	9.65%

I feel that the student's responses accurately reflect their understanding of the material.

Grade	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree	Do Not Know	Missing
5	29.87%	35.44%	17.22%	6.58%	3.80%	1.01%	6.08%
8	32.40%	35.43%	16.32%	7.93%	2.80%	0.93%	4.20%
11	33.17%	36.63%	13.12%	5.69%	2.23%	0.99%	8.17%

How much time did this student take on the assessment?

Grade	0–30 Minutes	31–60 Minutes	61–90 Minutes	91–120 Minutes	121–150 Minutes	151–180 Minutes	181–210 Minutes	≥211 Minutes	Missing
5	2.53%	41.77%	30.63%	11.39%	3.04%	1.27%	0.76%	1.27%	7.34%
8	1.17%	45.69%	34.50%	9.56%	3.03%	0.23%	1.40%	0.93%	3.50%
11	1.98%	24.75%	44.06%	14.60%	2.72%	1.98%	0.25%	0.74%	8.91%