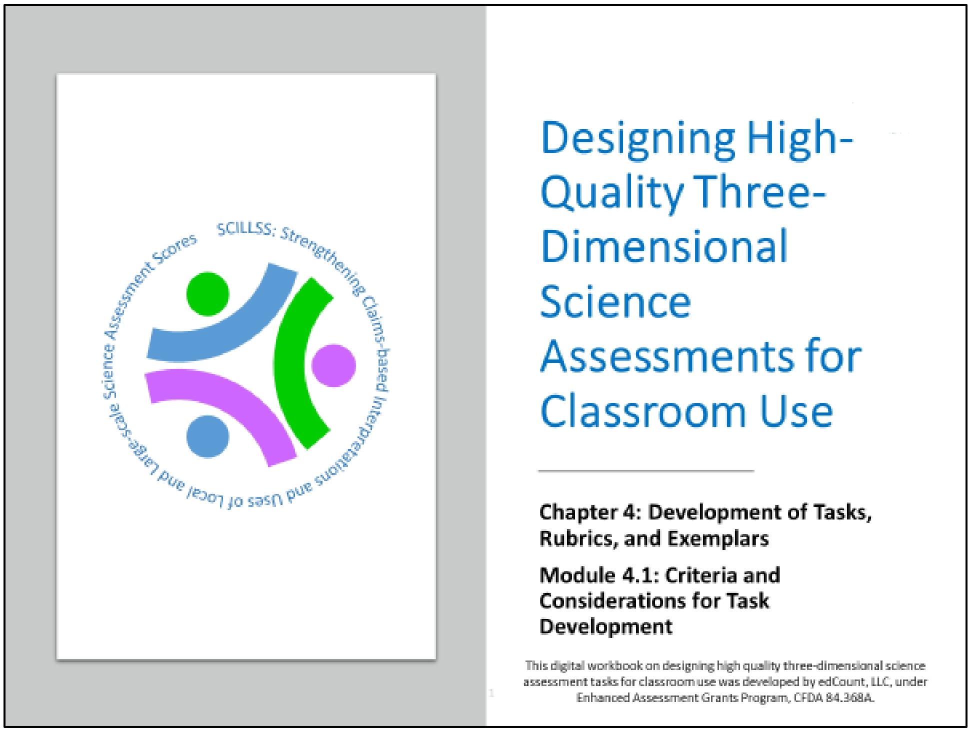
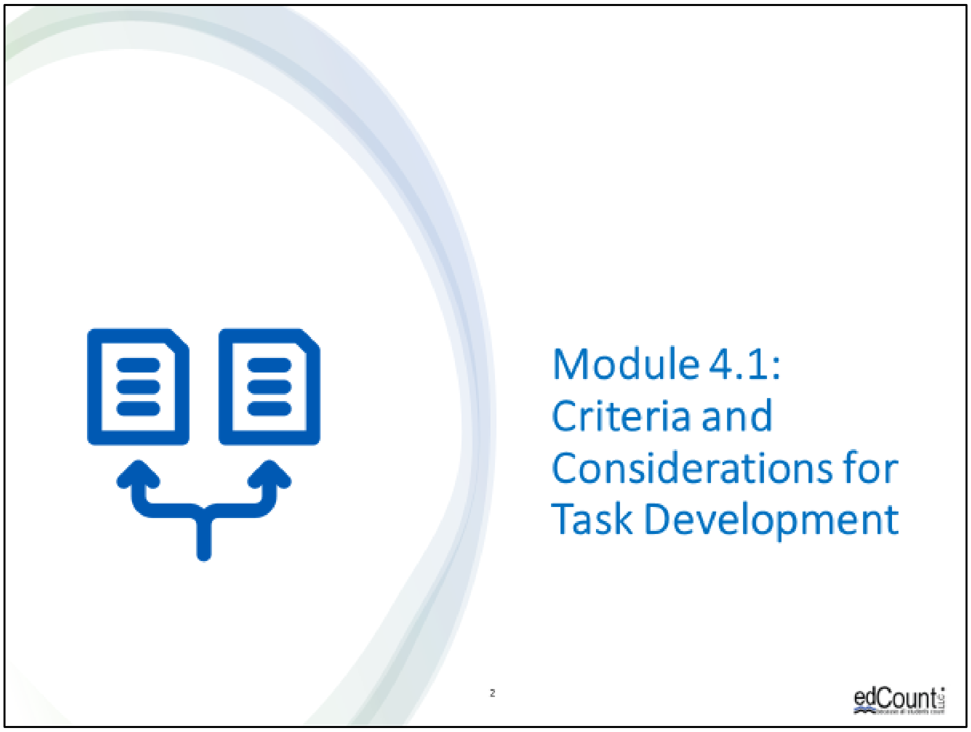


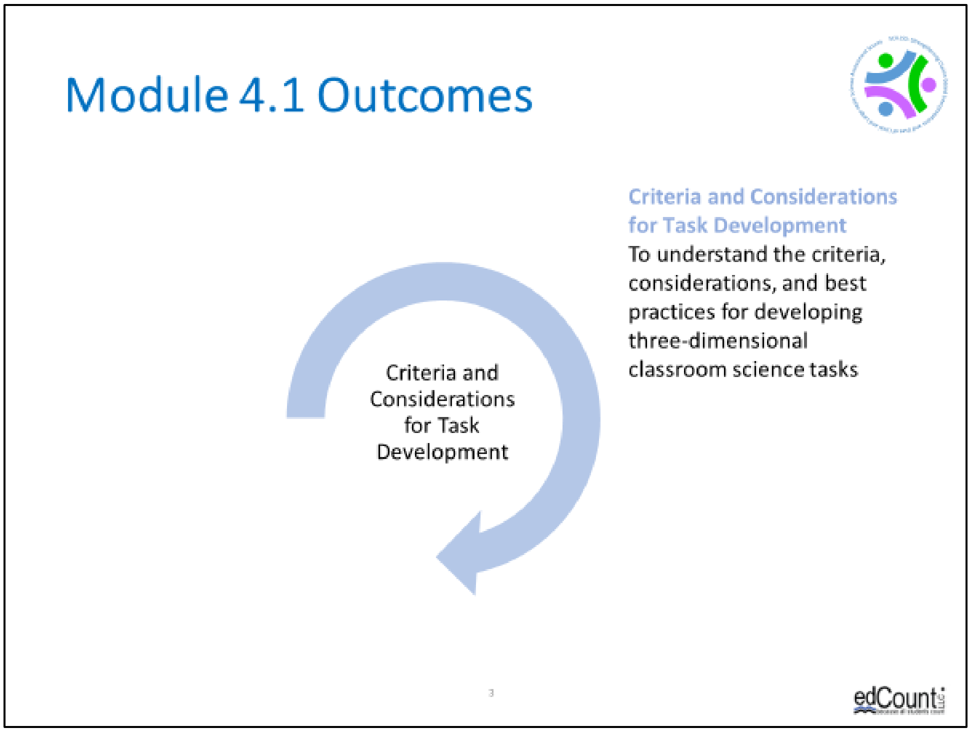
Welcome to the last of four chapters in a digital workbook on designing high-quality three-dimensional science assessment tasks for classroom use. This workbook is intended to help educators design and evaluate tasks that provide meaningful information about what students know and can do in science.

This digital workbook was developed by edCount, LLC, under the US Department of Education's Enhanced Assessment Grants Program, CFDA 84.368A.

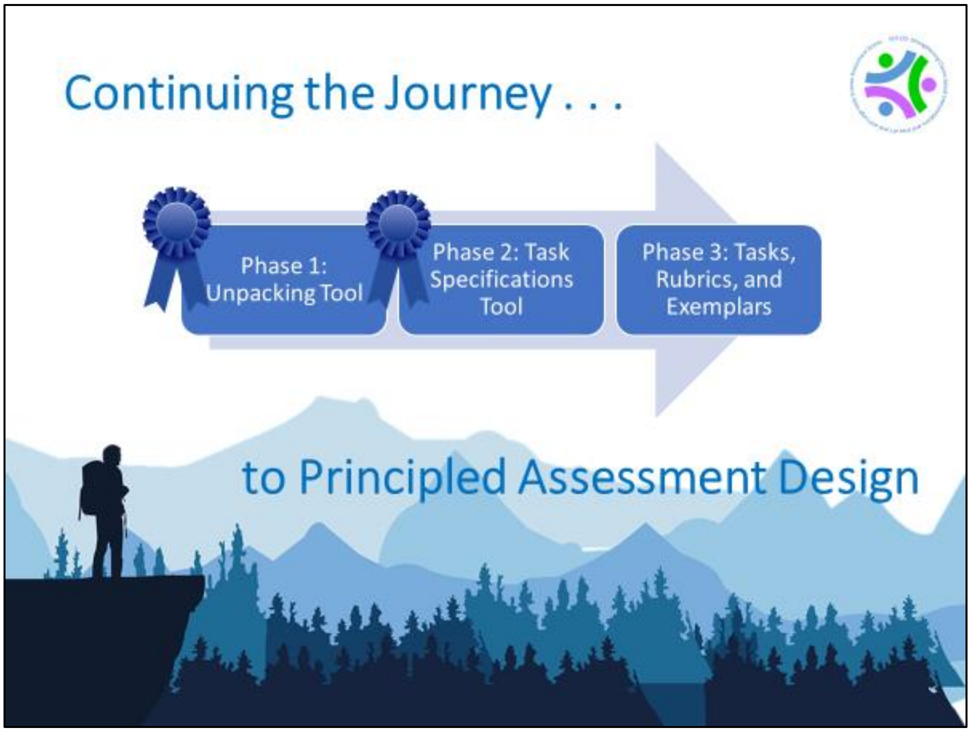


Chapter 4 of this workbook includes a series of six modules. Together these six modules provide an in-depth exploration of the third phase of principled assessment design: development of tasks, rubrics, and exemplars. In this chapter, we focus on translating the unpacking of the three dimensions of a specific performance expectation or indicator and the design elements in the task specifications tool into an assessment task and rubric. We provide opportunities for you to engage in interactive activities and explore and use our design template to complete your own task and rubric, and learn how to apply scoring guidelines for a three-dimensional standard.

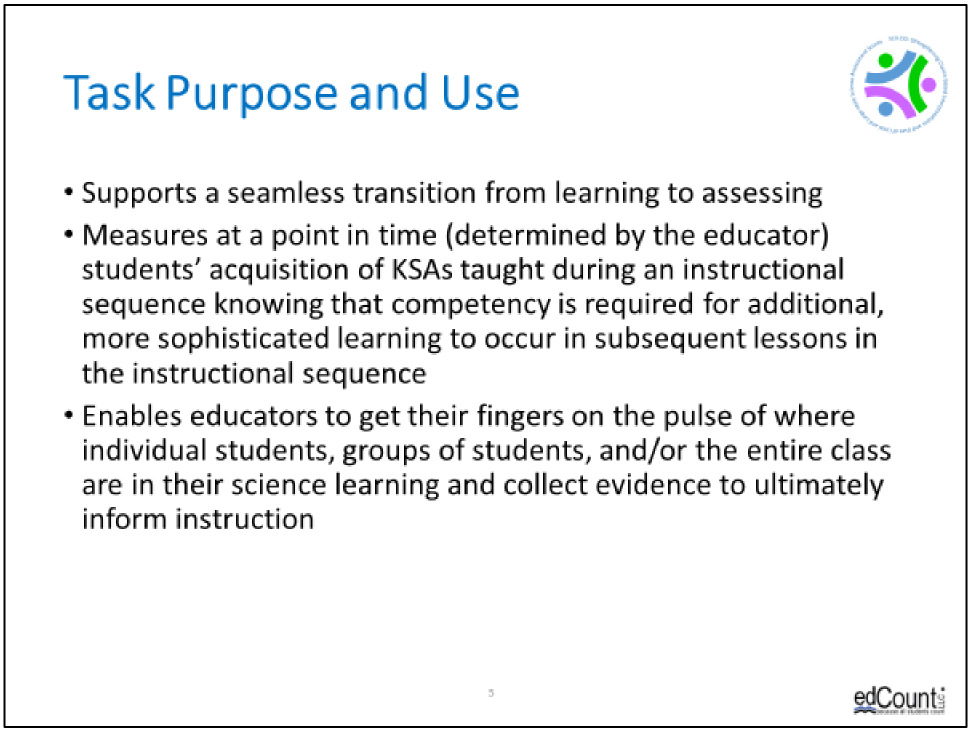
We begin this chapter with Module 4.1. In this module, we provide an overview of the purpose and use of three-dimensional science tasks and introduce criteria and considerations for developing high-quality tasks. In later modules, we offer model tasks and rubrics at the elementary, middle, and high school grade bands to illustrate the outcomes of the process as well as resources, key strategies, and guiding questions for designing high-quality three-dimensional classroom assessment tasks and rubrics.



In this module, you begin your journey into the third phase of principled assessment design. Our intent is to help you more deeply understand the characteristics of high-quality classroom science tasks and the criteria, considerations, and best practices that will guide you in the design of your own tasks. We present information, tools, resources, and guided activities to support your intentional use of the unpacking tool and task specifications tool to design tasks that generate interpretable, meaningful, and useful information.



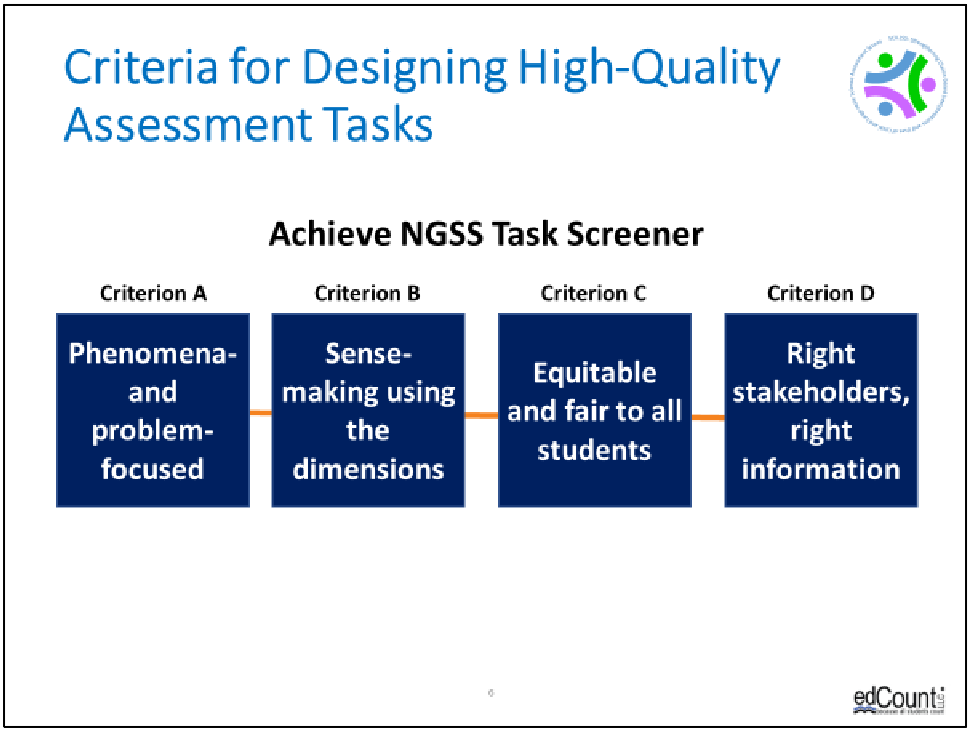
Congratulations on completing the first and second phases of principled assessment design: development of an unpacking tool and task specifications tool and choosing to continue your journey to design high-quality three-dimensional classroom science tasks. In this chapter, you will embark on the next phase of principled assessment design, development of the task and rubric, and identification of student exemplars through the application of the rubric. In Phase 3, we will draw upon the unpacking and task specifications tools you developed in phases 1 and 2 to construct a high-quality science assessment task.



Understanding your purpose for assessing, including what you are measuring and when, and how you will use the results, are foundational to principled assessment design. As an educator, you are likely well-accustomed to and adept at using both formative and summative assessment as part of your instructional toolkit. You understand the value of assessment and the wealth of information it provides for both students and educators. Therefore, **intentional** task design—the process of designing tasks by first understanding WHAT you want to understand about your students' science learning and WHEN you need to know it during the instructional sequence—is likely NOT a new or unfamiliar concept. To ensure that tasks elicit meaningful and useful evidence of student learning, task design and selection must begin with a consideration of your purpose for assessing and how you plan to use the results to inform instruction.

First and foremost, purposeful task design requires a clear definition of WHAT you intend to measure so that you can create or select a task that aligns with that intent and, thus, support a seamless transition from learning to assessing. Your task should neither under-represent the intended content or construct nor include content that is beyond or outside of what the task is intended to measure. If it does, the evidence from the task cannot be interpreted back to the intended PE or KSAs and cannot be used for purposes associated with those learning targets.

Purposeful task design also requires consideration of WHEN to assess. For example, you may choose to administer a task immediately following a lesson to ensure students have grasped an important foundational concept prior to advancing to more sophisticated learning within a series of subsequent lessons. Or perhaps you choose to administer a mid-unit or culminating assessment to determine if students are able to apply and transfer their learning of the dimensions into a new or novel context. By measuring student learning at a specific point in time during an instructional sequence, you can get your finger on the pulse of where individual students, groups of students, and/or the entire class is in their science learning and collect evidence to ultimately inform instruction.



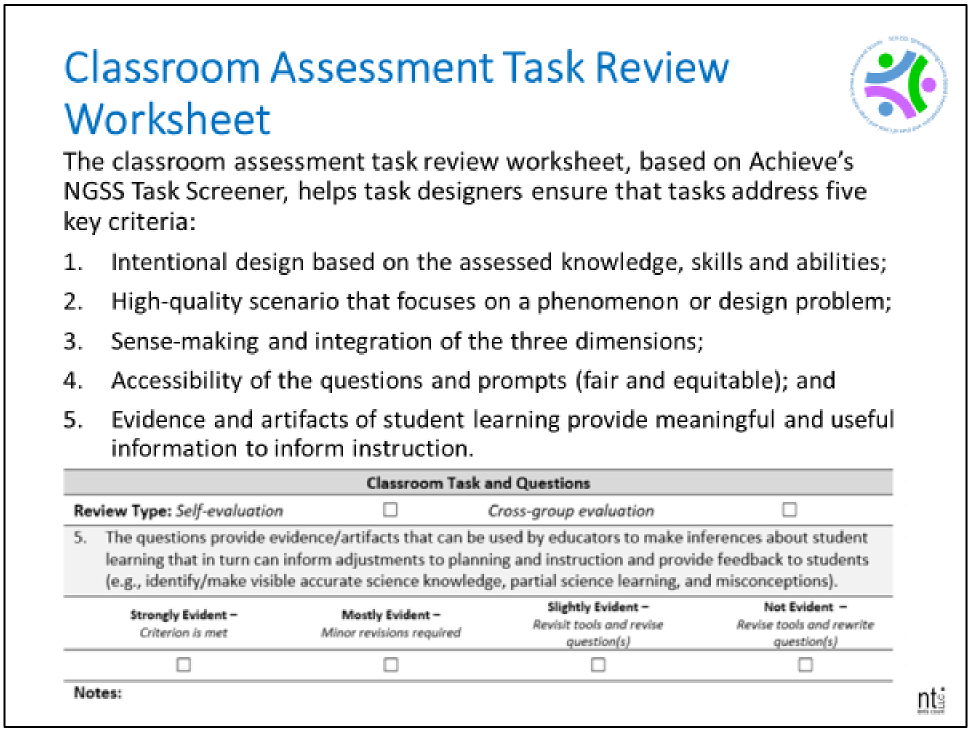
In this module, we explore important criteria and considerations for designing NGSS- and *Framework*-aligned tasks. As a task developer, it is important to understand the features and characteristics of high-quality tasks and to use these criteria to inform your design of classroom assessment tasks. First, let's explore four criteria defined in Achieve's NGSS Task Screener, which is provided for reference in the Web Links pod. Each criterion comes with a set of indicators to look for in a task.

The first criterion of the task screener emphasizes the importance of a high-quality scenario that focuses on a phenomenon or design problem. If we want students to show us that they can use their science ideas and practices to figure things out, we need to give them something to figure out. The scenario should be relevant to the task and should allow students to explain the phenomenon or design problem using the targeted dimensions and KSAs. It should be carefully crafted with students in mind. In other words, it should be puzzling and intriguing to students, compelling enough to motivate a wide range of students, and it should be authentic locally, globally, or universally. If students care about the outcome, they are more likely to be engaged and put forth effort to demonstrate their science learning.

The second criterion ensures that the task requires students to use reasoning to sense-make using the dimensions. As students figure out a phenomenon or solve a phenomenon-based problem, they must be drawing on multiple dimensions. Application of the DCI without a SEP or CCC is rote recall. Trying to use a SEP or CCC without the other dimensions results in losing the context that makes science science. The task should provide students opportunities to make their thinking visible as they integrate multiple dimensions, demonstrate their learning of the KSAs, and use reasoning to make sense of the phenomenon or design problem.

Tasks must also be equitable and fair. The third criterion requires that students have access to tasks, including the use of multiple points of entry and multiple types of supports, as well as the purposeful engagement of student agency and interest, making science relevant to them and their lives. The task should provide ways for students to make connections of meaningful local, global, or universal relevance, offer multiple modes for students to respond, and present scientifically-accurate content that is accessible, appropriate, and cognitively demanding for all learners, builds upon student interest and confidence in science and reflects students' opportunity to learn.

The fourth criterion emphasizes that all tasks must have an intended purpose or use. A task doesn't exist in isolation but rather is meant to produce evidence about student learning so that someone can do something with that evidence. Tasks must be designed with a purpose in mind and clearly connect evidence to use. The task should assess what it is intended to assess, and support the purpose for which it is intended, and should elicit student artifacts that provide evidence of how well students can use the targeted dimensions together to reason about a phenomenon or design problem. Criterion four also ensures that the task includes supporting materials, such as answer keys, rubrics, and scoring guidelines, that are clear and connected to the targeted KSAs and includes prompts and directions that provide sufficient guidance for administration for both teachers and students.



The Classroom Assessment Task Review Worksheet is a task evaluation tool based on Achieve's NGSS Task Screener. It is intended to help task developers design and evaluate classroom assessment tasks to ensure the tasks are designed for the purpose for which they will be used and contain questions that reflect an intentional design based on the assessed KSAs. The tool also helps task developers ensure that the four criteria within Achieve's NGSS Task Screener are addressed. High-quality tasks should be phenomena or problem-focused, elicit evidence of three-dimensional performances and sense-making, promote accessibility for ALL students, and provide clear and sufficient information to elicit meaningful and useful evidence to inform instruction. The task review worksheet is designed for flexible use. For example, an educator or a small group of educators might complete a self-evaluation of a task, or groups of educators might complete a cross-group review of a task developed by their peers. The task review worksheet is available for download in the Resources pod.

Let's take a brief look at the tool and its five task criteria. In later sections in this module, we will explore each criterion in greater depth to support an accurate interpretation and effective use of this tool.

First, the task should reflect an intentional design based on the assessed knowledge, skills, and abilities. When you use your unpacking tool and task specifications tool to design a task, you should observe direct alignment between the task and various aspects of the unpacking and task specifications tools. The questions within the task will vary in their degrees of complexity, types of demonstrations of student learning, and types of work products, but all features of the task will align back to the palette of design options included in your design tools.

The task should also include a high-quality scenario that focuses on a phenomenon or design problem and includes information that is necessary and adequate for students to successfully respond to the task. As mentioned previously, the scenario should be carefully crafted with students in mind and should allow students to explain the phenomenon or design problem using the targeted dimensions within the KSAs.

The questions within the task should provide opportunities for students to apply reasoning and sense-making in their integration of multiple dimensions within the PE. Traditional science assessments that measure discrete facts and ideas and focus on rote recall are a thing of the past. *Framework*-inspired assessments reflect a new vision for science education in which students actively engage in scientific and engineering practices and apply crosscutting concepts to deepen their understanding of the core ideas and engage with fundamental questions about the world and with how scientists and engineers have investigated and found answers to those questions (the *Framework*, pp. 8-9).

Questions and prompts within a task must also be fair and accessible to all students, including students with disabilities and English learners. Tasks should offer multiple entry points, supports, and modes for students to respond and should present information that is scientifically accurate and stated clearly, concisely, and at the appropriate readability level.

Finally, questions and prompts should elicit evidence and artifacts of student learning that can be used by educators to make inferences about student learning that, in turn, can inform adjustments to planning and instruction and provide meaningful feedback to educators and students.

Criterion 1 of the Classroom Assessment Task Review Worksheet addresses the notion of intentional design. In this section, we will provide additional information to support the design of your tasks and application of the task review worksheet.

Intentional Design based on the Assessed Knowledge, Skills, and Abilities

- Provide an indication of students' current understanding of the selected KSAs as set forth in the Task Specifications Tool
- In turn, must be designed using the created design tools (i.e., Unpacking Tool and Task Specifications Tool) to provide evidence of students' ability to use practices, to apply their understanding of the crosscutting concepts, and draw on their understanding of specific disciplinary ideas, all in the context of addressing specific phenomena or design problems

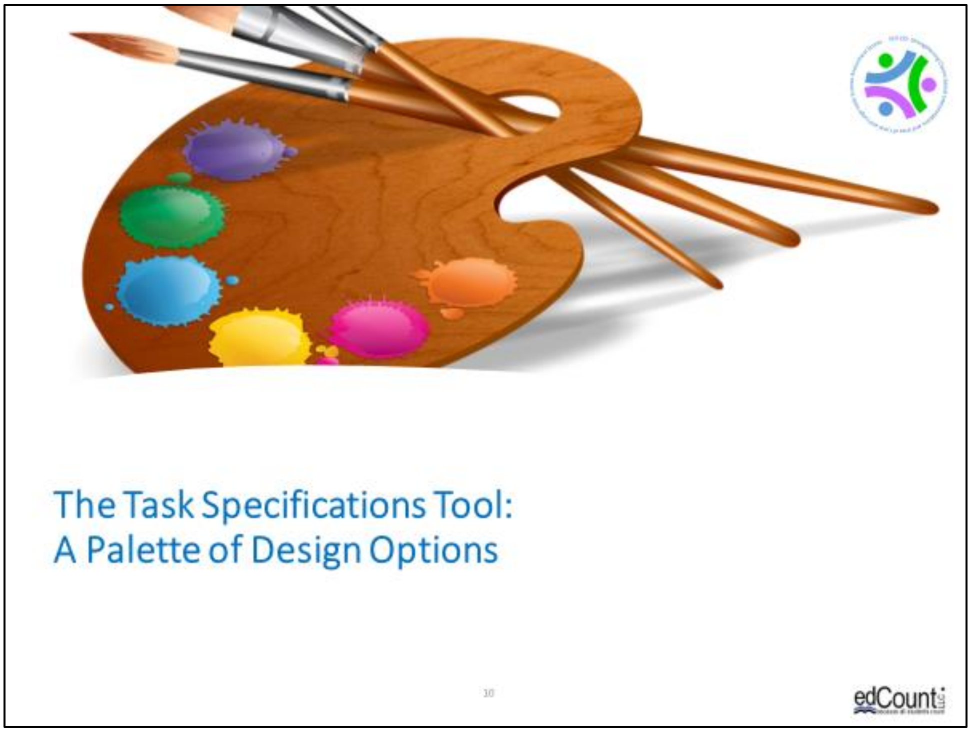
9

edCount

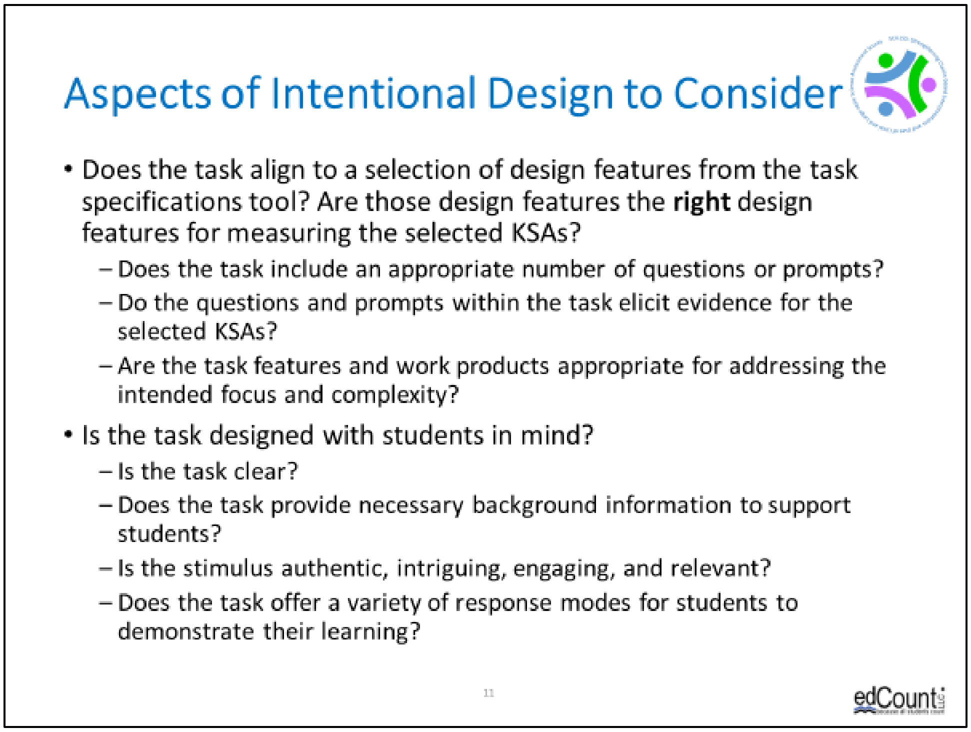
A cornerstone of principled assessment design is the notion that assessments are developed with the end goals for students in mind. Understanding the knowledge, skills, and abilities you aim to assess, and when during the instructional sequence you aim to assess them, is a critical first step to ensure the tasks you develop provide meaningful and useful information to address your students' learning needs.

The task specifications tool provides a wealth of information gleaned from the unpacking that you can use to create tasks. The tasks can range from a few interrelated questions with supports to a project-based task that requires using multiple skills, data, graphs, and sources of information or patterns.

That's the beauty of this tool; you can go from small to large, from shallow to deep, but it's all intentional. You can justify the evidence, gather the evidence, make inferences about what students know and can do, and make judgments about your own instruction and how you can adjust it to improve learning.



Just as an artist uses a paint palette to select from a range of colors to include in his or her artwork, a task developer has a palette of design options within the task specifications tool to create a classroom assessment task. Considering the elements of the task specifications tool and selecting the appropriate design features that match the intended purpose and use of the task is an essential part of intentional task design.



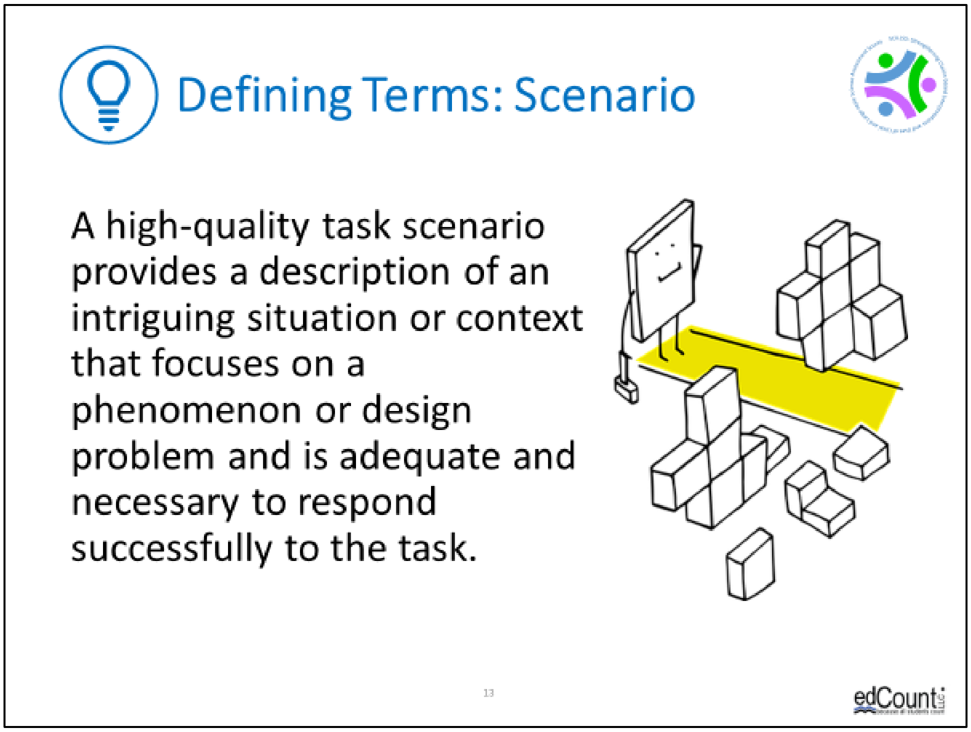
As you design or evaluate your classroom assessment task, consider if the design features of the task are reflected within the task specifications tool, and of equal importance, whether those design features are the right design features for measuring the selected KSAs.

Begin by considering the KSA or KSAs selected for measurement. Does the task include an appropriate number of questions or prompts to measure the KSAs? Do the questions or prompts elicit the **right** evidence to support accurate interpretations about what students know and can do in relation to the KSAs? Are the task features and work products appropriate for addressing the intended focus and complexity of the KSAs?

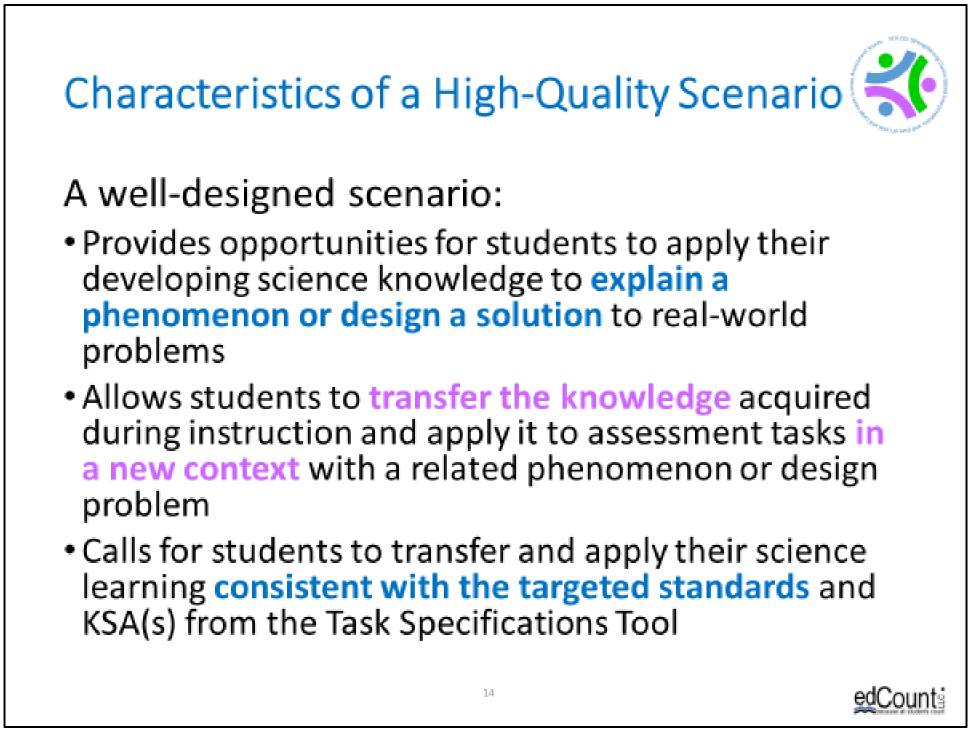
Another key aspect of intentional design is a consideration of the students who will engage with the task. Is the task clear? Does the task provide the necessary background information to support students? Is the scenario authentic, intriguing, engaging, and relevant? Does the task offer a variety of response modes for students to demonstrate their learning?

These considerations provide a glimpse into the thought process one must have when designing a task with intentionality in mind. A key indicator of intentional design is both ensuring alignment of the task to the unpacking and task specifications tools **and** ensuring that the **right** design features from these tools are reflected in the task based on the selected KSAs for measurement.

Criterion 2 of the Classroom Assessment Task Review Worksheet addresses the development of a high-quality task scenario. In this section, we will provide additional information to support the design of your tasks and application of the task review worksheet.



A scenario is an important part of an assessment task, but how do we define a scenario? A high-quality task scenario provides a description of an intriguing situation or context that focuses on a phenomenon or design problem and is adequate and necessary to respond successfully to the task.

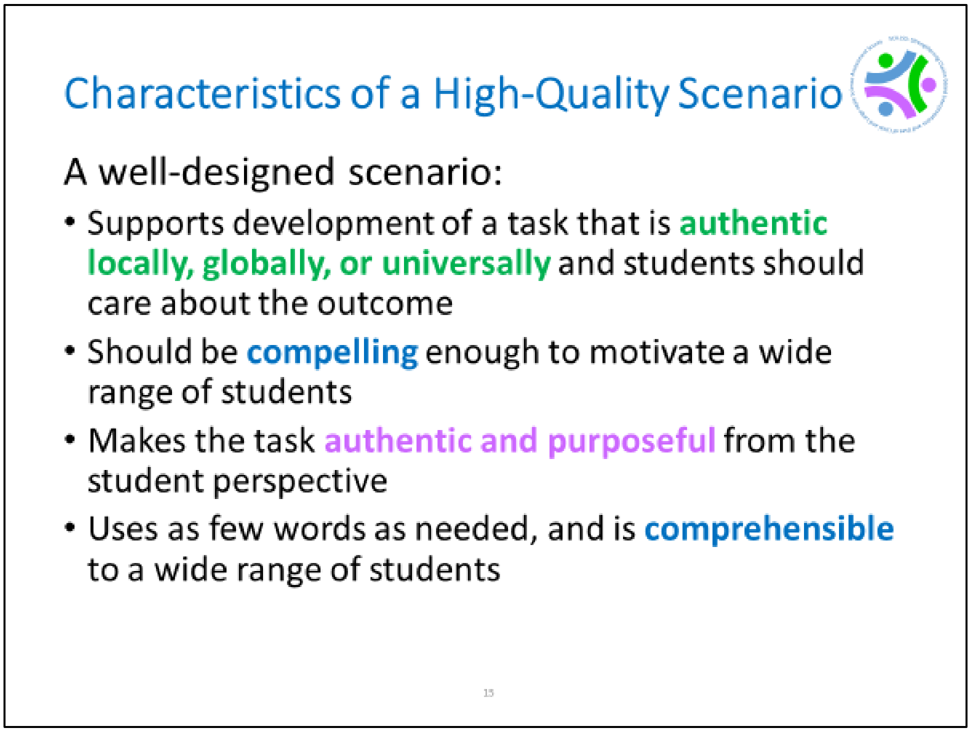


Let's consider the characteristics of a well-designed scenario.

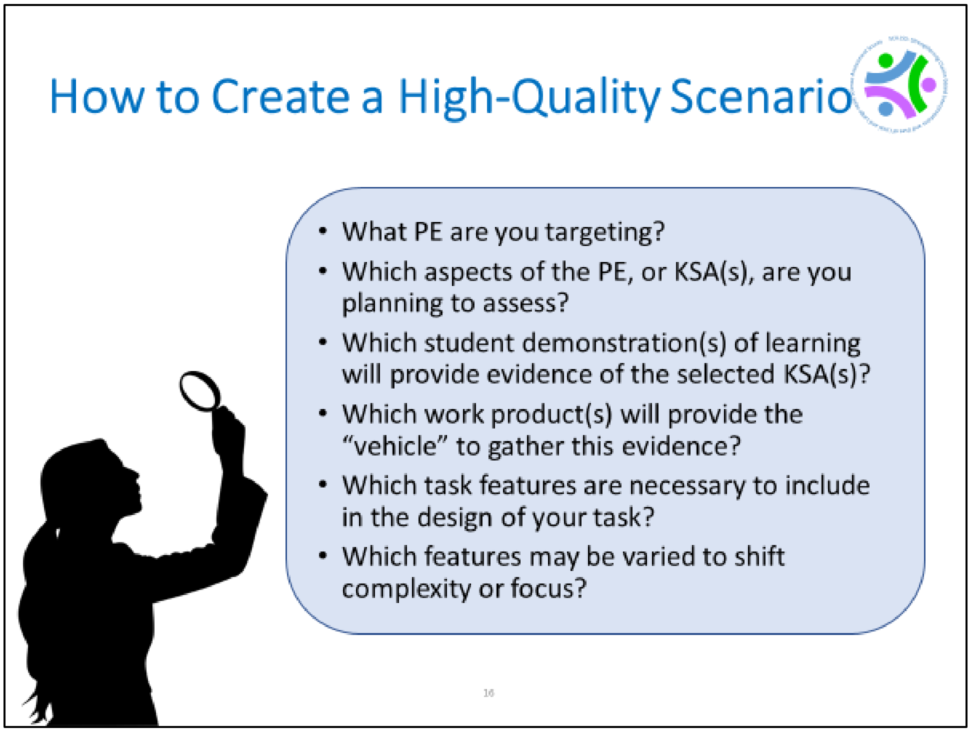
A task scenario provides opportunities for students to apply their developing science knowledge to make sense of a phenomenon or to address a design problem that is necessary to accomplish the task.

A scenario is sufficiently rich to drive the task and allows students to transfer the knowledge acquired during instruction and apply it to assessment tasks in a new context. A scenario supports students' use of the targeted grade-appropriate SEPs, CCCs, and DCIs and selected KSAs to explain the phenomenon or design problem.

According to Achieve's NGSS Task Screener, a scenario also effectively uses at least two modalities, such as images, diagrams, video, simulations, or textual descriptions, and if data are used, presents real and scientifically-accurate data. Also, a scenario should be based around at least one specific instance, not a topic or generally observed occurrence. For example, a scenario might focus on observations of a specific hurricane rather than "hurricanes" in general.



A scenario is also carefully crafted to support the development of a task that is authentic locally, globally, or universally and students should care about the outcome. For accessibility purposes, the scenario should be compelling to a wide range of students and authentic and purposeful from the student perspective. In addition, it should be stated clearly and concisely, using as few words as necessary, so that it is comprehensible to all students.



Let's assume that you recently completed work on your unpacking tool and task specifications tool for a selected PE. You have considered the purpose and use for assessing your students, and you are ready to begin designing your task. How might you create a high-quality scenario? What would you need to consider, or think about, to craft the right scenario?

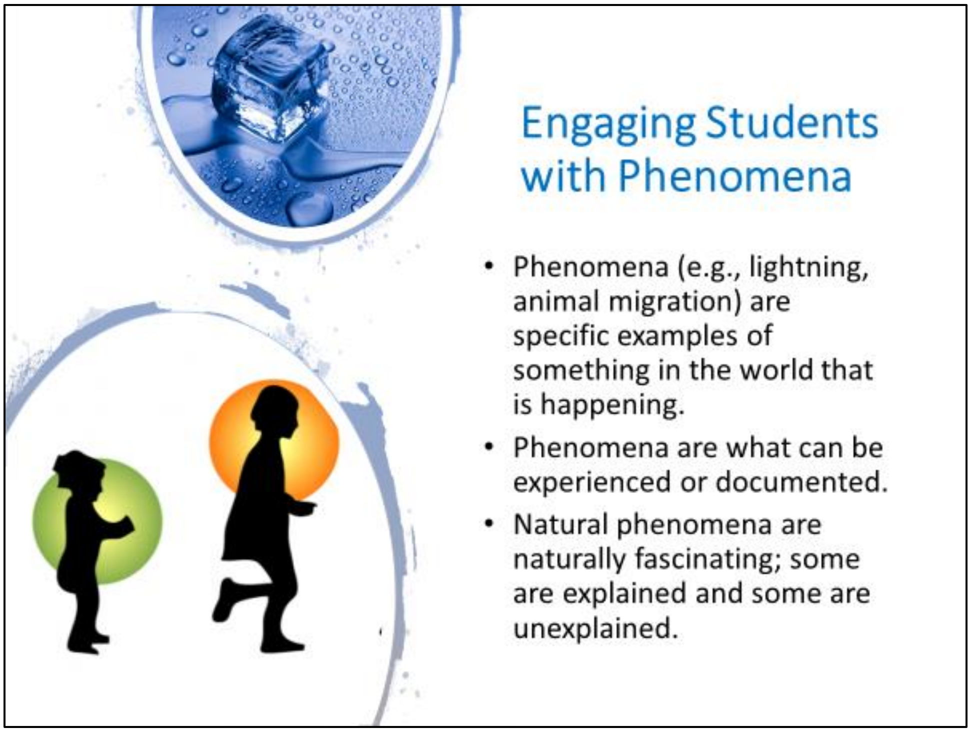
First, you must know what you want to measure and why. Carefully consider the elements of your unpacking tool and task specifications tool to support the intentional design of your task. Ask yourself:

- What PE am I targeting?;
- Which aspects of the PE, or KSA(s), am I planning to assess?;
- Which student demonstration(s) of learning will provide evidence of the selected KSA(s)?;
- Which work product(s) will provide the "vehicle" to gather this evidence?;
- Which task features are necessary to include in the design of my task?; and
- Which features may be varied to shift complexity or focus?

As you consider these questions and review the elements of your unpacking and task specifications tools, begin to identify the combination of design features that will support the measurement of the selected KSAs.

You may also find that as you review your task specifications tool, you notice inconsistencies, holes, or gaps in the information as defined in the tool. You may struggle to identify design features that would allow you to assess your selected KSA or KSAs. Remember, principled assessment design is an iterative process that requires continual review and refinement of your unpacking tool, task specifications tool, and tasks and rubrics. You may need to make adjustments to your design tools as you work to ensure alignment between your task and the full range of design choices that meet your purpose and use for assessing.

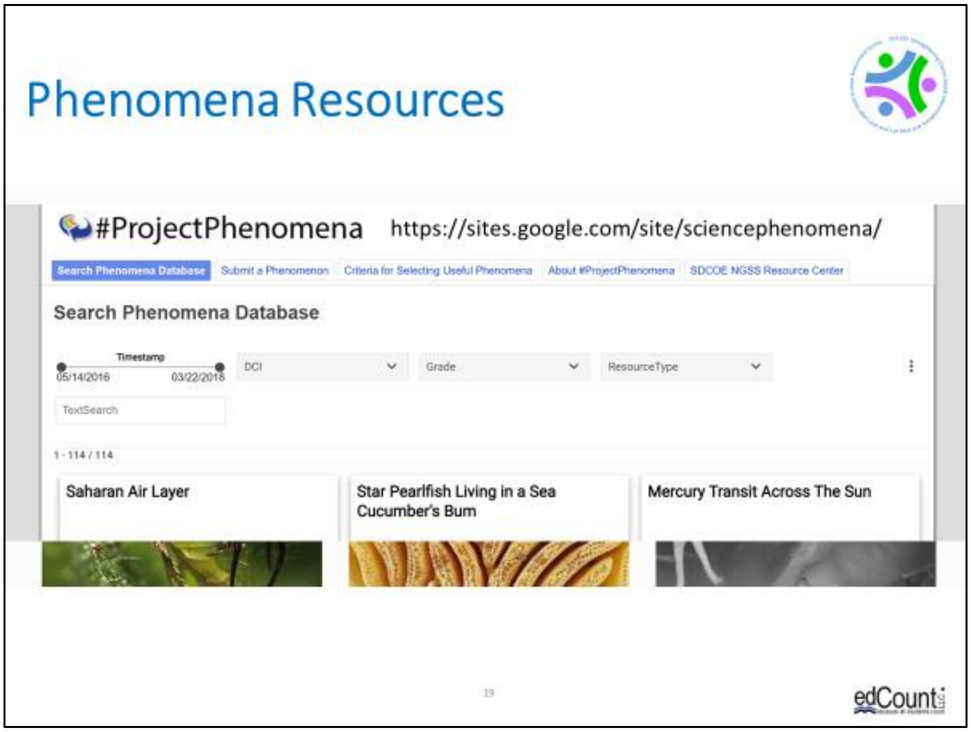
Now that you have carefully considered your design tools and your purpose and use for assessing, you'll need to select a phenomenon or design problem that will engage students in a meaningful way to demonstrate their learning of the selected KSAs. Remember, the phenomenon or design problem that you choose should support a scenario that is purposeful, authentic, meaningful, relevant, real-world, intriguing, puzzling, and motivating. Most importantly, the phenomenon or design problem that you choose must be appropriate for measuring students' science learning of the selected KSAs. A misaligned phenomenon or design problem would make it challenging to create questions or prompts that elicit the evidence you're looking for in students' responses.



Phenomena are specific examples of something in the world that is happening—an observable event that occurs in the universe and that students can use science knowledge to explain or predict. Phenomena are NOT the explanations or scientific terminology behind what is happening. They are observable. They are what can be experienced or documented.

Examples of natural phenomena, some explained and some yet to be explained, are all around us and your students all through life. They include a baby growing up and becoming an adult or ice melting on a hot summer day. They relate to concepts such as weather (e.g., fog, thunder, or tornadoes), biological processes (e.g., decomposition or germination), physical processes (e.g., wave propagation, erosion, or tidal flow), and natural disasters (e.g., electromagnetic pulses, volcanic eruptions, landslides, and earthquakes). And the list goes on.

Phenomena need not be flashy or jazzy. The key is that they are engaging and tap students' natural curiosity, encourage them to wonder, and ultimately, through their acquired habits of mind as a scientist or engineer, invite them to investigate, explore, test, and discover.

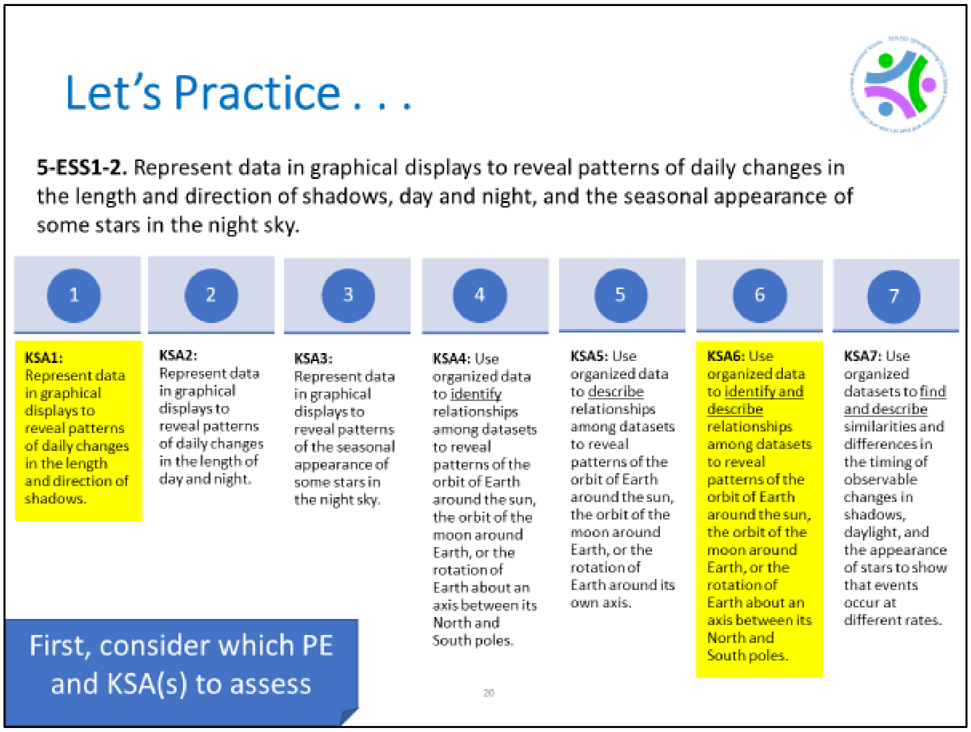


Several curated collections and databases of phenomena exist and are available for your use as a task developer.

The Phenomena for NGSS is a website that houses a collection of phenomena that you can search for by topic or disciplinary core idea. Content on the site was developed and/or curated by Chris Zieminski and TJ McKenna and is available for use under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.

The Wonder of Science offers a master list of phenomena organized in an open Google document by grade, course, and PE. The phenomena are in the process of being tagged and added to the website over time with relevant links, videos, and images. Similar to the previous website, this content is available for use under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.

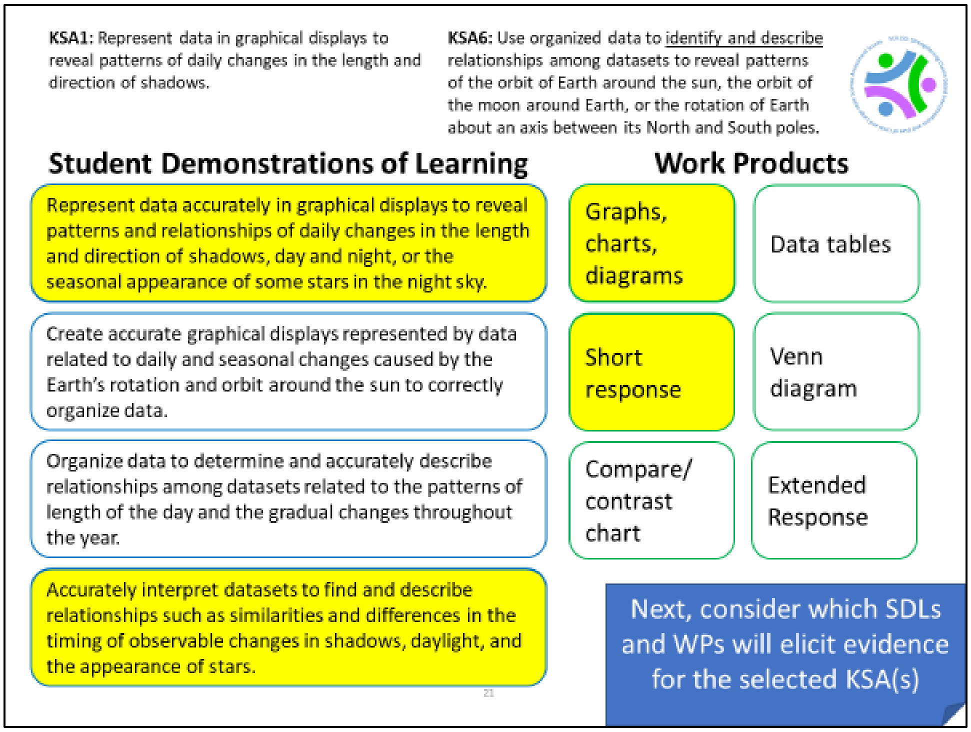
#Project Phenomena is a website database that houses phenomena organized by DCI, grade, and resource type. The website represents a collaboration of teacher, industry, university, and community organization leaders who want to help students engage in relevant, engaging, and meaningful phenomena. The website offers links to third-party websites and may contain copyrighted information.



Let's practice brainstorming a task scenario for the NGSS PE *5-ESS1-2: Represent data in graphical displays to reveal patterns of daily changes in the length and direction of shadows, day and night, and the seasonal appearance of some stars in the night sky.*

We'll begin with the task specifications tool. Let's presume that we just finished a series of lessons in which students conducted observations of daily changes in the length and direction of shadows and organized and represented their data in graphical displays. We want to measure how well students can use an organized data set to identify and describe patterns in the appearance of shadows based on the position of the sun in the sky at various times of day.

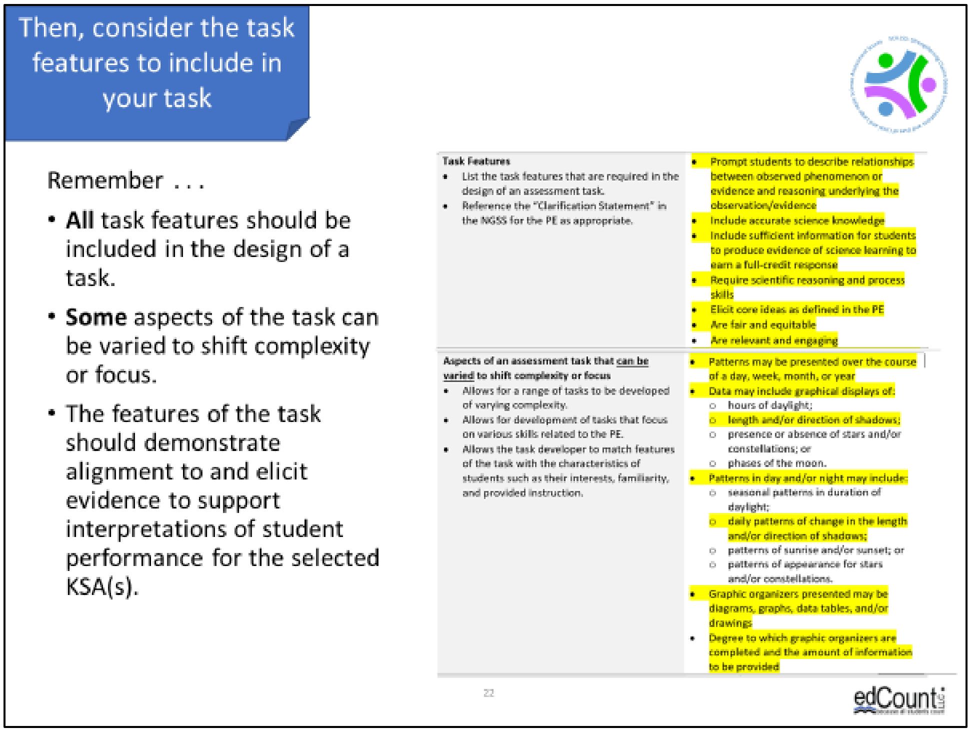
This particular task specifications tool identifies seven KSAs for measurement. Remember, a task can measure one or multiple KSAs. Based on the purpose for assessment we just identified, which KSA or KSAs might be appropriate as the focus of our task? Let's select KSA1 and KSA6 as they closely align to what we want to measure in terms of students' acquired knowledge, skills, and abilities. KSA1 requires students to *"Represent data in graphical displays to reveal patterns of daily changes in the length and direction of shadows."* KSA6 states that students should *"Use organized data to identify and describe relationships among datasets to reveal patterns of the orbit of Earth around the sun, the orbit of the moon around Earth, or the rotation of Earth about an axis between its North and South poles."* For the purposes of this task, we will focus on patterns that emerge as a result of Earth's rotation about an axis between its North and South poles.



Now that we have identified KSA1 and KSA6 for measurement, let's consider the student demonstrations of learning and work products that would elicit and represent the evidence we need to make accurate interpretations of students' science learning.

First, let's consider these four student demonstrations of learning. Which SDLs would represent and qualify the types of performances we might expect to see from students for these KSAs? That's right. We might expect students to be able to represent data accurately in graphical displays to reveal patterns and relationships of daily changes in the length and direction of shadows. We would also expect students to accurately interpret datasets to find and describe relationships in the observable changes in shadows. These two SDLs represent the evidence we are planning to elicit from the classroom assessment task and the student expectations we would describe within our rubric.

Next, let's think about the types of work products we might include in our task. Remember, the work products should be the **right** work products. They need to align with the SDLs we just identified and provide appropriate "vehicles" for students to demonstrate evidence that they have learned and can apply the selected KSAs. Since we want students to identify and describe patterns in the appearance of shadows due to the rotation of Earth on its axis, it might be appropriate to have multiple items of varying complexities. The first item might require students to identify patterns from a provided data set in a partially completed graph, chart, or diagram. A second, more complex item might require students to write a short response to describe or explain the graph, chart, or diagram using the observed relationship between the length of the shadow and the apparent motion of the sun as recorded at different times of day.



As we move through the task specifications tool, we begin to narrow our design choices based on the KSA or KSAs we've selected for measurement. Now that we have identified our SDLs and WPs to include within our task, let's consider the task features and aspects of the task that can be varied to shift complexity or focus. It is important to emphasize that the features of a task must demonstrate alignment to and elicit evidence to support interpretations of student performance for the selected KSA or KSAs. In this case, we consider those features appropriate for measuring KSA1 and KSA6.

Let's start by reviewing our task features. Recall that all task features should be included in the design of our task. According to the task specifications tool, the task should: prompt students to describe relationships between observed phenomenon or evidence and reasoning underlying the observation, include accurate science knowledge, include sufficient information for students to produce evidence of science learning to earn a full-credit response, require scientific reasoning and process skills, elicit core ideas as defined in the PE, and be fair, equitable, relevant, and engaging to all students.

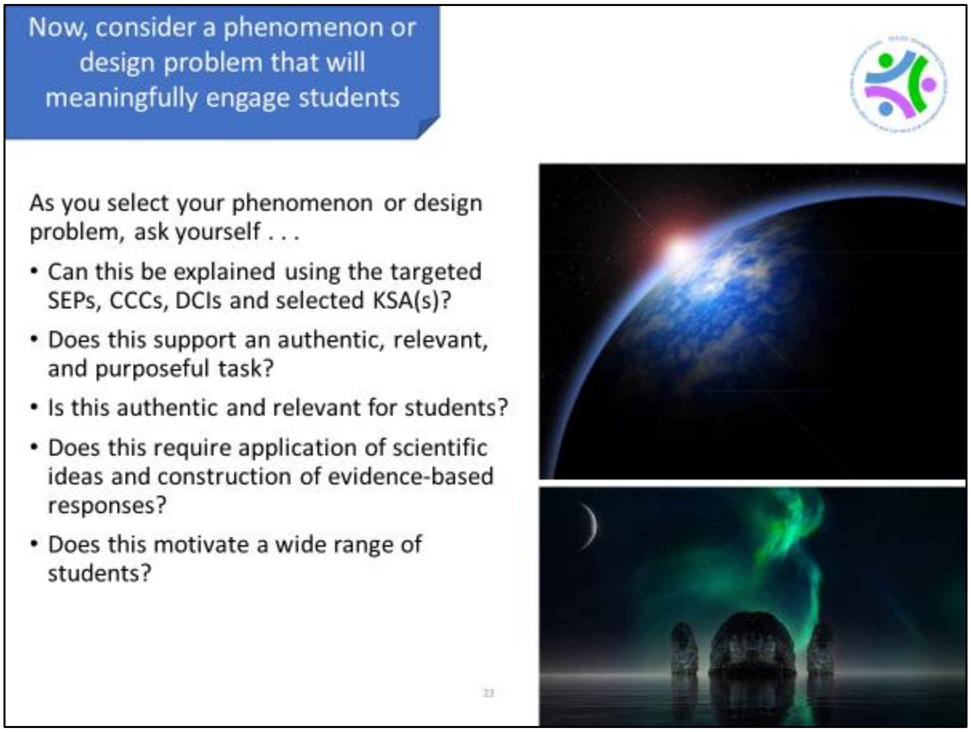
Now let's consider those aspects of the task that might be varied to shift complexity or focus. The first bullet, *"Patterns may be presented over the course of a day, week, month, or year"* is an important consideration as we select the phenomenon for our task. Seasonal patterns would require at least a year's worth of data in comparison to shadow patterns that require a day's worth of data. Since the particular focus of our task requires students to identify and describe the observable patterns of shadows, we'll likely want to select a phenomenon that presents patterns over the course of a day.

We would also expect our task to present data that includes graphical displays of length and/or direction of shadows and include daily patterns of change in the length and/or direction of shadows.

Our task may utilize a diagram in which students illustrate the changes in the patterns of shadows over the course of a day, or it might represent the data in a table or chart. For example, we might ask students to compare data that presents the length and timing of shadows in two locations. In this instance, a table of length in standard units might be simpler and clearer to interpret than drawings.

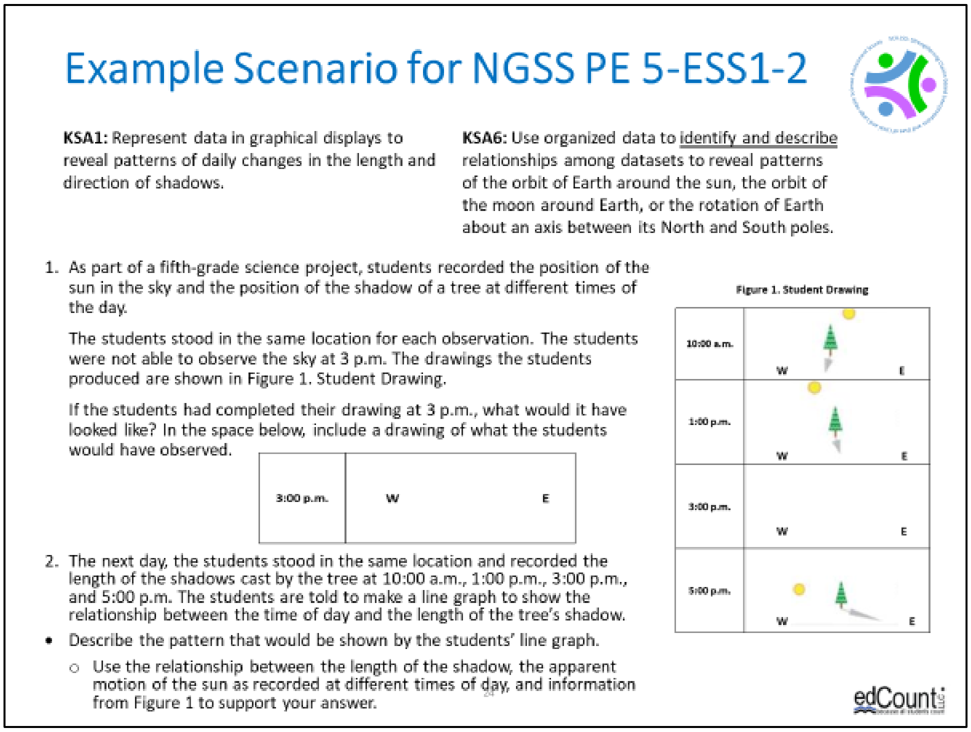
Finally, to vary the complexity of the task, we can vary the degree to which graphic organizers are completed or the amount of data we provide to students. We might ask students to create and label their own diagram of the positions of shadows over the course of a day, or we may ask students to finish a partially completed diagram with an incomplete data set.

As we consider these task features, we need to think not only about the KSA or KSAs we are trying to measure but also about our students who will be engaging with the task. What prior instruction have they received? How familiar are they in demonstrating their learning using the various work products? Do the ways in which students demonstrate their learning in the classroom align to the ways in which you are eliciting evidence in the task? Students' opportunities to learn in the classroom should align to what they might experience in an assessment situation.



We are almost ready to begin crafting our task scenario, but first, we need to select a phenomenon or design problem as the basis for our scenario. Let's consider the phenomenon that the position and length of shadows change as the position of the sun changes in the sky. Can this phenomenon be explained using the targeted KSAs and dimensions of the PE? Does it support an authentic, relevant, and purposeful task? Is it authentic and relevant to students? Does it require the application of scientific ideas and construction of evidence-based responses? Does it motivate a wide range of students? If the answer to these questions is yes, then we are ready to begin crafting our task scenario!

Remember, a phenomenon must be appropriate for measuring the targeted KSAs and dimensions, but it also must be appropriate for students who are engaging with the task.

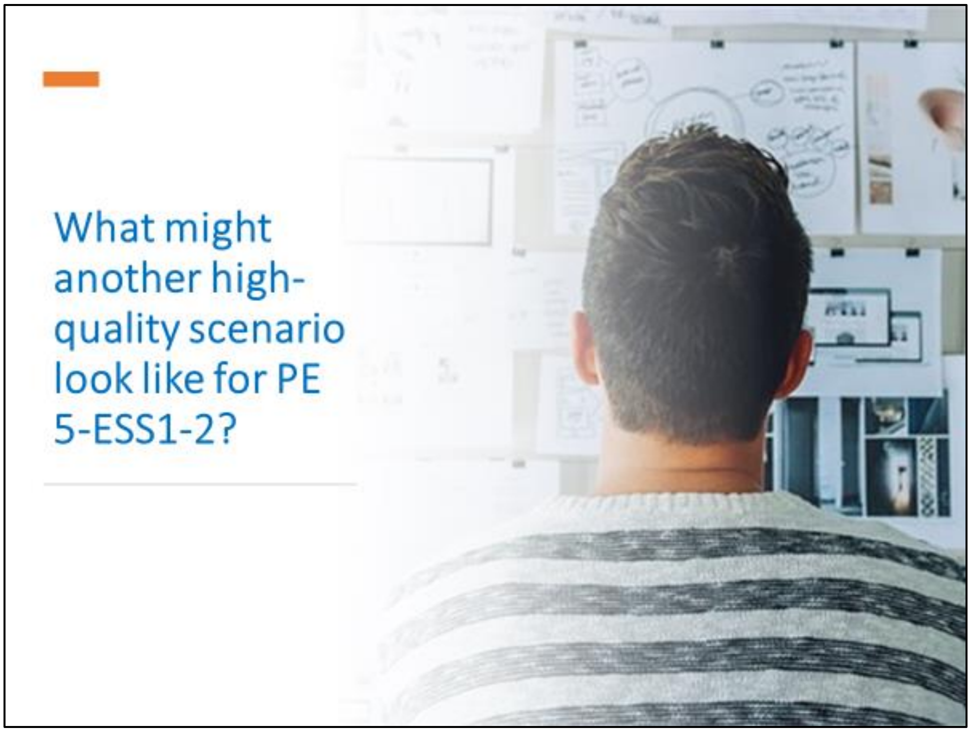


Here is a sample scenario for the task specifications tool for *5-ESS1-2. Represent data in graphical displays to reveal patterns of daily changes in the length and direction of shadows, day and night, and the seasonal appearance of some stars in the night sky.*

Pause the presentation to review the scenario. Consider how well the scenario aligns to KSA1 and KSA6 and the aspects of the task specifications tool that we identified previously. What does this scenario do well? Are there aspects of the scenario that might be improved?

Remember, a well-crafted scenario:

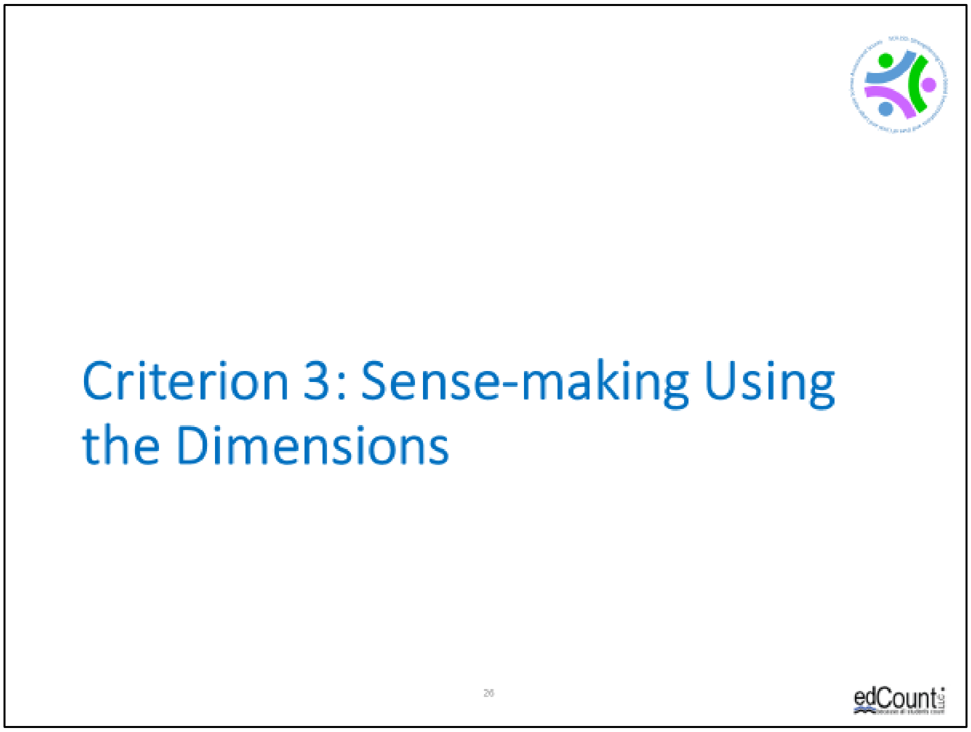
- Provides opportunities for students to apply their developing science knowledge to explain a phenomenon or design a solution to a real-world problem;
- Allows students to transfer the knowledge acquired during instruction and apply it to assessment tasks in a new context with a related phenomenon or design problem;
- Calls for students to transfer and apply their science learning consistent with the targeted standards and KSA(s) from the task specifications tool;
- Supports development of a task that is **authentic locally, globally, or universally** and students should care about the outcome;
- Should be **compelling** enough to motivate a wide range of students;
- Makes the task **authentic and purposeful** from the student perspective; and
- Uses as many words as needed, and no more, and is **comprehensible** to a wide range of students.



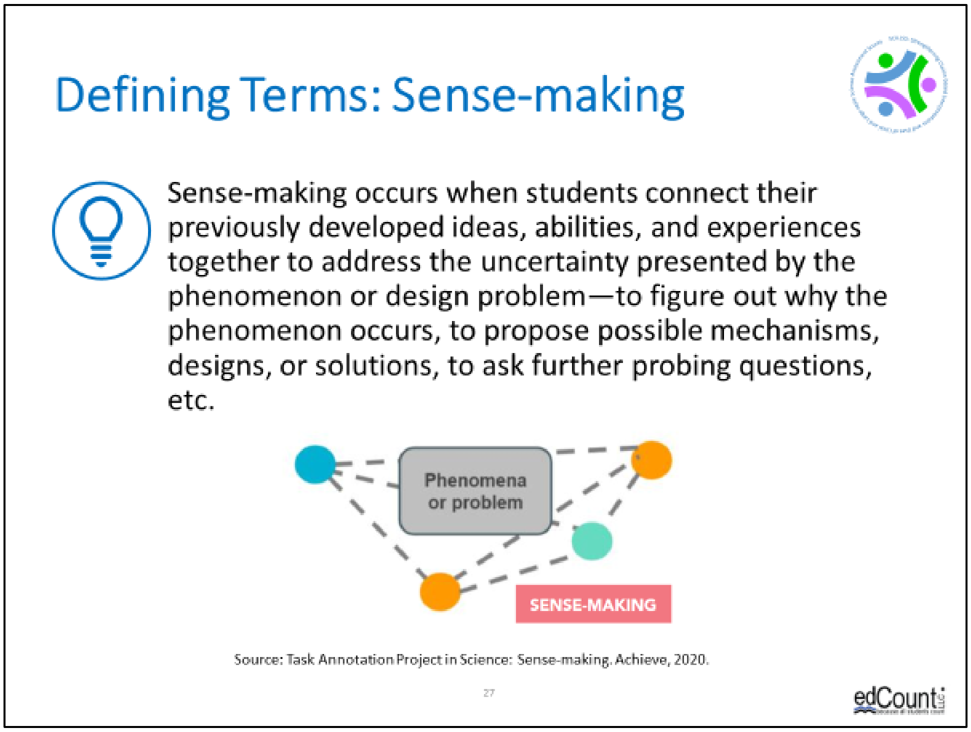
A task specifications tool provides a palette of design options that a task developer might choose from to design a range of tasks that vary in focus and complexity for a targeted PE. What might other tasks look like for this grade 5 PE?

Whether you are reviewing this module individually or with colleagues, we invite you to download the task specifications tool for 5-ESS1-2 from the Resources pod. As you review the tool, repeat the steps that we just completed. First, consider which KSA or KSAs to assess. Then, identify which student demonstrations of learning and work products would elicit evidence for the selected KSA(s). Next, think about the task features you would include in your task. Finally, select a phenomenon or design problem that would meaningfully engage students to demonstrate their learning of the selected KSAs.

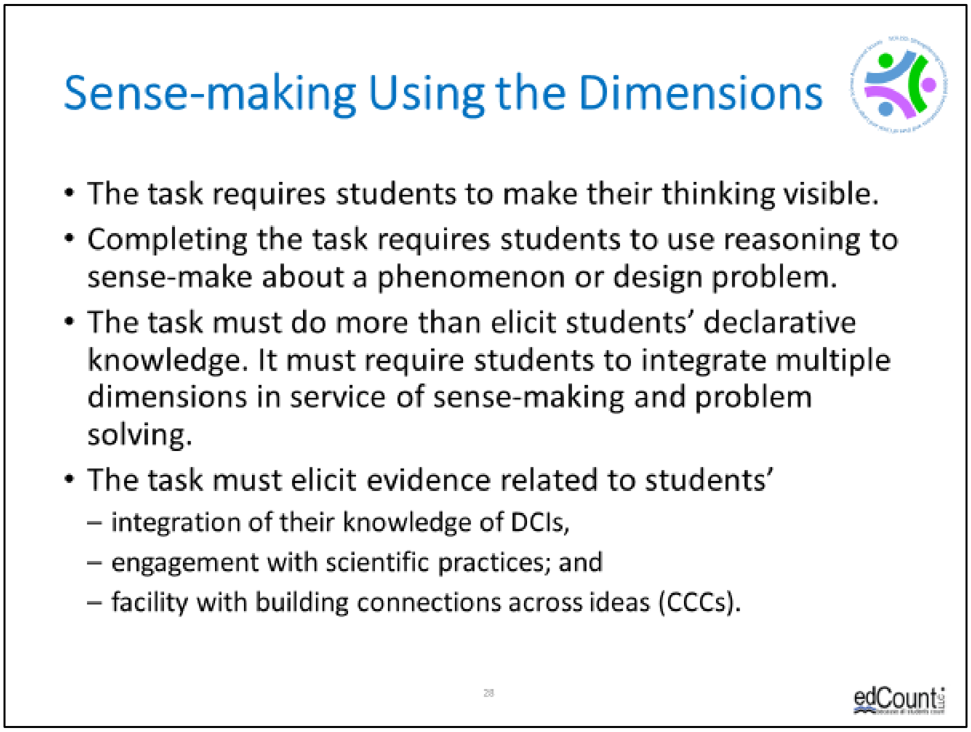
How might you craft a scenario that aligns back to various elements of the task specifications tool? Resume the presentation when you are ready to learn about the next criterion for high-quality tasks: sense-making using the dimensions.



Criterion 3: Sense-making Using the Dimensions defines another set of features of a high-quality task. This criterion helps educators identify what it "looks like" for students to integrate the dimensions for the selected KSAs as they reason and sense-make to explain a phenomenon or design a solution to a problem. Applying this criterion verifies the inclusion of critical task features and the identification of areas for improvement.

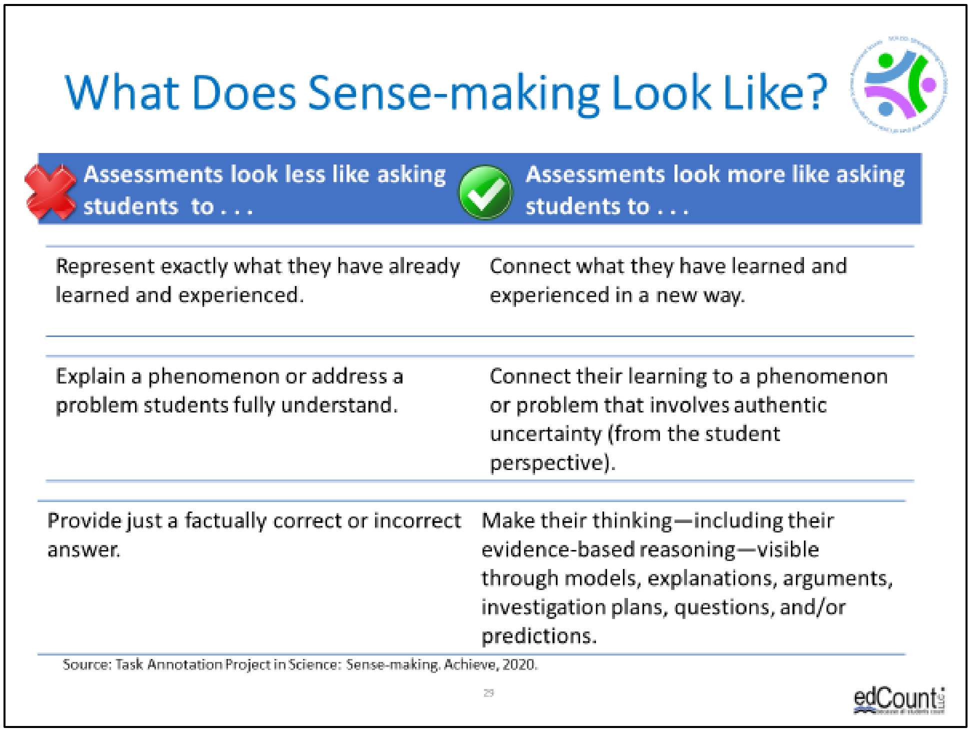


We begin with a definition of sense-making. Tasks require students to make their thinking visible as they integrate the three dimensions in the service of sense-making or problem-solving. Sense-making occurs when students connect their previously developed ideas, abilities, and experiences together to address the uncertainty presented by the phenomenon or design problem—to figure out why the phenomenon occurs, to propose possible mechanisms, designs, or solutions, or to ask further probing questions.



The tasks you develop provide opportunities for students to apply their developing science knowledge to explain phenomena or design solutions to real-world problems.

The tasks must make students' thinking visible. This can be achieved by developing tasks that elicit evidence of what students know and can do and reveal how they are integrating multiple dimensions—the core ideas, the practices, and the crosscutting concepts—and using reasoning to make sense of phenomena and design solutions to problems. These tasks will help you to evaluate if a student, some of your students, or all of your students are learning and understanding what has been taught and can apply or transfer this learning to new scenarios or contexts.



What Does Sense-making Look Like?

Assessments look less like asking students to . . .	Assessments look more like asking students to . . .
Represent exactly what they have already learned and experienced.	Connect what they have learned and experienced in a new way.
Explain a phenomenon or address a problem students fully understand.	Connect their learning to a phenomenon or problem that involves authentic uncertainty (from the student perspective).
Provide just a factually correct or incorrect answer.	Make their thinking—including their evidence-based reasoning—visible through models, explanations, arguments, investigation plans, questions, and/or predictions.

Source: Task Annotation Project in Science: Sense-making. Achieve, 2020.

29

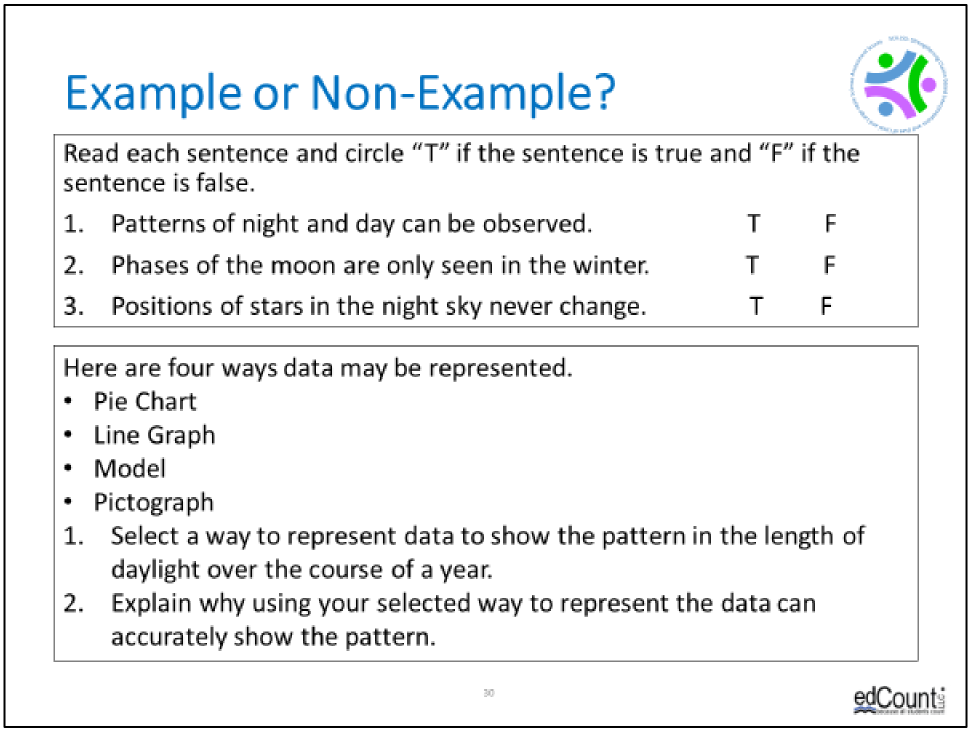
edCount

So, what must a task or assessment ask students *not* to do vs. to do in order to ensure that students are using reasoning to sense-make using the three dimensions to demonstrate grade-appropriate understanding of the assessed KSAs? What are the characteristics of tasks that should be less represented, and what are the characteristics of tasks that should be well represented?

Tasks should not ask students to represent exactly what they have already learned; rather, tasks should ask students to connect what they have learned and experienced in a new way or in novel contexts.

Tasks should not ask students to explain a phenomenon or address a problem students fully understand and that you have observed them competently demonstrating during instruction or investigations; rather, tasks should ask students to connect that learning to a phenomenon or problem that involves authentic uncertainty to which they will apply or transfer their understanding and reasoning to sense-make using the three dimensions.

And, tasks should not provide just a factually correct or incorrect answer such as a key aspect of a DCI presented in a multiple-choice item; rather, tasks should include questions that explicitly prompt students to make their thinking visible through a range of different work products—for example, models, explanations, arguments, investigation plans, questions, and/or predictions. Through these "carriers" of student evidence, you will be able to look to see how the task surfaces students' current understanding, abilities, gaps, and accurate ideas vs. incorrect ideas or misconceptions.



Now that you are well acquainted with criterion 3: sense-making using the dimensions, let's begin a guided activity. In this activity, you will analyze two sets of items to determine which set meets the expectations for criterion 3—an example, and which set does NOT meet the expectations for criterion 3—a non-example. Pause the presentation to evaluate each set and consider whether it is an example or a non-example of sense-making using the dimensions. Then, resume the presentation to consider the explanation provided for each set.

Here is the first set of items:

Read each statement and circle "T" if the sentence is true and "F" if the sentence is false.

1. Patterns of night and day can be observed. T / F
2. Phases of the moon are only seen in the winter. T / F
3. Positions of stars in the night sky never change. T / F

This set of items is a non-example of the application of criterion 3. It does not require the student to sense-make using the dimensions. The student is asked to determine if a statement is true or false, which provides a "correct" or "incorrect" answer about students' recall of core ideas. It provides no information about a student's ability to integrate his or her science knowledge with a SEP or CCC to make sense of a phenomenon. Evidence of correct understanding of these science ideas should be revealed in a task that requires a student to integrate the dimensions to reason about why a phenomenon occurs.

Here is the next set of items:

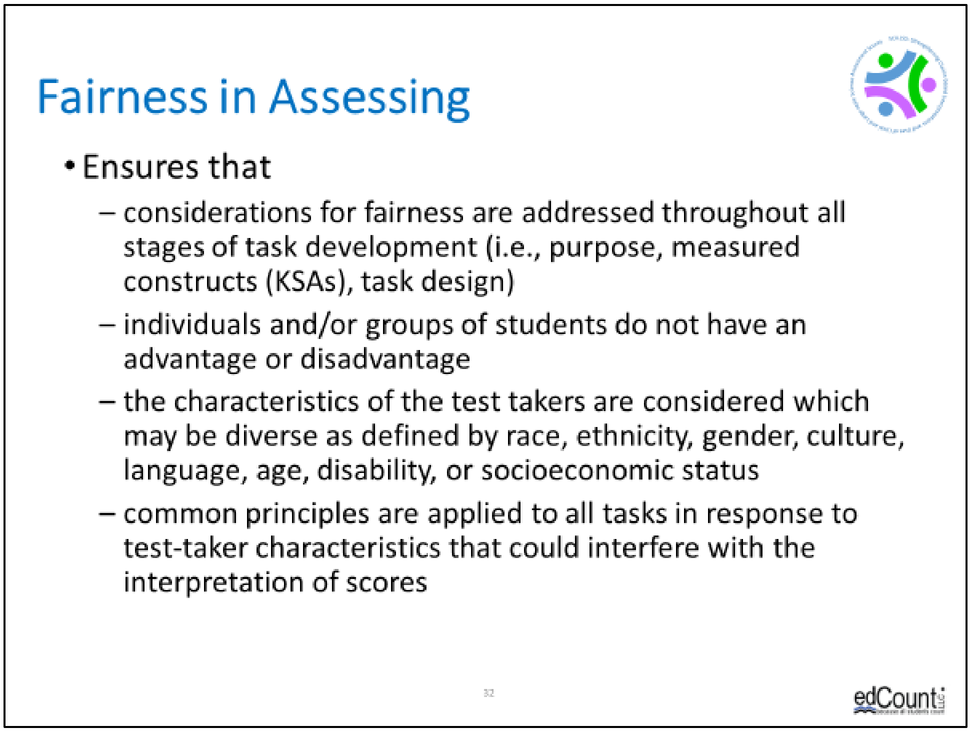
Here are four ways data may be represented.

- Pie Chart
- Line Graph
- Model
- Pictograph

1. Select a way to represent data to show the pattern in the length of daylight over the course of a year.
2. Explain why using your selected way to represent the data can accurately show the pattern.

This set of items is an example of the application of criterion 3. It does require the student to integrate multiple dimensions in the service of sense-making to reveal his or her thinking about why the phenomenon occurs. The student demonstrates understanding of the core ideas by selecting a way to represent data in a graphical display, which addresses a SEP, analyzing and interpreting data. The student is also asked to explain how the selected representation of data accurately shows or reveals the pattern in the length of daylight, which addresses a CCC, *Patterns*.

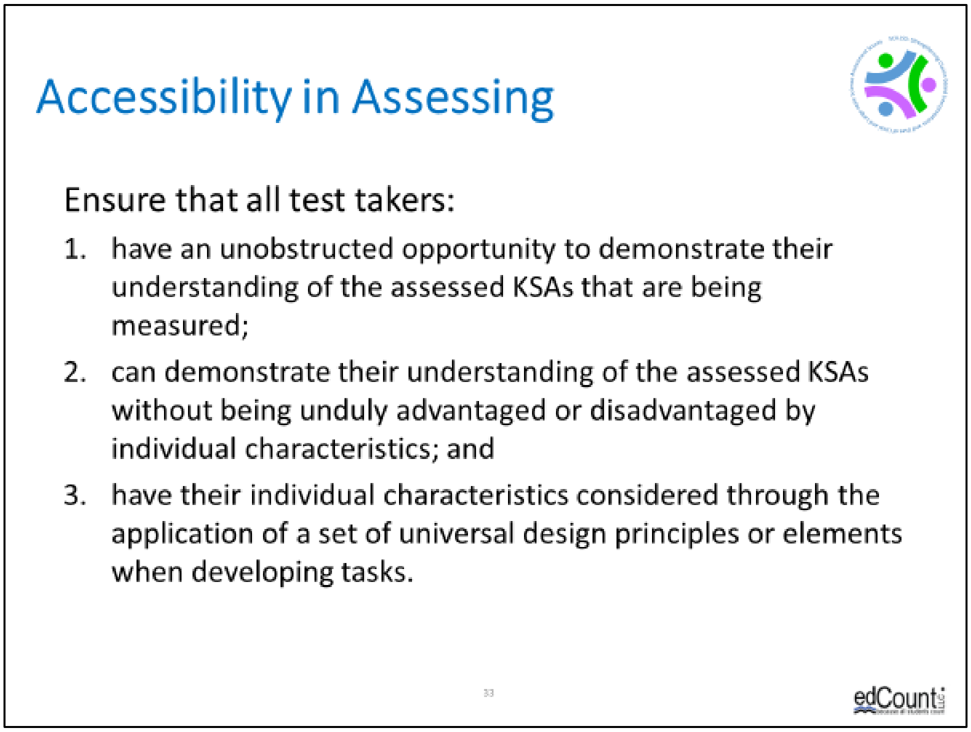
Criterion 4: Fairness: Accessibility, Universal Design, and Bias and Sensitivity defines another set of features of a high-quality task. This criterion helps educators focus on how to enable students to make local, global, or universal connections to the phenomenon or design problem and task at hand and provide ways for students to find meaning and relevancy in the task. Applying this criterion will support evaluation of whether undertaking the task is a meaningful and valuable experience that has real-world relevance for your students and is designed in such a way to enable all students to demonstrate their science learning.



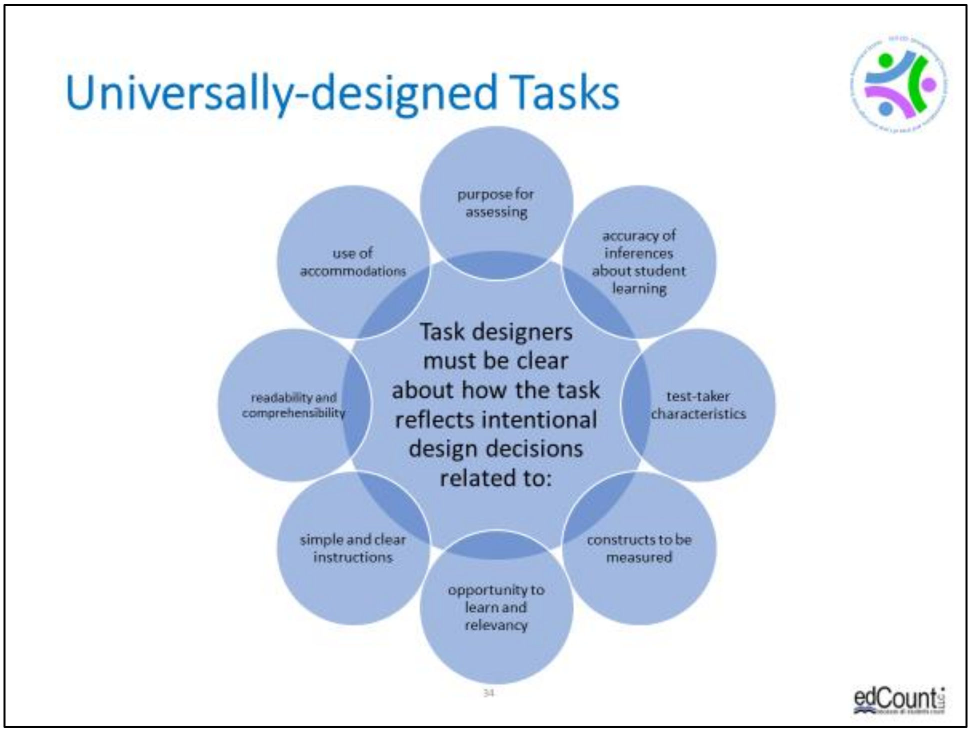
Fairness to all individuals in the intended population of test takers is an overriding, foundational concern. To address fairness in the design of tasks, we must apply common principles to address a range of test-taker characteristics. If we do not, the interpretation of student responses and evaluation of those responses may be threatened (*Standards for Educational and Psychological Testing* (AERA, APA, NCME, 2014, p. 49)).

Fairness in assessing is not an afterthought. Considerations for fairness are addressed throughout all stages of task development, beginning with a clear purpose for assessing, which informs the selection of the knowledge, skills, and abilities to be measured through an intentional task design. Remember, your use of a principled approach to assessment design supports you all along your journey to create high-quality assessment tasks that yield the right information to make informed instructional decisions. This journey begins with the completion of the first and second phases of principled assessment design, development of an unpacking tool and task specifications tool, and continues to phase 3—the design of high-quality three-dimensional classroom science tasks and rubrics.

Your ability to design high-quality tasks that yield meaningful results and support accurate interpretations of scores rests on your attention to fairness in assessing. Fairness in assessing ensures that tasks are not designed to provide an advantage or disadvantage to any individual student or group of students. In this module, we aim to build your understanding of fairness by exploring how to consider the diverse characteristics of test takers and how to apply common accessibility principles to ensure that tasks are designed with all students in mind.



Accessibility in assessing is necessary to promote fairness for all test takers. What do we mean when we talk about accessibility when assessing students' knowledge? There are three key aspects. As science task designers, you need to consider how to ensure that all test takers 1) have an unobstructed opportunity to demonstrate their understanding of the assessed KSAs that are being measured; 2) can demonstrate their understanding of the assessed KSAs without being unduly advantaged or disadvantaged by individual characteristics; and 3) have their individual characteristics considered through the application of a set of universal design principles or elements when developing tasks. Let's take a closer look at principles for creating accessible tasks.

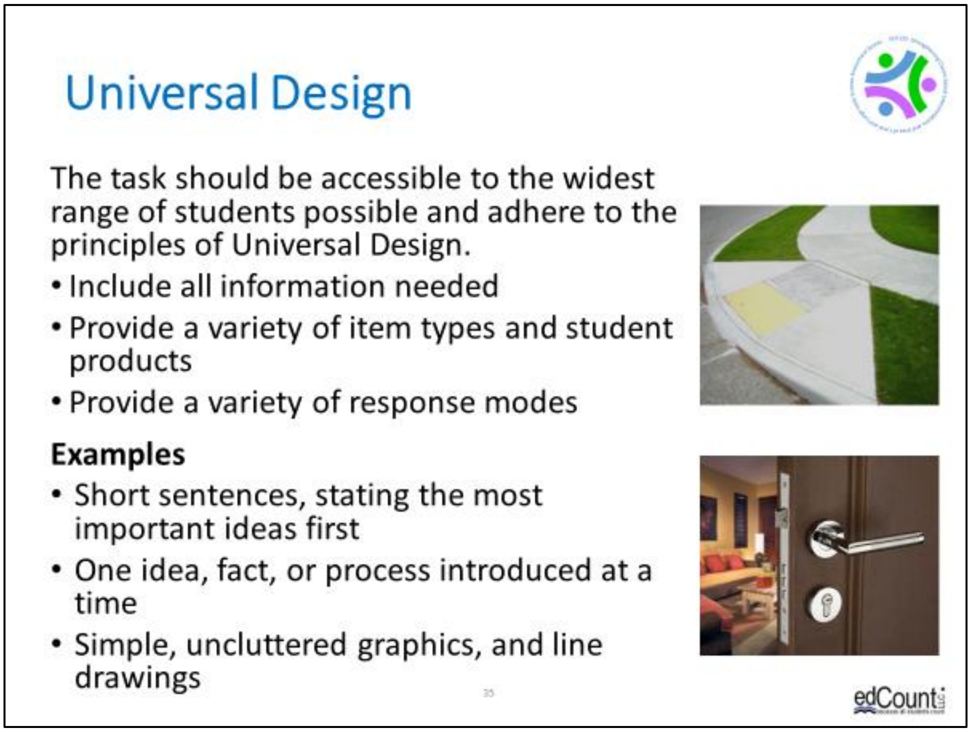


Universal design principles address policies and practices that are intended to improve access to learning and assessments for all students. Universal design principles are important to the development and review of assessments because some assessment designs pose barriers that bar students with disabilities (or possibly other test takers) from showing what they know. Universal design techniques can result in a more accurate understanding of what students know and can do (NCEO, n.d.).

As task designers, you must be clear about how the task reflects intentional design decisions related to the purpose for assessing and for collecting the right evidence and sufficient evidence to make accurate inferences about student learning. You must take into consideration the characteristics of the test taker and determine how to promote engagement, interest, and motivation; provide different methods for students to demonstrate what they have learned, such as through writing, drawing, or demonstration; and present items by representing them in different formats such as text, graphs, images, or objects.

Be clear on the selection of constructs or KSAs to be measured at a particular point during instruction and ensure students have had sufficient opportunity to learn the constructs being measured. Clear learning goals result in the development of construct-focused assessment tasks that measure whether students have achieved those goals. This results in more authentic tasks that allow students to transfer their knowledge to related or new phenomena and design problems. When students recognize the connection between instruction and the tasks you develop, the purpose and relevancy of the assessment is made clear and student engagement is promoted (CAST, 2009).

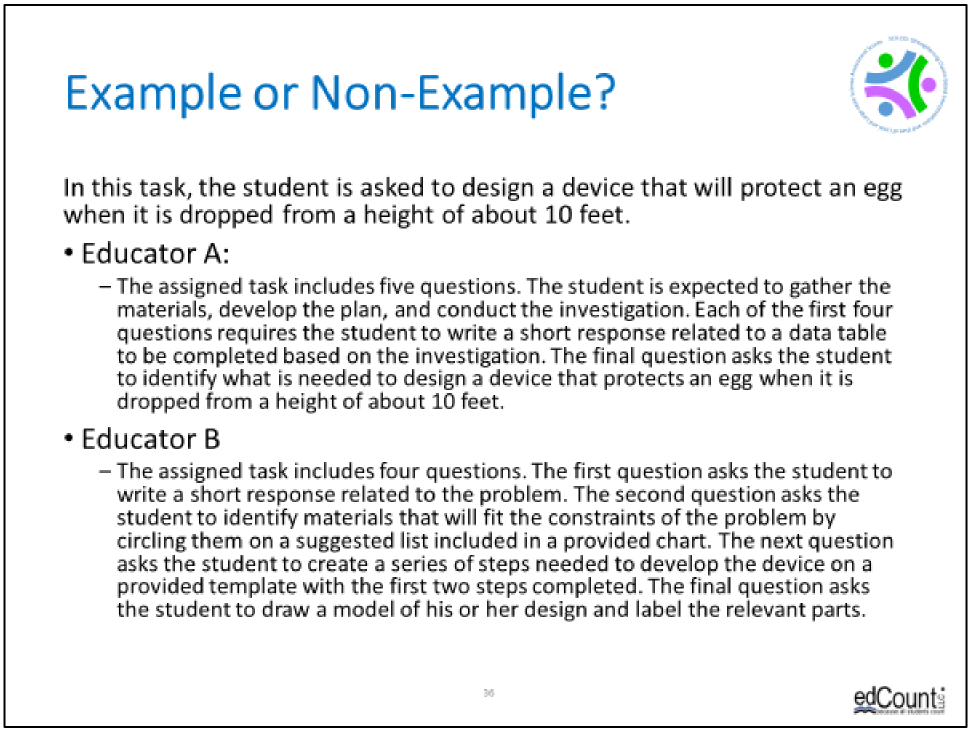
Instructions and questions that students will use to guide their responses must be simple and clear, and their readability and comprehensibility must be considered to ensure accessibility by all test takers. In general, the amount of reading should be kept to a minimum, given the context of the problem. Tasks should, for example, minimize unnecessary or confusing language and overly complex sentence structures and should include relevant vocabulary that has been taught and used during investigations and science learning. Ensure task language is at- or below- grade level. Note that sometimes domain-specific language may be above-grade-level but is appropriately used in a task if a term is commonly used during instruction. Also, consider how the use of accommodations for students with disabilities and English learners, utilized during instruction, can be made available and implemented during task administration.



The term universal design has been applied to a variety of educational approaches over the past several years. More than 20 years ago, Ron Mace, an architect who was a wheelchair user, began to actively promote a concept he termed "universal design." Mace was adamant that his field did not need more special-purpose designs that serve primarily to meet compliance codes and may also stigmatize people. Instead, he promoted design that works for most people, from the child who cannot turn a doorknob to the elderly woman who cannot climb stairs to get to a door (Mace, 1998).

Universally-designed assessments add a dimension of fairness to the testing process. The goal is that the task should be accessible to the widest range of students possible and adhere to the principles of Universal Design. As task designers, we must create universally-designed assessments by considering how to minimize barriers. This includes, for example, providing all of the necessary information in the task to elicit students' background knowledge and to enable students to produce a complete and accurate response or incorporating a variety of item types and student work products to allow students to demonstrate their learning using a variety of response modes, such as verbal, written expression, and/or drawings, diagrams, graphs, presentations, or videos.

To address readability and comprehensibility, we must attempt to use short sentences, stating the most important idea first, introduce one idea, fact, or process at a time, and use simple, uncluttered graphics and line drawings that convey the appropriate information to be interpreted by the test takers to respond fully to a question. Consider your students and how you may more deliberately address universal design in your assessments.



Now that you are more familiar with criterion 4: fairness: accessibility, universal design, and bias and sensitivity, here is another activity to evaluate an example and non-example of the application of this criterion to a classroom science assessment task. As before, in this activity, you will analyze two statements, one from Educator A and one from Educator B, to determine which task description meets the expectations of criterion 4—an example, and which task description does NOT meet the expectations of criterion 4—a non-example. Pause the presentation to evaluate each statement and consider whether it is an example or a non-example of fairness. Then, resume the presentation to consider the explanation provided for each statement.

In these task descriptions, the student is asked to design a device that will protect an egg when it is dropped from a height of about 10 feet.

Here is the statement from Educator A:

The assigned task includes five questions. The student is expected to gather the materials, develop the plan, and conduct the investigation. Each of the first four questions requires the student to write a short-response related to a data table to be completed based on the investigation. The final question asks the student to identify what is needed to design a device that protects an egg when it is dropped from a height of about 10 feet.

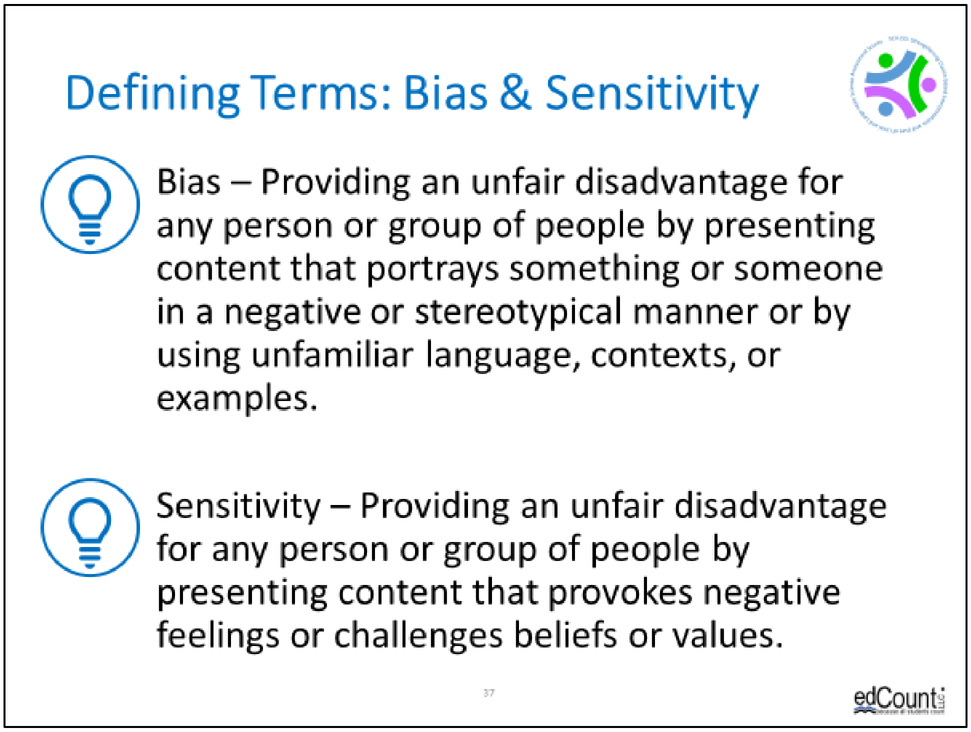
This statement is a non-example of the application of criterion 4 to a task. It does not reflect a task that was designed using principles of universal design. Educator A has designed a complex task that requires multiple steps to complete, which may not be easily accomplished within a class period as intended for a formative assessment. Fairness may be compromised for some test takers, given that the students are asked to gather the materials to conduct the investigation. What if the materials determined by the test taker are not available in the classroom or if an insufficient number of materials are available? Four of the five questions require the same work product, written short-response, and do not allow for students to provide other ways in which their science knowledge can be elicited. Given potential time constraints and the number and types of questions, it may be challenging as students may not have sufficient time to fully produce evidence to show their understanding of the measured constructs. This could then lead to incorrect inferences about what students know and can do.

Here is the statement from Educator B:

The assigned task includes four questions. The first question asks the student to write a short response related to the problem. The second question asks the student to identify materials that will fit the constraints of the problem by circling them on a suggested list included in a provided chart. The next question asks the student to create a series of steps needed to develop the device on a provided template with the first two steps completed. The final question asks the student to draw a model of his or her design and label the relevant parts.

This set of statements is an example of the application of criterion 4. The task description reflects various principles of universal design. Educator B has designed a task that includes a reasonable number of interconnected questions that vary in format and are of varying complexity to be addressed by the student during an instructional period. The first question asks the student to write a short response related to the problem, which creates a focus for the design of the solution and may help the student then consider what the constraints might be. Rather than generating a list of constraints in a written response, the second question asks the student to identify materials that will fit the constraints of the problem by circling them on a suggested list included in a provided chart—an alternative presentation and required response mode. The next question asks the student to complete a template by adding additional steps to carry out an investigation. The student response is scaffolded by the provision of a template and two initial steps from which subsequent steps can be developed. This may also help the student to structure their response for accurate interpretation. The final question asks the student to draw a model of their design and label the relevant parts, thus providing a different response mode for students to demonstrate what they know and understand about designing a solution to a design problem.

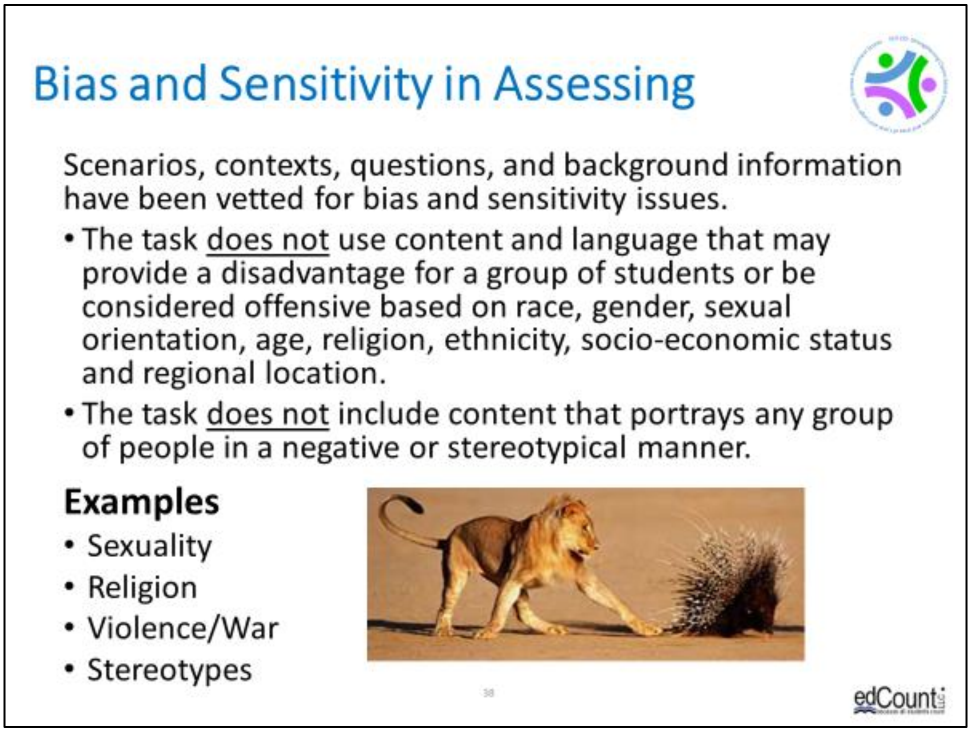
Now let's examine the aspect of bias and sensitivity as it relates to criterion 4.



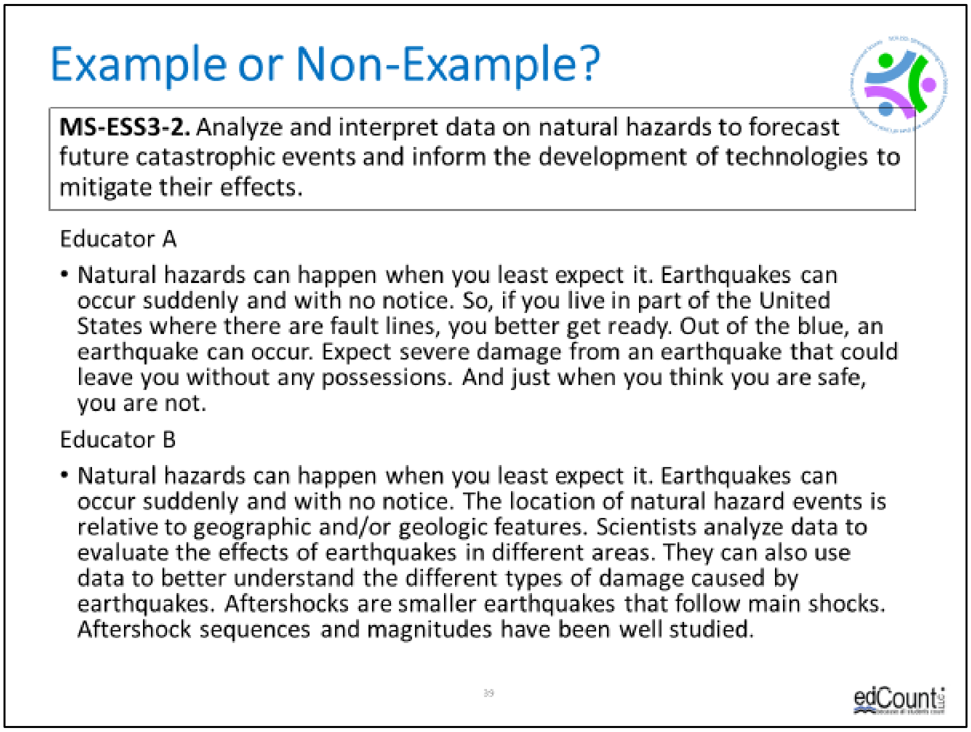
As we have shared, the design of a high-quality task that yields meaningful results and supports accurate interpretations of scores rests on your attention to fairness in assessing. You want to ensure that groups of students do not have an advantage or disadvantage, and that bias and sensitivity issues are minimized. Your ability to consider potential bias and sensitivity issues from the earliest stages of task development is enhanced by having a common understanding of definitions of these terms.

Bias refers to providing an unfair disadvantage for any person or group of people by presenting content that portrays something or someone in a negative or stereotypical manner or by using unfamiliar language, contexts, or examples.

Sensitivity refers to providing an unfair disadvantage for any person or group of people by presenting content that provokes negative feelings or challenges beliefs or values.



Attention to bias and sensitivity in assessing provides another avenue by which you can ensure that you are maximizing accessibility for all of your students. When preparing assessment items, be sensitive to the possibility of unintentionally placing groups of students at an unfair disadvantage. When the task avoids the use of content and language that disadvantages a group of students and avoids negative or stereotypical content, again, you are maximizing accessibility. Student engagement with the task is more likely to be increased if you design the task from beginning to end with these potential issues in mind. When students deem tasks to be relevant, non-offensive, and engaging, you can expect the student to produce evidence that is commensurate with their science learning and understanding.



Let's examine another aspect of criterion 4: bias and sensitivity. As before, in this activity, you will analyze two statements, one from Educator A and one from Educator B, to determine which task scenario meets the expectations of criterion 4 with respect to bias and sensitivity—an example, and which task scenario does NOT meet the expectations of criterion 4—a non-example. Pause the presentation to evaluate each statement and consider whether it is an example or a non-example of fairness with respect to bias and sensitivity. Then, resume the presentation to consider the explanation provided for each statement.

These task scenarios are based on the NGSS PE, *MS-ESS3-2: Analyze and interpret data on natural hazards to forecast future catastrophic events and inform the development of technologies to mitigate their effects.*

Here is the task scenario from Educator A:

Natural hazards can happen when you least expect them. Earthquakes can occur suddenly and with no notice. So, if you live in part of the United States where there are fault lines, you better get ready. Out of the blue, an earthquake can occur. Expect severe damage from an earthquake that could leave you without any possessions. And just when you think you are safe, you are not.

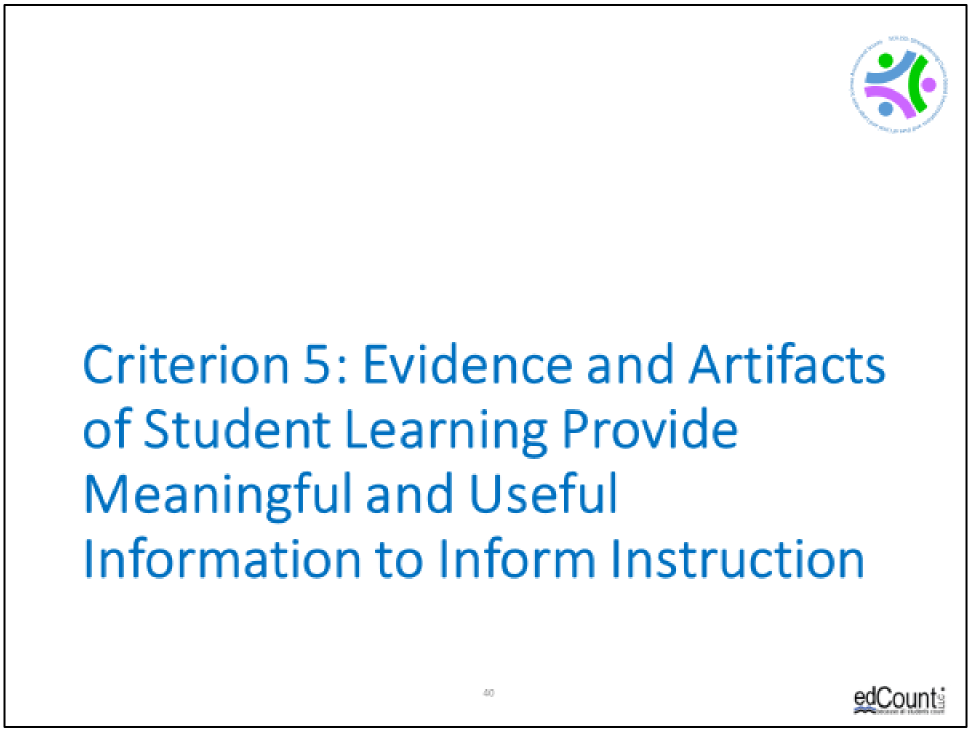
This scenario is a non-example of the application of criterion 4 to a task scenario. It does not reflect a task that was designed and vetted with attention to bias and sensitivity. Educator A has designed a scenario that begins with accurate information, but it then presents a natural hazard in a context that is dramatic and includes little information related to phenomenon-based scientific ideas. Rather, the scenario relates natural hazards to tragedy. It is accurate that natural hazards may result in catastrophe for groups of people depending on where they live. However, the use of exaggeration and fear, such as by including the phrase, "you better get ready" may be upsetting for some students whether they live in a geographic area that is more or less likely to experience earthquakes. The inclusion of the idiom, "Out of the blue," may not hold meaning for some students and may be misleading or confusing when considering the phenomenon. The scenario uses content that may disadvantage students who live in particular regional locations, and the provided information may cause concern or a strong reaction that may interfere with students' ability to remain engaged to fully respond to the task. This could then lead to incorrect inferences about what students know and can do.

Here is the task scenario from Educator B:

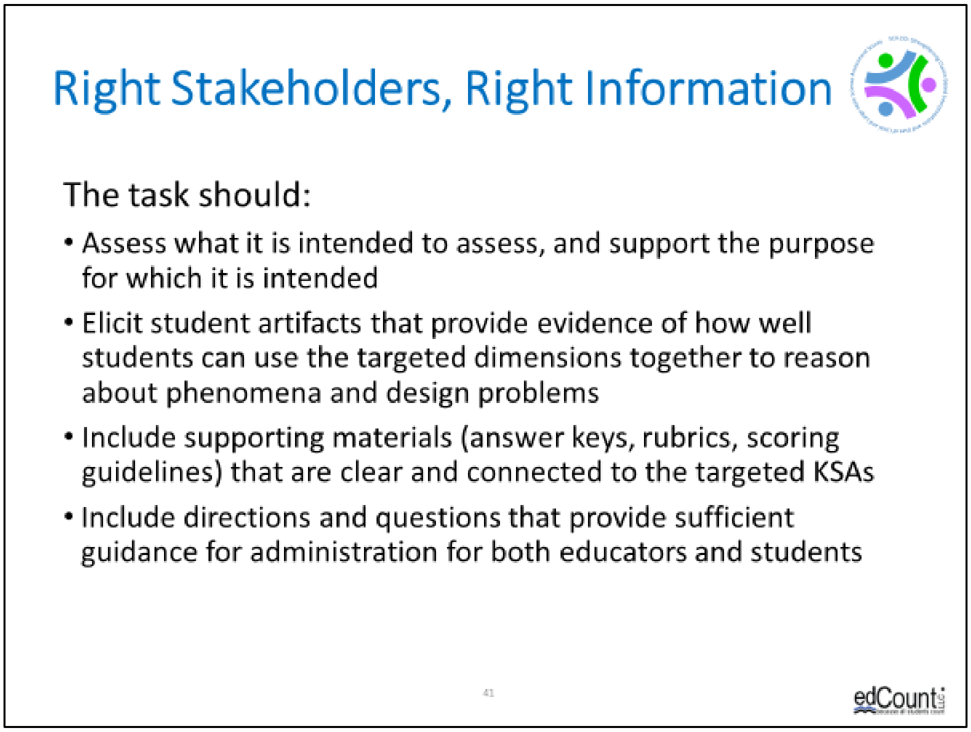
Natural hazards can happen when you least expect them. Earthquakes can occur suddenly and with no notice. The location of natural hazard events is relative to geographic and/or geologic features. Scientists analyze data to evaluate the effects of earthquakes in different areas. They can also use data to better understand the different types of damage caused by earthquakes. Aftershocks are smaller earthquakes that follow mainshocks. Aftershock sequences and magnitudes have been well studied.

This scenario is an example of the application of criterion 4 to a task scenario. It does reflect a task that was designed and vetted with attention to bias and sensitivity. Like Educator A, Educator B has designed a scenario that begins with accurate information. Unlike Educator A, Educator B then goes on to expand on factual information related to the phenomenon, earthquakes. The scenario makes connections between that information and that scientists have learned about earthquakes through analyzing and interpreting data—the SEP associated with this PE. The scenario is factually accurate and is devoid of exaggeration or phrases or expressions that may be unfamiliar to students or provoke a negative reaction, and thus promotes engagement and increases comprehensibility. The scenario is more likely to support students' ability to use prior knowledge and to remain engaged with the task in order to produce evidence that is more fully representative of their science learning.

Now we will address criterion 5: Evidence and Artifacts of Student Learning Provide Meaningful and Useful Information to Inform Instruction and consider how this criterion contributes to the design of high-quality tasks.



Tasks must be designed with a clear purpose in mind based on the KSAs to be assessed and must produce evidence and artifacts of student learning that result in useful information to support educators' decisions about the need for adjustments to instruction for individual students, groups of students, or the class. Criterion 5: Evidence and Artifacts of Student Learning Provide Meaningful and Useful Information to Inform Instruction supports educators in making accurate inferences about students' science learning and in evaluating students' readiness to learn new, more sophisticated science concepts stemming from previous understandings.



Educators and other stakeholders, such as students and parents, need the right information to gain an understanding of where students are in their science learning of the assessed knowledge, skills, and abilities.

Gaining the "right information" requires the use of classroom assessment tasks that do not exist in isolation, disconnected from the implemented curriculum and instruction, but rather tasks that produce meaningful evidence about student learning so that educators can make well-informed decisions to adjust instruction or continue with the instructional sequence. These tasks are also formative. They provide information about whether students learned what they have experienced during recent instruction, whether students can apply what they learned to similar but new contexts, or whether students can generalize their learning to a different context. The tasks support educators' identification of students' misconceptions, misunderstandings, or lack of understanding related to the integration of the SEPs, CCCs, and/or DCIs.

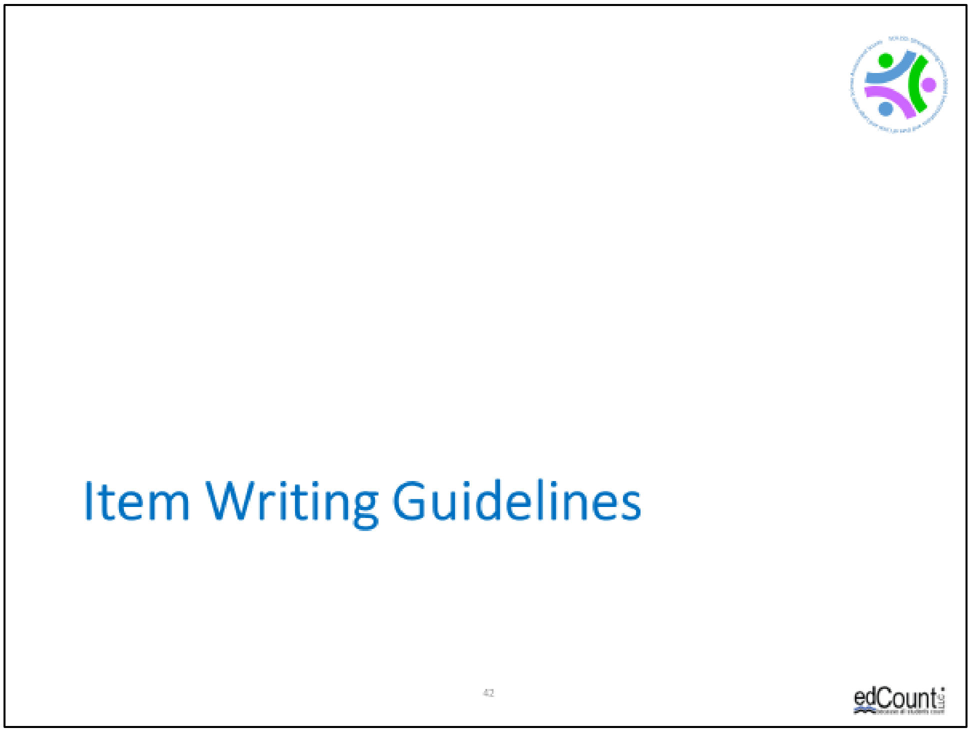
Well-designed tasks also elicit student "artifacts" or work products that provide clear and strong evidence of how well students integrate the targeted dimensions to make sense of phenomena and design solutions to problems. As a task designer, it is critical to consider how the student responses elicited support the purpose of the task. For example, if a task is intended to help educators determine if students understand the distinction between physical and chemical reactions, does the task support this inference?

Gaining the "right information" also requires that all materials associated with administering, evaluating, and scoring student evidence, such as answer keys, rubrics, and/or scoring guidelines, are clear and sufficient. Consider how well these support materials are developed to guide accurate evaluations and interpretations of a range of student responses that may include incomplete scientific understandings or misconceptions or to reveal barriers to students' abilities to demonstrate their acquired science knowledge. For example, are there language or vocabulary barriers or a lack in the variety of question types or response modes that limit or prevent students with specific characteristics from responding or responding fully?

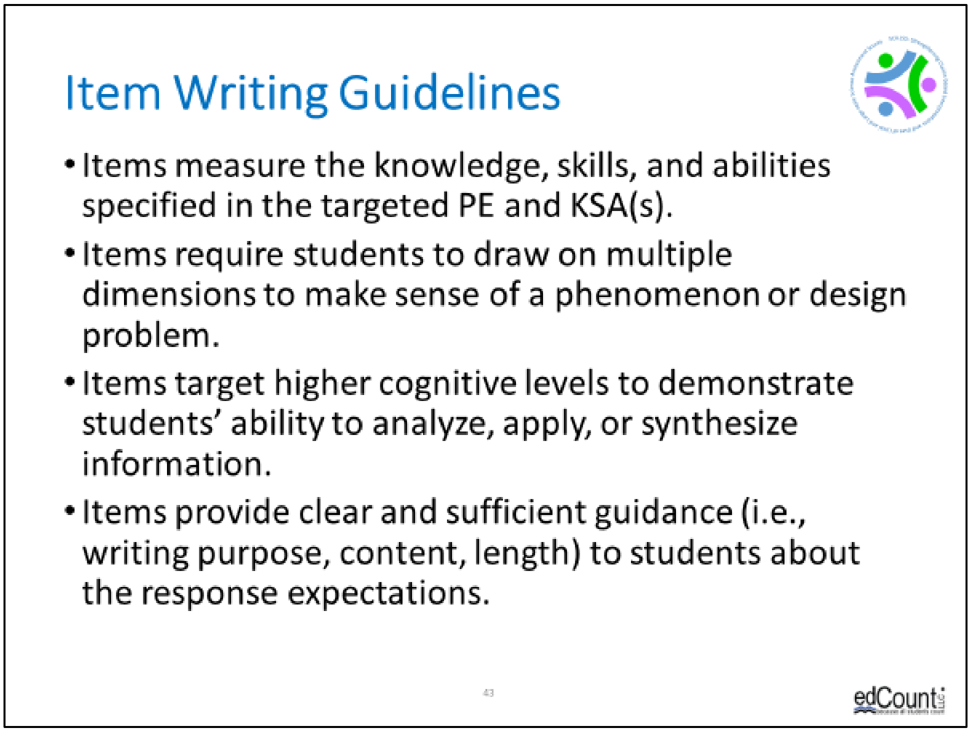
And finally, to achieve the intended purpose for undertaking the development of classroom science tasks using a principled design approach, the directions and questions must be written with clarity to promote accurate administration by educators and interpretation by students. This will promote accessibility for both groups!

Tasks that are intentionally designed with these criteria in mind will yield accurate and useful interpretations of students' knowledge, skills, and abilities at a given point in time during instruction and will provide educators with the evidence they need to make defensible, well-informed decisions about the instructional needs of their students.

Now let's look at how the application of item writing guidelines can promote the generation of high-quality science tasks that are fair and accessible and produce the right information for the intended purposes of the tasks.



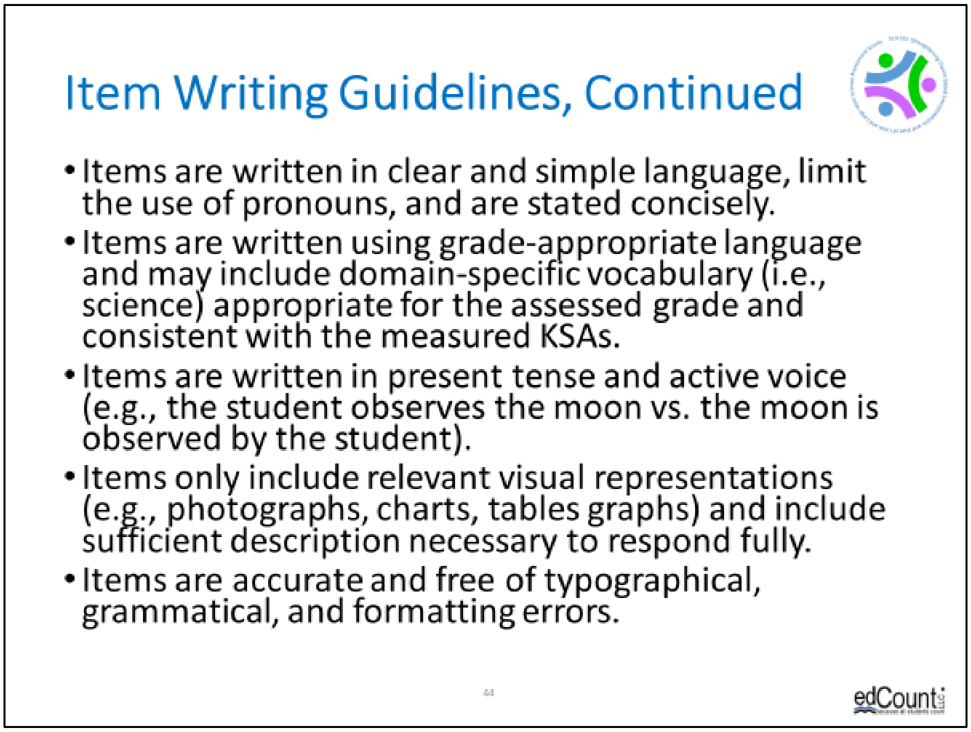
Writing good test questions is both an art and a science. It requires imagination and creativity, but also requires following guidelines to accurately measure students' science learning. Now that you have a good understanding of fairness in assessing, we will introduce basic item writing guidelines to contribute to your ability to write clear and comprehensible questions and produce a range of items and item types that elicit evidence of a wide range of student abilities and science understandings.



Here we present basic item writing guidelines. As the guidelines are discussed, consider how you are currently applying them in the development of your classroom science assessment tasks. Consider how you might further apply the guidelines to promote fairness and accessibility for all test takers to ensure you are getting the "right information" at the "right time" consistent with your purpose for assessing. And keep in mind that as a task designer, you need to evaluate test length and determine the number of items to include in a task based on the estimated time required for students to complete the task balanced against the allotted time for task administration.

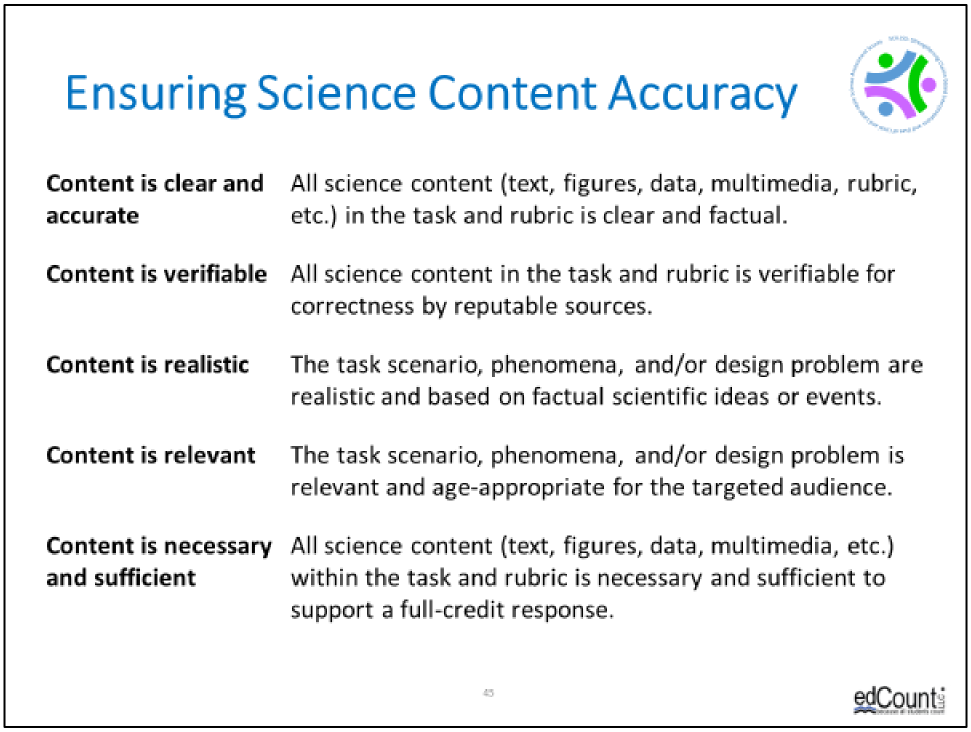
As we have shared, formative science tasks that are created using a principled-design approach include a high-quality scenario or context that couches the phenomenon or design problem associated with a selected PE or indicator. The scenario serves as the driver from which the items flow. Each item includes a stem or question and describes the nature of the expected response.

Items must be aligned to the targeted PE and selected KSAs. Remember, a clear and accurate understanding of the KSAs associated with a PE is prerequisite to instruction and to designing assessment tasks. To measure student understanding and science learning based on three-dimensional science standards, items must require students to draw on and reason using multiple dimensions to make sense of a phenomenon or design problem. This requires that items target higher cognitive levels to demonstrate students' ability to analyze, apply, or synthesize information. We expect students to have different degrees of competency with the assessed KSAs and that students have a range of ability. A high-quality task should provide items of varying complexity levels to elicit different types of student evidence. In addition, items must provide clear and sufficient guidance (i.e., writing purpose, content, length) to students about the response expectations to promote fairness and equity. Clear and sufficient guidance ensures that students have a clear understanding of what is being assessed and the response expectations to produce accurate and complete evidence that meets the highest level of student performance.



Students' ability to demonstrate what they know and can do is enhanced when items are written in clear and simple language, the use of pronouns is limited, and statements and directions are stated concisely. Writing the scenario and items in present tense and avoiding clauses and conditional words such as "if" or "suppose" can also promote clarity and comprehensibility. Items should include grade-appropriate language and may include domain-specific vocabulary appropriate for the assessed grade and consistent with the measured KSAs. Visual representations, for example, photographs, charts, tables, and graphs, can serve as relevant sources of information to support students' full and complete responses to a question. Task designers must include sufficient written descriptions of these visual representations, commonly referred to as alternative text, for accurate use and interpretation by all test takers. Finally, test developers must carefully copyedit tasks to ensure that all items and all aspects of the task are accurate and free of typographical, grammatical, and formatting errors.

It may seem like a lot to consider, but with practice, the art and science of writing items will emerge, and your items will increase in clarity and comprehensibility and thus promote fairness and accessibility to all test takers.



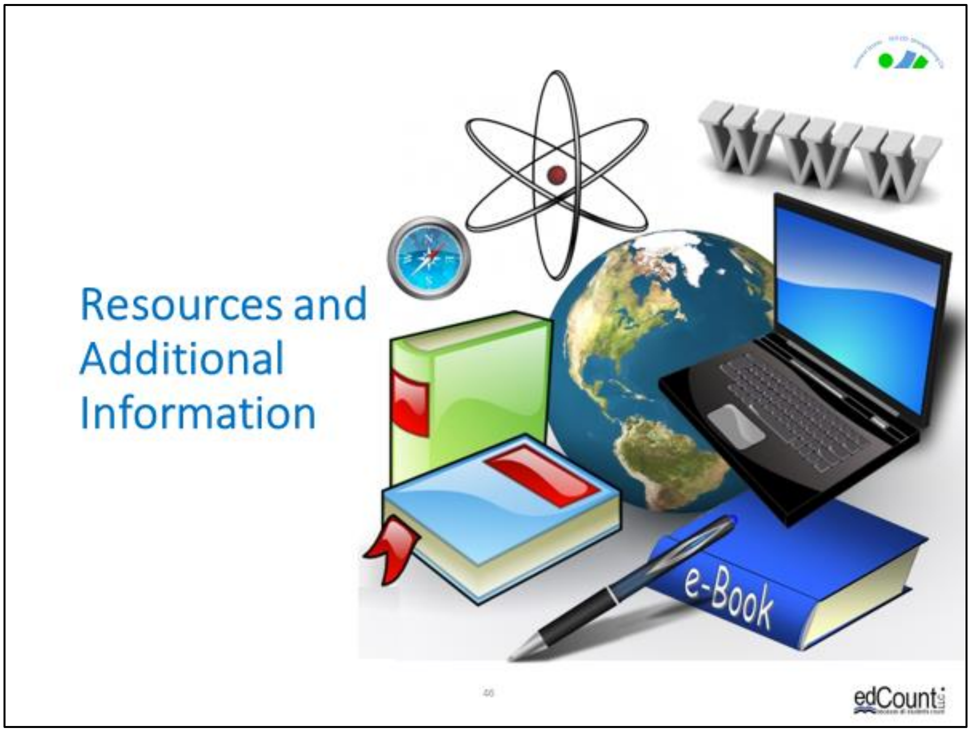
The NGSS represent a significant shift from previous state science standards because they require a multi-dimensional approach to teaching and assessing science learning. For many school districts and science educators, both the number and degree of changes required to implement the NGSS or new state science standards may be significant. Time is needed by educators to understand these changes or shifts in the standards with respect to their teaching goals, structure, content, and use of the NGSS to evaluate and address teaching strategies and curriculum and, as is our focus in this chapter, to assess science learning aligned to three-dimensional science standards. These changes require dedicated time for full implementation of the standards.

Science education has traditionally focused on content—basic facts and science ideas—and vocabulary, which has proved inadequate and insufficient in teaching students a deeper understanding of key scientific concepts and the application of these concepts to daily life and understanding of our world. The NGSS calls for refocusing K–12 science instruction, and we would add the assessment of three-dimensional science learning to explain phenomena or design solutions to design problems, to improve college preparation, Science, Technology, Engineering, and Mathematics (STEM) career readiness, and the ability of all members of society to make informed decisions. Your command of the DCIs, SEPs, and CCCs is critical to ensure that the science content included in tasks is accurate. This is a required feature of every component of an assessment task to promote fairness as students interpret and respond to scenarios, questions, and expectations for demonstrating what they know and can do.

To ensure that the science content is accurate in the tasks that you design and to provide all students a fair and equitable opportunity to demonstrate what they know and can do, consider:

- including content that is clear and factual, as represented in text, figures, data, multimedia and in the rubric;
- verifying the correctness of included science content using the NGSS and your expertise, and through consultation with an array of information obtained from reputable data sources;
- basing the scenario, selection of phenomena, and/or design problems on realistic and factually correct scientific ideas or events;
- attending to the targeted audience to promote engagement and relevancy with content that is age-appropriate for the targeted audience; and
- supporting student production of a full-credit response by including necessary and sufficient science content in the scenario to reference and elicit background knowledge to enable students to produce accurate and complete evidence that meets the criteria for a full-credit response.

Rest assured, implementing the vision of the NGSS, and the *Framework* is challenging, as is creating assessments that measure students' understanding of three-dimensional science standards, but it is well worth the effort. As we incorporate the shifts and address the challenges inherent in understanding, teaching, learning, and assessing the NGSS, we are moving forward to achieve the vision of the *Framework* and its goals for K–12 science education to educate all students in science and engineering and to provide the foundational knowledge for those who will become the scientists, engineers, technologists, and technicians of the future (the *Framework,* p. 10).

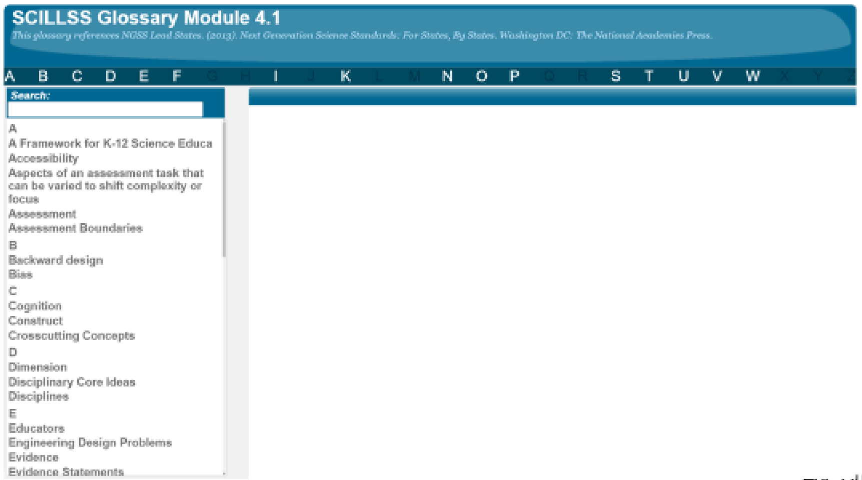


Finally, we offer additional resources that may be helpful to anyone interested in learning more about the concepts presented in this module. A glossary of terms and our reference list follow.

Thank you for your engagement in this fourth chapter of the SCILLSS digital workbook on designing high-quality three-dimensional science assessment tasks for classroom use.

SCILLSS Glossary

Please refer to the SCILLSS Glossary for operational definitions of terms used.



Resources

In the Web Links pod, you can find the following resources:

- A Framework for K-12 Science Education
- NGSS Science Task Screener
- Phenomena for NGSS
- The Wonder of Science Phenomena
- Project Phenomena

In the Resources pod, you can find the following resources:

- Classroom Assessment Task Review Worksheet
- Unpacking Tool for 5-ESS1-2
- Task Specifications Tool for 5-ESS1-2

References

Achieve. (2018). NGSS Task Screener. Retrieved from https://www.nextgenscience.org/sites/default/files/resource/files/Achieve%20Task%20Screener_Final_9.21.18.pdf

Achieve. (2020). Task Annotation Project in Science: Sense-making. Retrieved from https://www.achieve.org/publications/task-annotation-project-science-sense-making

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

CAST. (2009). UDL Tips for Assessment. Retrieved from http://www.cast.org/binaries/content/assets/common/publications/downloads/cast-udltipsforassessment-20200920-a11y.pdf

Mace, R. (1998). A perspective on universal design. An edited excerpt of a presentation at Designing for the 21st Century: An International Conference on Universal Design. Retrieved January, 2002, from the World Wide Web: www.adaptenv.org/examples/ronmaceplenary98.asp?f=4.

National Research Council. (1999). High stakes: testing for tracking, promotion, and graduation (J. Heubert & R. Hauser editors, Committee on Appropriate Test Use). Washington, DC: National Academy Press.

NCEO. (n.d.). Universally Designed Assessments. Retrieved from https://nceo.info/Resources/publications/TopicAreas/UnivDesign/UnivDesignTopic.htm

NCEO. (n.d.). Universally Designed Assessments: Frequently Asked Questions. Retrieved from https://nceo.info/Resources/publications/TopicAreas/UnivDesign/UnivDesignFAQ.htm

Thompson, S. J., Johnstone, C. J., Anderson, M. E., & Miller, N. A. (2005). Considerations for the development and review of universally designed assessments (Technical Report 42). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.

edCount