NM-Assessment of Science Readiness: Test Specifications

NEW MEXICO ASSESSMENT OF SCIENCE READINESS

Purpose

- Part of a Balanced Assessment System
- Claims/Score Interpretation and Use Statements

Test Specifications

- Reporting Categories
- Test Design
- Blueprint
- Fairness
- Universal Design for Assessments
- Bias
- Sensitivity

Item Specifications

- Alignment
- Item Types
- Sample Items

Purpose

Part of a Balanced Assessment System

The NM-ASR is one assessment among many that are a part of New Mexico's Balanced Assessment System, designed to provide point-in-time information about the academic achievement and progress of New Mexico students. Student results are reported according to academic achievement descriptors utilizing scale scores for each of four performance levels: Advanced, Proficient, Nearing Proficiency, and Novice. The results from these assessments provide educators and the public with information to guide the creation of future educational practices to meet the needs of students, while monitoring the continuous improvement efforts of schools, districts, and the state in achieving a world-class education system for all students.

The Assessment of Science Readiness focuses on the integration and application of disciplinary core ideas, science and engineering practices, and crosscutting concepts to engage in sense-making around scientific phenomena and engineering design problems. Students demonstrate their acquired skills and ability by answering various types of questions such as multiple-choice items, multiple-select items, technology-enhanced items, and open-ended items. Many of the items were grouped together in clusters with a common stimulus, to allow for better assessment of the depth of the constructs in the standards.

Claims/Score Interpretation and Use Statements

The NM-ASR provides reliable and valid information about whether students are on track to be ready for college or career, as defined by the State, by showing they have mastered the New Mexico *STEM Ready! Science Standards*. In addition to the overall scale score, student performance on three science domains: life science, physical science, and Earth and space science is also reported. Scale scores can be used to compare an individual student's performance to the performance of other students in the school, district, and state. Additionally, student subgroups (e.g., English learners, students with disabilities, racial/ethnic subgroups) are also reported at the school, district, and state level.

Test Specifications

Reporting Categories

The reporting categories for the NM-ASR Assessment are based on the science domains in the New Mexico STEM Ready! Science Standards as noted below.

Grade 5

- Practices and Crosscutting Concept in Life Sciences
- Practices and Crosscutting Concept in Physical Sciences
- Practices and Crosscutting Concept in Earth and Space Sciences

Grade 8

- Practices and Crosscutting Concept in Life Sciences
- Practices and Crosscutting Concept in Physical Sciences
- Practices and Crosscutting Concept in Earth and Space Sciences

Grade 11

- Practices and Crosscutting Concept in Life Sciences
- Practices and Crosscutting Concept in Physical Sciences
- Practices and Crosscutting Concept in Earth and Space Sciences

Test Design

The NM-ASR test design is based on the three content domains of Physical Sciences, Life Sciences, and Earth and Space Sciences. Items are expected to align to the multiple dimensions of the standards (Disciplinary Core Ideas, Science and Engineering Practices, Crosscutting Concepts) in each domain, such that every item is at least two-dimensional, if not three-dimensional. To emphasize this multi-dimensional nature of the items, the names of the reporting categories incorporate the three dimensions (Practices and Crosscutting Concepts in Physical Sciences, Practices and Crosscutting Concepts in Life Sciences, Practices and Crosscutting Concepts in Earth and Space Sciences). Students are expected to demonstrate sense-making by using core ideas, practices, and crosscutting concepts together to respond to items.

Items assessing Engineering, Technology, and Applications of Science as well as the New Mexico-specific content domain of Science and Society are reported within the Physical, Life, or Earth and Space Sciences category, depending on the content match of the design problem presented in the item.

Blueprint

Science Reporting Category	Grade 5		Grade 8		Grade 11	
	# of Points	% of Points	# of Points	% of Points	# of Points	% of Points
Practices and Crosscutting Concept in Earth and Space Sciences	18-22	28-35%	18-22	28-35%	18-22	26-32%
Practices and Crosscutting Concept in Life Sciences	18-22	28-35%	20-24	31-38%	22-26	32-39%
Practices and Crosscutting Concept in Physical Sciences	24-28	38-44%	20-24	31-38%	22-26	32-39%
Total	64	100%	64	100%	68	100%

Fairness

Fairness is defined as the extent to which the test scores are valid for different groups of test takers. Consideration of universal design, bias, and sensitivity guidelines support the construction of fair, valid assessments.

Universal Design for Assessments

The concept of Universal Design for Assessments focuses on developing content and assessments that reach the widest population of students possible. Stimuli and items on the NM-ASR are designed to simply and clearly present tasks and to provide maximum readability, comprehensibility, and legibility. The Universal Design process facilitates item development that is sensitive, fair, and free of bias. This minimizes construct-irrelevant variance due to differences in a test-taker's socioeconomic status, cultural background, disability, or linguistic background. The Universal Design approach includes:

- Minimizing features that are irrelevant to what is measured
- Using accessible authoring and accessible technologies from the inception of content
- Designing items to address the full range of complexity of content standards
- Using technology for accessible test delivery
- Using simple sentence structure
- Minimizing words with multiple meanings
- Avoiding slang and regional dialects
- Avoiding complicated names or names with dual meanings
- Labeling graphics clearly

As part of the general assessment development process, items undergo reviews by accessibility experts and a dedicated bias and sensitivity panel. Items are then reviewed externally by the Bias and Sensitivity Educator Review Committee. This ensures that the language and context used is appropriate and understandable for the intended grade-level of the test takers. Passages go through several additional rounds of review and analysis, both qualitative and quantitative, to ensure texts are engaging, understandable, and appropriate for the grade where they are intended to be used.

Bias

The concept of Bias is defined as the presence of some characteristic of an item that results in differential performance for two individuals of the same ability but from different ethnic, sex, cultural, or religious groups.

Bias can occur whenever content offends or disadvantages a student or group of students due to gender, race, regional background, socioeconomic status, or any other such classification.

Test developers take care to craft content in a way that does not misrepresent specific groups or rest on assumptions made about specific groups, that in turn could negatively impact how students interpret content.

- Stimulus and item content on the NM-ASR must not present stereotypes or unfair representations of gender, race, ethnicity, disability, culture, or religion.
- Stimulus and item content on the NM-ASR should not depend on overly experiential information such as knowledge of technology, consumer goods, pop culture, geographic locations, or sports and extracurricular activities. While these topics are not completely excluded from use, care must be taken to ensure that the items are presented in a way that does not require a level of knowledge that would not be held by all students.

Sensitivity

Sensitivity refers to the presence of content that is contrary to the acceptable norms of the students, educators, parents, or other members of the community that may interact with the assessment. Sensitive subject matter can impact student performance or attitudes toward testing, and hence, their test scores.

Consideration of bias and sensitivity issues is very important when developing content for an assessment. Test developers must ensure that stimuli and items are free of content that will negatively affect a student's performance not because of what the student knows and can do but because the content evokes an emotional response from that student (or is in some other way distracting to the student).

Subjects/contexts that are likely to prompt emotional distress on the part of students cannot be used on the NM-ASR (e.g., war, violence, human death or debilitating disease, animal-based medical research). Careful judgment should be applied to PEs that cover topics that may be considered controversial by some groups (e.g., evolution examples, population dynamics including death/extinction, environmental impact). Those PEs represent content knowledge to be assessed, but the assessment must be done in a sensitive, unbiased way. Careful attention was also paid to a list of topics to avoid provided by New Mexico.

Item Specifications

Alignment

The items on the NM-ASR are aligned to the New Mexico *STEM Ready! Science Standards*. Each item is aligned to one of Performance Expectations (PEs) of the content standards.

Item Types

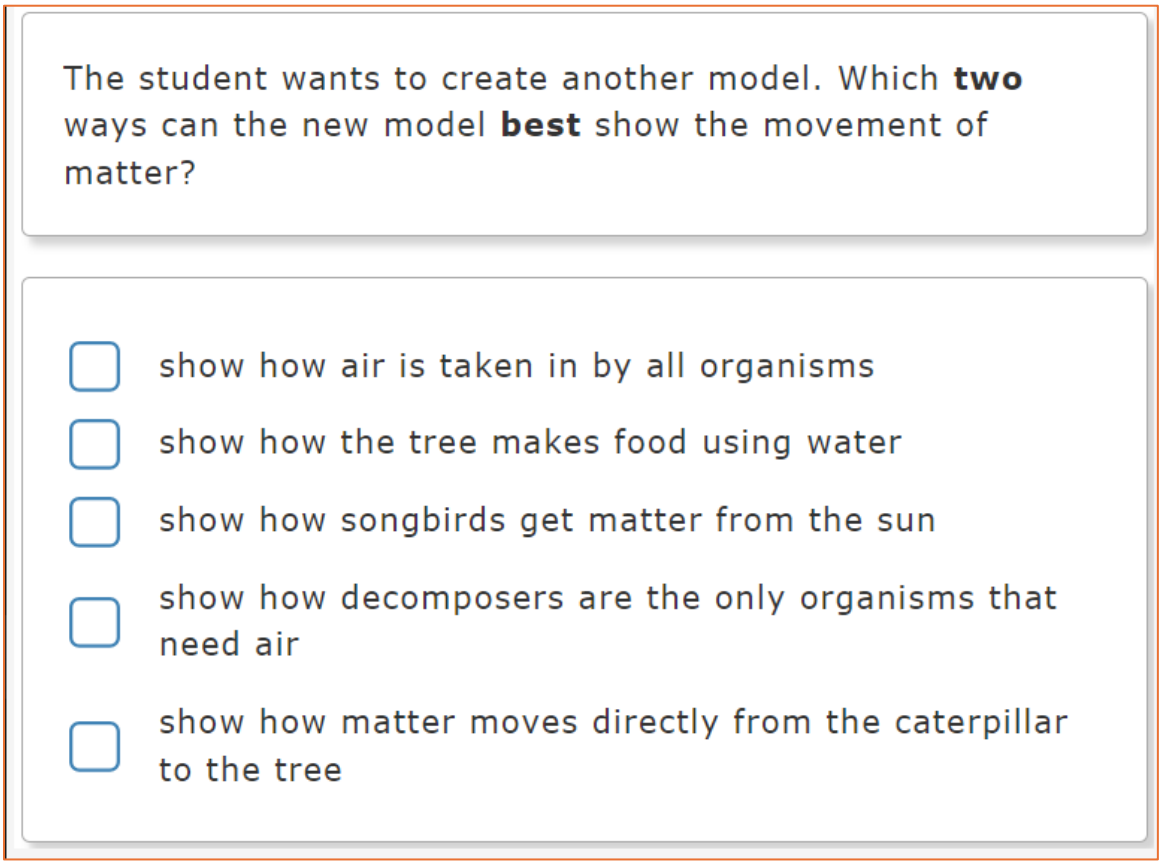
The types of items on the NM-ASR are machine-scored items worth 1-2 points and human-scored items worth up to four points:

- **Stimulus-based item blocks** are items associated with a science phenomenon. There are 2–6 machine-scored items associated with the stimulus; each associated item may be worth 1 or 2 points.
- **Stand-alone items** are machine-scored items that are self-contained; each item may be worth 1 or 2 points.
- **Open-ended items** are human-scored items that are worth up to 4 points and in a specific reporting category. An open-ended question is associated with each of the three reporting categories: life sciences, physical sciences, Earth & space sciences.

Sample Items

- Stimulus-based and stand-alone items include many different item types. All of these items are machine scored and may be worth 1 or 2 points. Among the specific item types are multiple choice, drop-down, matrix, matching lines, background graphic, hot spot, multiple drop bucket, ordering, and composite.

Multiple Choice Item:



Drop-down Item:

Select the words that correctly describe Pluto's movement in the solar system.

The effect of gravity on Pluto's movement is best shown in the Select because it shows how Pluto's Select

Matrix Item:

The student wants to show how blight changes the bur oak tree ecosystem. Choose how blight affects the movement of matter in each relationship.

Relationship	Increase	Decrease	Stay the Same
bur oak tree to caterpillars	○	○	○
bur oak tree to decomposers	○	○	○
bur oak tree to mammals	○	○	○
bur oak tree to songbirds	○	○	○

Matching Lines Item:

Use figure 1. Match each organism with its role in transfer of matter.

Organism	Role
caterpillar	producer
fungi	consumer
bur oak tree	decomposer

Background Graphic Item:

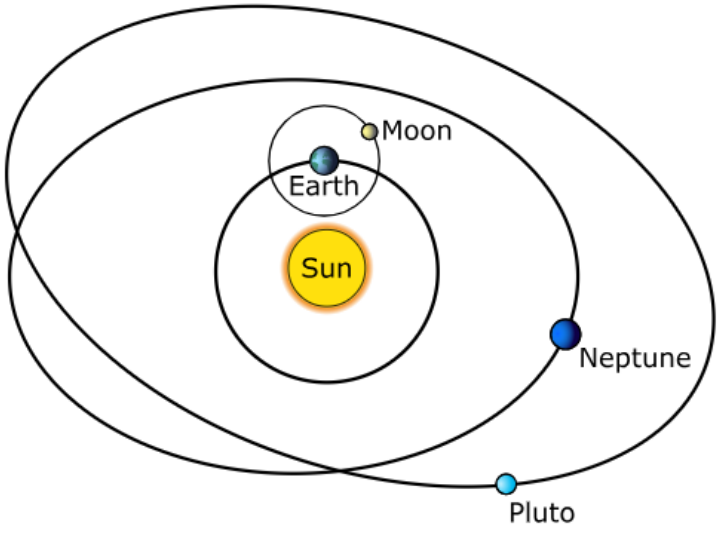
Complete the model to compare the lifespan and the rate of energy produced by each star.

10 thousand	2 billion	10 billion
greatest rate	second-greatest rate	third-greatest rate

Star Name	Lifespan (years)	Rate of Energy Produced
Sun		
Sirius A		
Westerhout 51-57		

Hot Spot Item:

The figure shows the Sun, Earth, the Moon, Neptune, and Pluto. Select the object that plays the **greatest** role keeping Pluto in orbit.



Multiple Drop Bucket Item:

Place the transportation systems that meet each criterion. More than one transportation system may be placed for each criterion. Transportation systems may be used more than once or not at all.

Transportation System

- Public bus
- Electric bike share
- Car share
- Electric scooter share

Pollution

Cost

Years until Replacement

Ordering Item:

The steps for the events in a star's lifespan are out of order. Put the steps in chronological order to show the events in a star's lifespan.

- star produces light
- fusion of hydrogen begins
- core of star contracts and collapses
- stellar nebula reaches a critical mass
- star expands to become red giant or red supergiant

Composite Item:

This question has two parts. First, answer part A. Then, answer part B.

Part A

Which trait did the common ancestor of the two birds **most likely** have?

- ○ solid bones
- ○ short flippers
- ○ no flying ability
- ○ thick and pointy beak

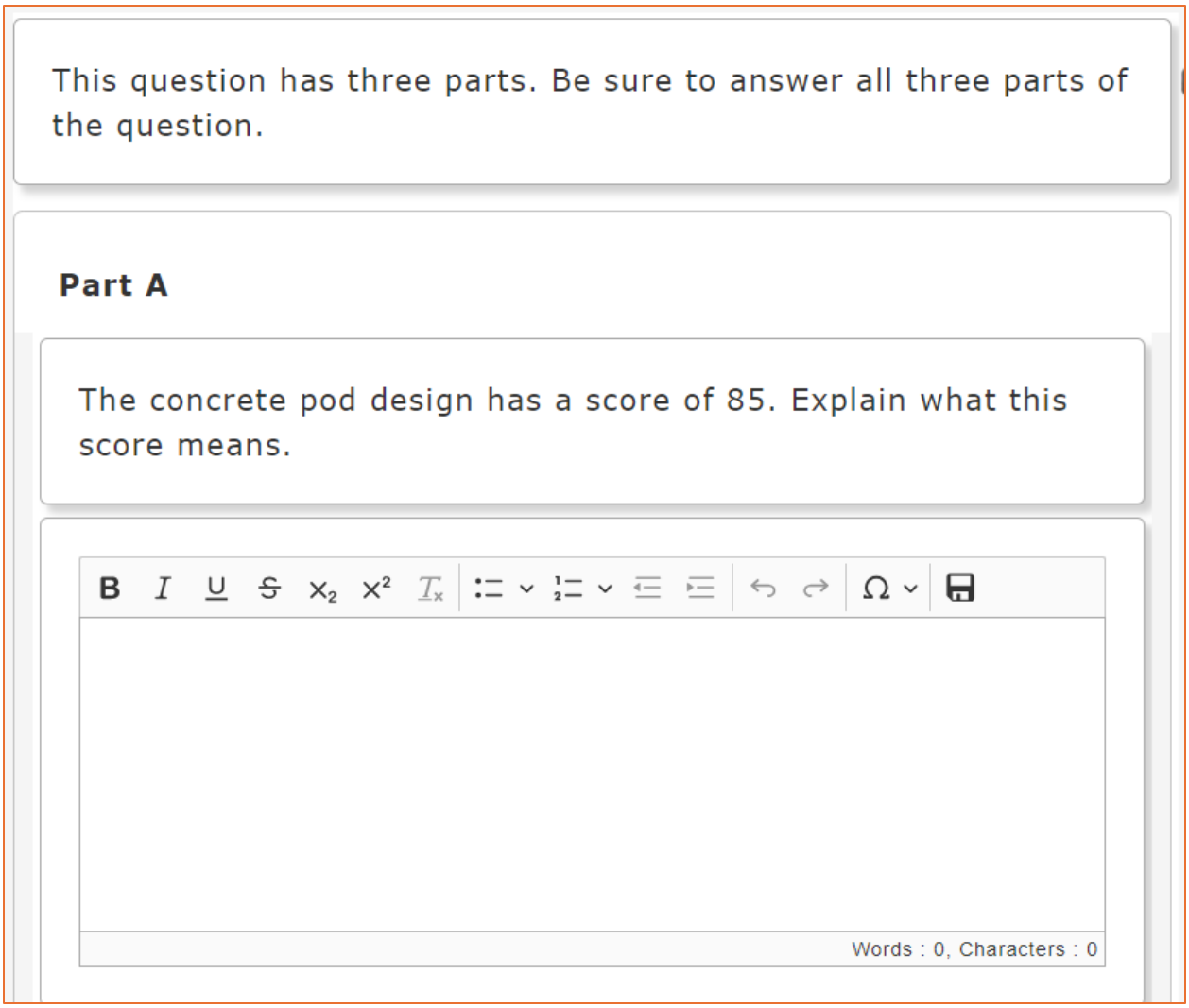
Part B

Which statement **best** explains your answer from part A?

- ○ The common ancestor would have had the trait shared by both species.
- ○ The common ancestor would have evolved the best trait for survival.
- ○ The common ancestor would have been more like the albatross because the fossil was not found on Antarctica.
- ○ The common ancestor would have been more like the penguin because penguins are able to survive in harsh climates.

Extended Response Item:

Open-ended items are human-scored questions that are worth up to 4 points. One open-ended question is presented for each of the three domains: life sciences, physical sciences, Earth & space sciences.



Part B

Describe what will likely happen if a strong wave hits the wood fence design.

Words : 0, Characters : 0

Part C

Identify the design that **best** meets Sid's criteria. Explain your reasoning.

Words : 0, Characters : 0