Designing High-Quality Three-Dimensional Science Assessments for Classroom Use

Chapter 1: Introduction to Principled Assessment Design

Module 1.2: What is Principled Assessment Design and Why Use It?

This digital workbook on designing high quality three-dimensional science assessment tasks for classroom use was developed by edCount, LLC, under Enhanced Assessment Grants Program, CFDA 84.368A.

Welcome to the first 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 high-quality classroom science assessment 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.

Module 1.2:
What is Principled
Assessment
Design and Why
Use It?
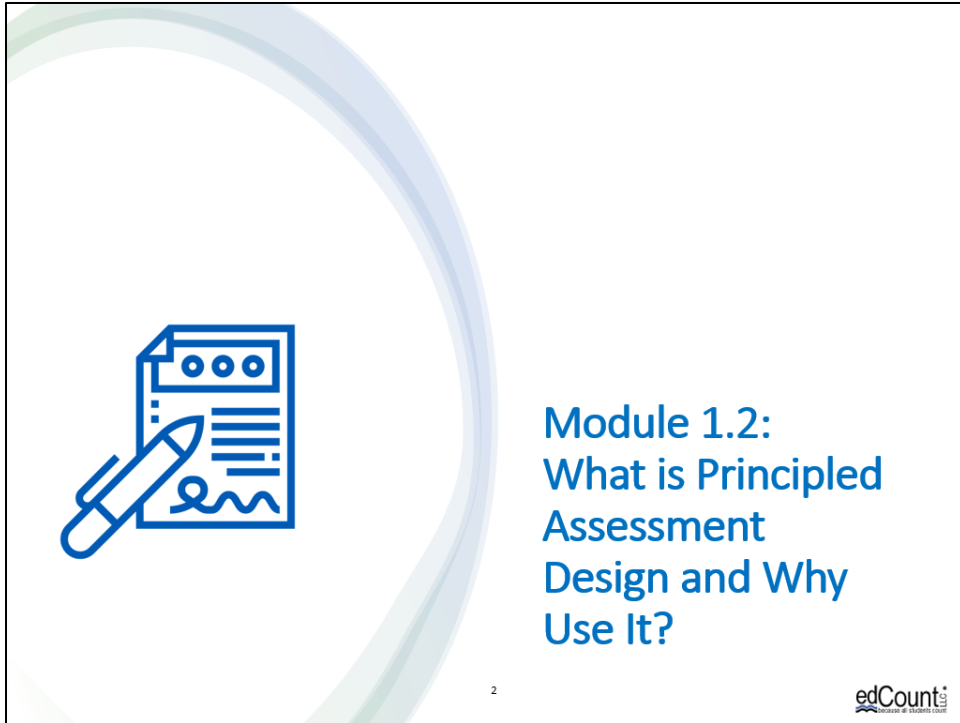
Chapter 1 of this workbook includes a series of three modules. Module 1.1 focuses on classroom-based assessments, their relationship to other forms of assessment, and their purposes and uses in a standards-based system of curriculum, instruction, and assessment. Modules 1.2 and 1.3 introduce principled-assessment design, a backward design approach for developing three-dimensional tasks aligned to the Next Generation Science Standards (NGSS) for use within classrooms.

In these Chapter 1 modules, we lay the groundwork for educators to engage in intentional assessment design by considering the purposes and uses of classroom assessments and the types of tasks and situations that elicit evidence that supports meaningful interpretations of students' science learning.

Let's take a closer look at the outcomes associated with Module 1.2. What is Principled Assessment Design and Why Use It?
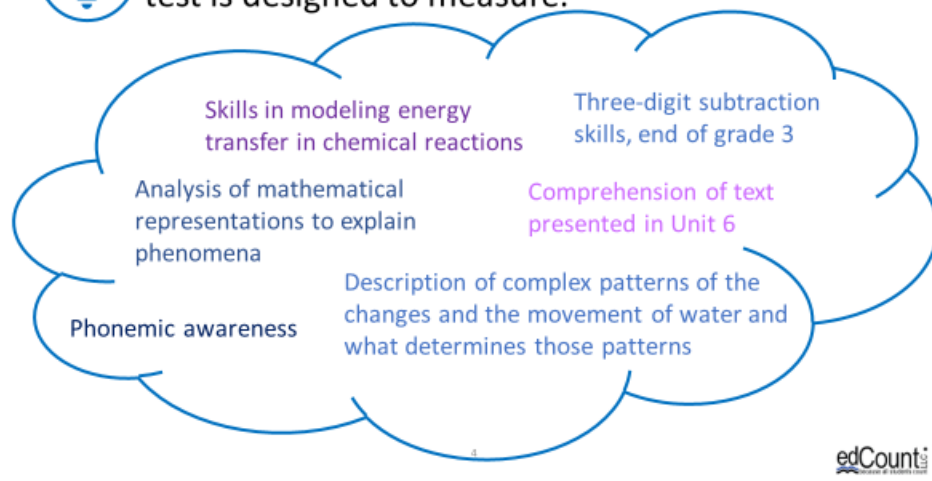
The new vision for K-12 science education calls for engaging students in three-dimensional learning and, thus, requires us to use new ways to assess this learning. Meaningful measurement of increased and more in-depth understanding of these complex, three-dimensional standards over the course of students' K-12 educational experience requires that assessment tasks ask students to apply knowledge of disciplinary core ideas and crosscutting concepts while engaging in a science or engineering practice.

This module begins with an introduction to principled assessment design and the reasons and benefits for using it to design and evaluate classroom science assessment.

A construct is the concept or characteristic that a test is designed to measure. In education settings, the constructs of most interest have to do with content knowledge and skills or personal or social characteristics that often relate to academic performance. We cannot directly observe these constructs and must present students with opportunities—such as tests—when we can observe them demonstrate their knowledge and skills. If well-designed and well-implemented, tests can provide samples of performance that reflect the underlying constructs that are our real targets in education.

In science, for example, constructs that an educator might measure through assessment include students' ability to model energy transfer in chemical reactions, analyze mathematical representations to explain phenomena or describe complex patterns of the changes and the movement of water and what determines those patterns.

Defining Terms: Validity

Validity, by definition, is the quality of being logically or factually sound; soundness or cogency

A test elicits a sample of a student's knowledge and skills.

What a score on the test means depends on how well the test was designed and administered and on the quality of how students' responses were scored, analyzed, reported, and used.

Validity evidence

(Oxford English Dictionary, 2018)

Validity is the cornerstone concept in testing, and its meaning in this area is essentially the same as what is found in the dictionary:

The quality of being logically or factually sound; soundness or cogency.

No test can be valid in and of itself. Validity depends on the strength of the evidence regarding what a test measures and how its scores can be interpreted and used.

We create tests to elicit samples of what students know and can do and then make interpretations based on those samples. What a score on the test means depends on how well the test was designed and administered and on the quality of how students' responses were scored, analyzed, and reported.

What may be surprising to some is that assessment validity relates to assessment scores, not to the assessments themselves. Assessment scores may support valid interpretations and uses, but a test itself cannot be inherently valid or invalid.

Principled design is a disciplined approach aimed at designing assessment systems while keeping in mind the inferences end users wish to make based on test scores. If assessment information is expected to have value and usefulness for educators, then early in the design phase, assessment developers must have a clear sense of how the test scores will be used to support inferences about teaching and learning. Principled design is an approach to constructing assessments that ensures the evidence, and interpretations of evidence from the assessment, align with and support the intended student learning outcomes, purposes, and uses of the assessment.

Using a principled-design approach for the assessment development process helps ensure that the assessment yields valid and reliable results. Because this approach to development centers around the assessment purpose and intended use of scores, alignment and accessibility are built into the assessment from the beginning. Principled design focuses on the end goal and uses for the assessment and supports educators in designing tasks that are accessible and engaging to a wide range of students.

A principled approach helps link your assessment to your curriculum and instruction, which is important. Having a process of thinking about what you want to know about students can help build coherence. It is important to focus on what it is you really want to get out of the classroom assessments and what tools you would use.

The goal is to think critically about the domain and what you're trying to measure and then work your way through the task. This helps you think about the outcomes and what you are trying to measure. This helps you not only think about what you want to assess, but also what you want to teach. This process can benefit both assessment and instruction.

Utilizing a principled-design process begins with a clear intent or purpose for assessing. This process helps educators clearly define the purpose for the assessment, the construct, or the knowledge, skills, and other attributes we aim to measure, as well as when the assessments can be administered along the instructional sequence.

Another benefit to principled design is that we consider the behaviors or performances that will elicit evidence to reveal those constructs, which contributes to the meaning and validity of assessment scores. What is it during the teaching and learning experience that signals the need for more "evidence" of students' learning and what would that evidence look like? We must define the "what," the "why," the "when," and the "how" for assessment—what standard is being taught, why student learning needs to be measured, when to assess student learning during the instructional sequence, what evidence needs to be gathered and interpreted, and how the results will be used.

Further, the process provides flexibility. Tasks can be intentionally designed to measure students' knowledge and skills in ways that promote accessibility and engagement and to address a wide range of learner characteristics, needs, and understandings. When we consider how tasks should be designed, we must carefully select the features of the assessment tasks that would be necessary to appropriately measure one or more of the learning outcomes. The task features can influence the design of the task in terms of structure, content, and complexity.

Finally, teachers can use and reuse a common set of design tools to create a variety of tasks for a variety of purposes. The design templates provide a menu of options from which educators can choose to design their tasks. By selecting appropriate task features and work products, educators can create sets of tasks that vary in focus and complexity, and that can serve a range of purposes

and uses along the instructional pathway as students build their understanding of a topic over time.

In Module 1.3, we will begin to explore the four phases of principled assessment design and discuss the tools, templates, and resources associated with each phase.

Defining Terms: Task Features and Work Products

Task features are features of assessment tasks that would be necessary to appropriately measure what students have learned. The task features can influence the design of the task in terms of structure, content, and complexity.

Work products are the vehicles (e.g., a model, an argument, a description, a graph, a chart) that contain observable evidence of what students say, do, or make to produce evidence of their learning.

As task designers, we make a number of decisions based on what we understand about the standard or indicator, and what we know about the evidence that must be produced by students to demonstrate their learning of that standard or indicator. We must consider the key elements of a task that are going to promote student engagement, ensure accessibility, and provide interpretable evidence. Here we highlight and define two of these key elements: task features and work products.

Task features are features of assessment tasks that would be necessary to appropriately measure what students have learned. The task features can influence the design of the task in terms of structure, content, and complexity.

Work products are the vehicles (e.g., a model, an argument, a description, a graph, a chart) that contain observable evidence of what students say, do, or make to produce evidence of their learning.

Teaching to Assessing – A Closing Activity

_____
[Title]

During the school year, I teach my _____ students about _____
[grade level]   [science topic]

_____. Often, my students have challenges with_____

_____.
[misunderstanding, misconception, lack of knowledge]

I want to understand how well my students _____.
[purpose of assessment; knowledge, skills, abilities to measure]

I could ask students to complete _____ or
[work product(s)]

_____to gain evidence about what they know.

I would administer this assessment_____
[e.g., before, during, or after an instructional unit]

To ensure that all of my students can show what they know and can do, I can design the task to include

_____.
[task features]

Based on the evaluation of the evidence, if shows that students lack understanding of the topic, I could adjust my

instruction by _____.
[adaptation, modification, different method or materials]

edCount

In this closing activity, consider yourself a writer, recording on paper an example of a teaching, learning, and assessment experience you might encounter as an educator, which would enable you to get your finger on the pulse of your students' learning along an instruction pathway.

Think about a science topic you teach. Are there specific challenges that students typically encounter when exploring the topic? Are those challenges due to the complexity of the topic, or a misconception or a gap in prior knowledge or understanding?

Then, consider how you might identify and address those challenges throughout the course of your instructional sequence. What knowledge, skills, and abilities would you want to measure? What work products—what students actually say, do, or make to produce evidence—would provide the necessary evidence for evaluating student learning? How might you design your assessment tasks with appropriate task features to promote accessibility and engagement for your students? Perhaps it is through the selection of an engaging and culturally-relevant phenomenon or design problem, or by ensuring that all terminology and science concepts are grade-appropriate and presented clearly and accurately; or perhaps it is by offering students multiple ways to show what they know and can do, whether it is through a written response, oral presentation, or demonstration.
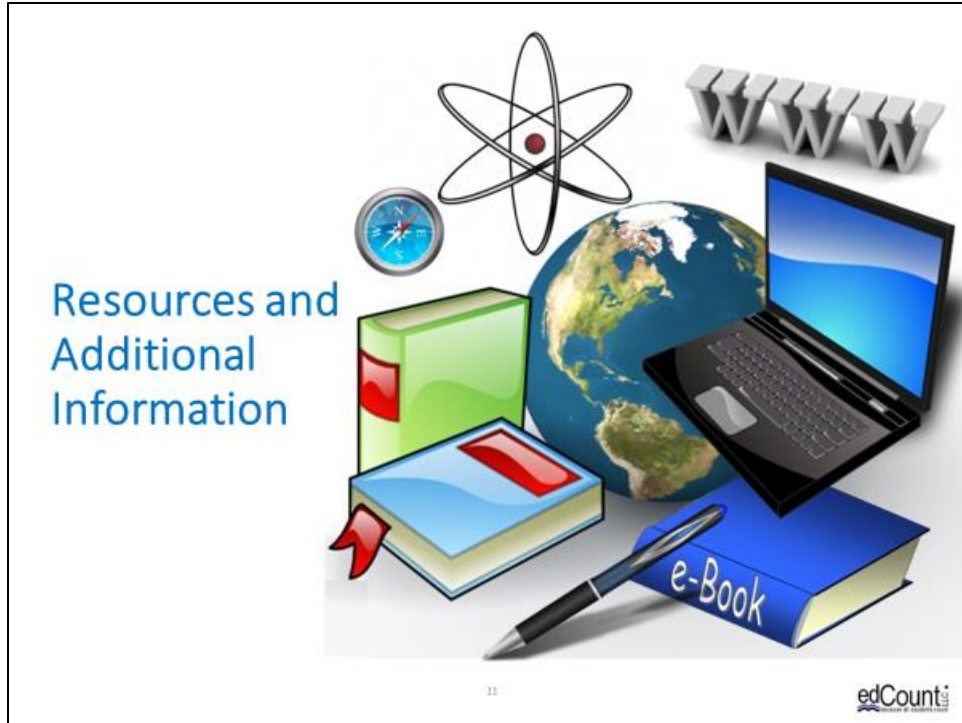
Finally, we offer additional resources that may be helpful to anyone interested in learning more about the concepts presented in this module. A glossary of terms and our reference list follow.

Thank you for your engagement in this first 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.

Web Links

In the web links pod, you can find the following resources:

- A Framework for K-12 Science Education
- Next Generation Science Standards

edCount

References

Oxford English Dictionary. (n.d.). Validity. In *oed.com dictionary*. Retrieved January 22, 2018.