Strengthening Claims-based Interpretations and Uses of Local and Large-scale Science Assessment Scores (SCILLSS)

Advancing Multidimensional Science Assessment Design for Large-scale and Classroom Use

2020 NCSA Annual Conference

June 2020

SCILLSS Project Overview

About SCILLSS

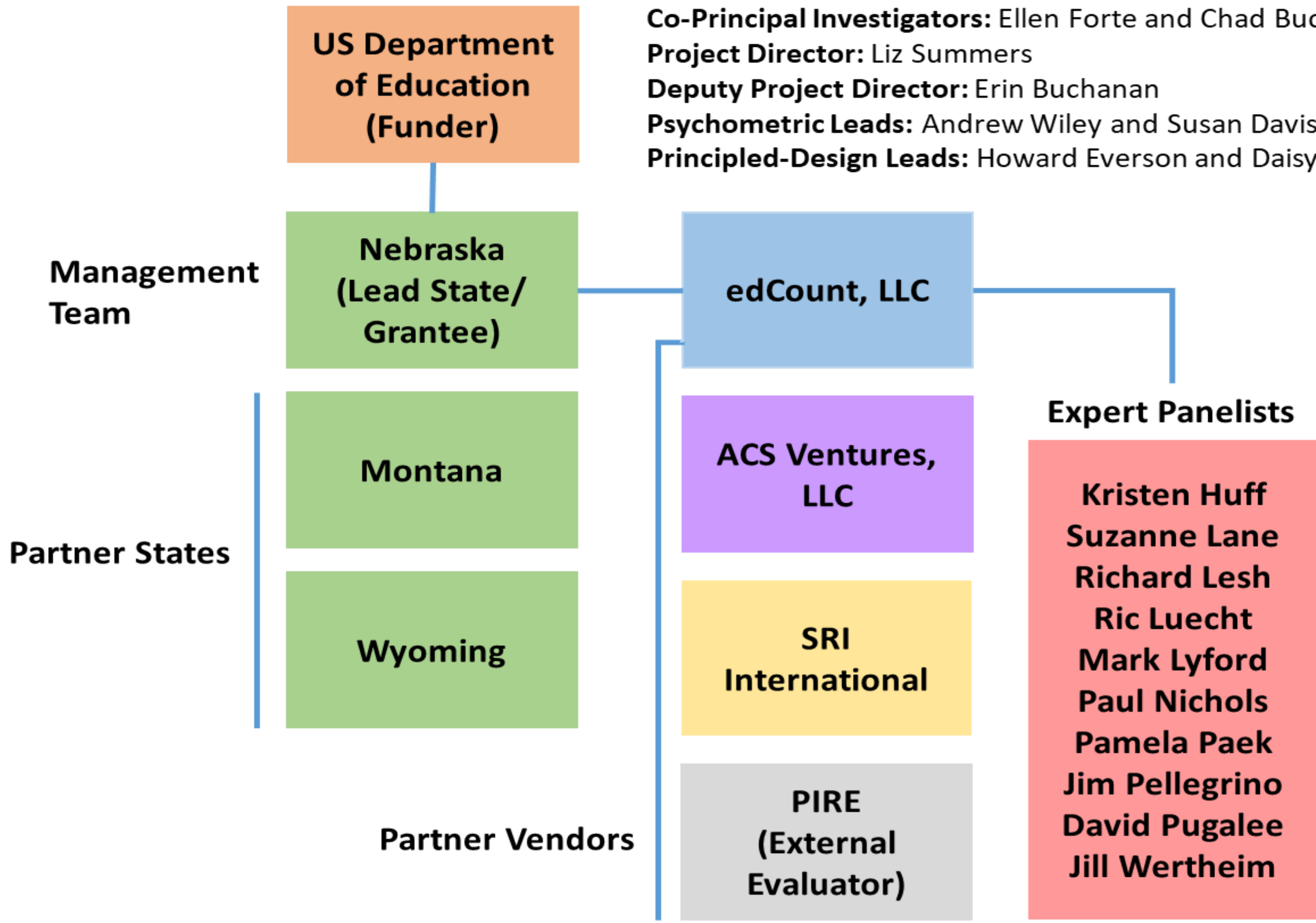
- **S**trengthening **C**laims-based **I**nterpretations and Uses of **L**ocal and **L**arge-scale **S**cience Assessment **S**cores
- One of two projects funded by the US Department of Education's Enhanced Assessment Instruments Grant Program (EAG), announced in December 2016
- Four-year timeline (April 2017 – December 2020)
- Collaborative partnership including three states, four organizations, and 10 expert panel members
- Nebraska is the grantee and lead state; Montana and Wyoming are partner states

SCILLSS Project Goals

- Create a science assessment design model that **establishes alignment with three-dimensional science standards by eliciting common construct definitions** that drive curriculum, instruction, and assessment
- Strengthen a shared knowledge base among instruction and assessment stakeholders for using **principled-design approaches** to create and evaluate science assessments that **generate meaningful and useful scores**
- Establish a means for state and local educators to **connect statewide assessment results with local assessments and instruction** in a coherent, standards-based system

SCILLSS Partner States, Organizations, and Staff

Co-Principal Investigators: Ellen Forte and Chad Buckendahl
Project Director: Liz Summers
Deputy Project Director: Erin Buchanan
Psychometric Leads: Andrew Wiley and Susan Davis-Becker
Principled-Design Leads: Howard Everson and Daisy Rutstein



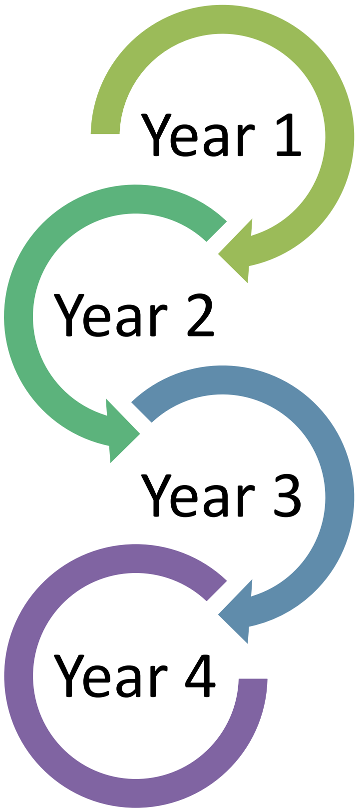
Project Deliverables

2 - Large-scale assessment resources

- Three sets of claim-specific resources:
  - PLD white paper
  - measurement targets, task models, and design patterns
  - sample items
- Assessment literacy modules 2-5

4 - Reporting and Dissemination

- Database of student artifacts corresponding to the performance levels
- Post-project survey
- Post-project action plans for each state
- Final project report



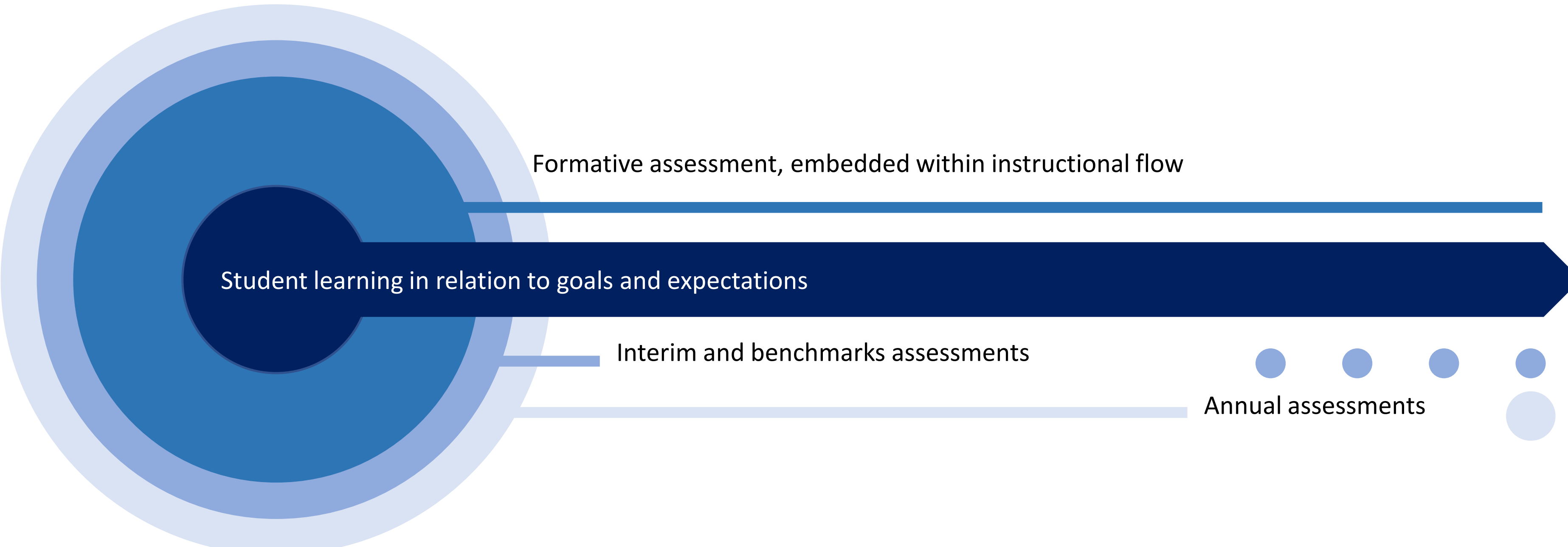
1 - Project Foundations

- SCILLSS website
- Theory of Action for the project and for each state
- Local and state needs assessment tools
- Assessment literacy module 1
- Three prioritized science claims

3 - Classroom-based assessment resources

- Six task models
- Six tasks
- Six sets of student artifacts

Assessment of Student Learning



SCILLSS Resources and Student Learning

SCILLSS Goal 1, Coherence: Establish a means for states to strengthen the meaning of statewide science assessment results and to connect those results with local science curriculum, instruction, and assessment

SCILLSS Goal 2, Support Implementation of Principled-Design: Strengthen the knowledge base and experience among stakeholders in using principled-design approaches to create and evaluate quality science assessments that generate meaningful and useful scores

Student learning in relation to goals and expectations

Interim and other classroom assessments

Annual assessments

SCILLSS resources

A Principled-Design Approach to Creating PLDs and Building Score Scales

Purpose: To explain how and why to develop PLDs and score scales using a principled-design approach

Audience: State and local educators; vendors

Format: White paper

Guide to Developing Three-Dimensional Science Tasks for Large-Scale Assessments

Purpose: To guide implementation of principled-approaches for developing three-dimensional tasks aligned to NGSS-like standards for large-scale science assessments

Audience: State administrators; vendors

Format: Guidebook; templates; tasks; exemplars

Guide to Developing Three-Dimensional Science Tasks for Classroom Assessments

Purpose: To guide implementation of principled-approaches for developing three-dimensional tasks aligned to NGSS-like standards for use within classrooms

Audience: Local educators and administrators

Format: Guidebook; templates; tasks; exemplars

Professional Learning Sessions on Using a Principled Approach to Designing Classroom Assessment Tasks

Purpose: To support local educators in applying principled-design in the development of classroom assessment tasks that link to curriculum and instruction

Audience: Local educators and administrators

Format: Workbook; templates; PPT slides; guiding questions

Assessment Fundamentals

Self-Evaluation Protocols

Purpose: To support educators in evaluating the quality of the assessments in their assessment systems

Audience: State and local educators; vendors

Format: Protocol

Assessment Literacy Workbook

Purpose: To strengthen educators' understanding of and ability to make good decisions about assessments

Audience: State and local educators; vendors

Format: Digital workbook

Theory of Action Principles

- Assessment systems are developed such that they can inform improvements to curriculum and instruction
- Assessments are equitable, accessible, and culturally relevant for widest range of students
- Educators use student performance data appropriately to monitor progress toward CCR and to inform teaching
- State assessments connect coherently to local C-I-A in a way that provides comprehensive coverage of the standards
- Stakeholders collaborate to effectively coordinate alignment of C-I-A systems

Strengthening Claims-based Interpretations and Uses of Local and Large-scale Science Assessment Scores

Home | Policy | NGSS | Principled-Design | Universal Design | SCILLSS Resources

- SCILLSS Resources
- Validity Resources
- NGSS Resources
- Assessment Resources
- SCILLSS Contacts

SCILLSS Resources

The SCILLSS project is producing a wide range of resources for public access and use.

These resources connect visitors with relevant research, instruments, and theoretical frameworks to help inform further work in this field, during and beyond the life of the SCILLSS project period.

Quarterly Newsletters

Year 1

- Quarter 1 (April-May 2017) click here
- Quarter 2 (June-September 2017) click here
- Quarter 3 (October-December 2017) click here
- Quarter 4 (January-March 2018) click here

Year 2

- Quarter 1 (April-June 2018) click here
- Quarter 2 (July-September 2018) click here
- Quarter 3 (October-December 2018) click here

State Assessment Resources

Ensuring Rigor in State Assessment Systems: A Self-Evaluation Protocol click here

SCILLSS Resources

www.scillsspartners.org/scillss-resources/

Overview of the SCILLSS Principled-design Approach

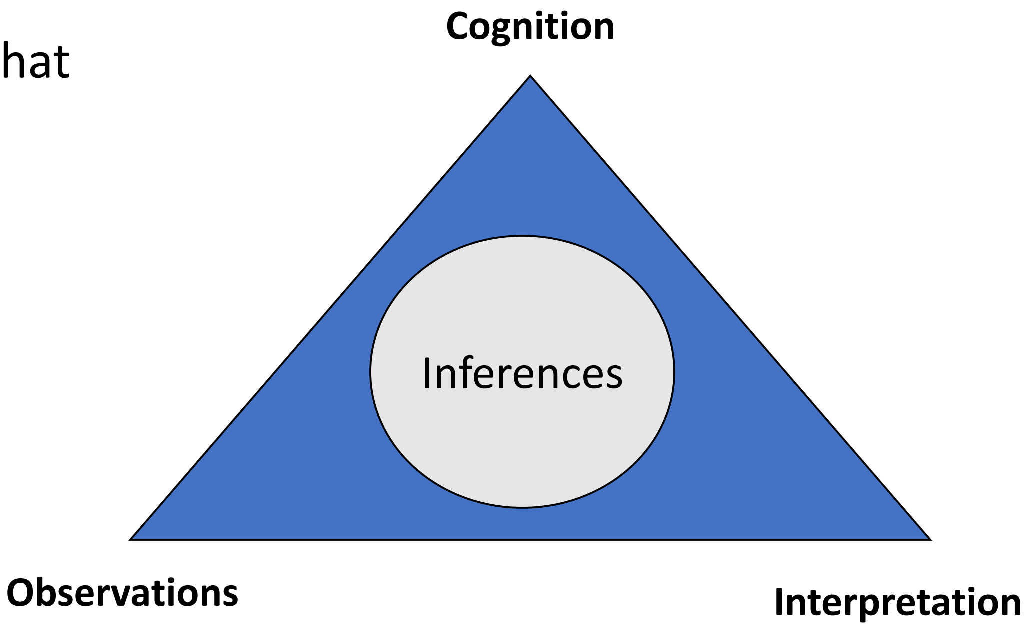
Assessment: A Process of Reasoning from Evidence

Cognition-model of how students represent knowledge

Observations-tasks or situations that allow us to observe students' performance

Interpretation-method of making sense of the data

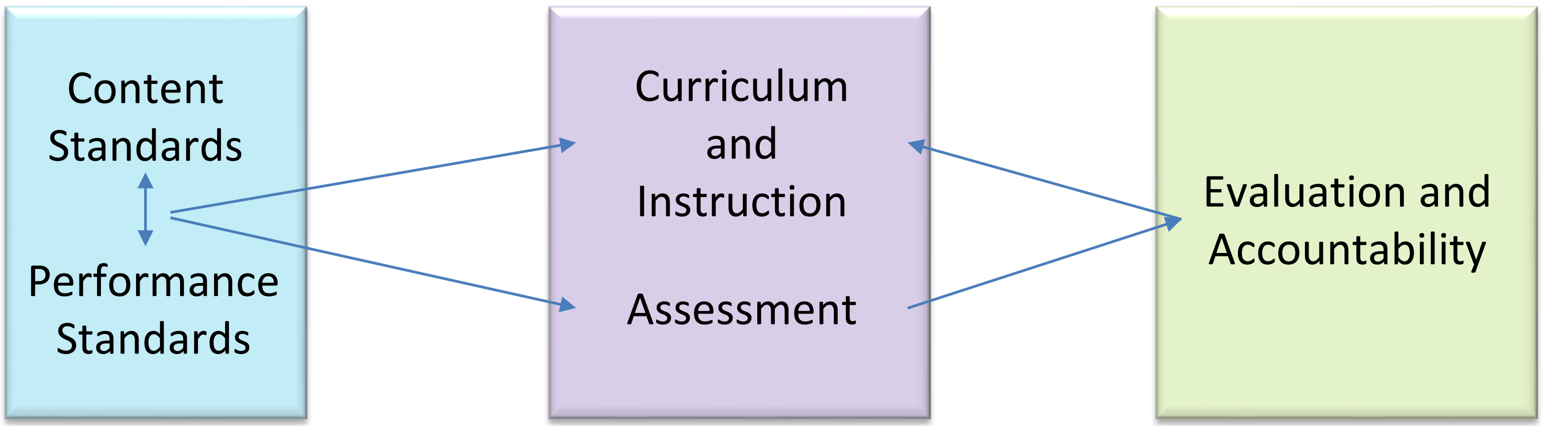
Inference-judging what students know and can do



Coherence is Key

If	Then scores may reflect…	And may be used…
Constructs are well-defined	what students know and can do	to build and deliver instruction aligned with academic expectations
Construct definitions are shared across the system	or	to monitor or track student progress
The system is well-designed	what students have learned this year/ in this course;	for school accountability decisions and program evaluation
The system is well-implemented		

Standards-Based Assessment and Accountability Model



- Standards define expectations for student learning
- Curricula and assessments are interpretations of the standards
- Evaluation and accountability rely on the meaning of scores
- Without clear alignment among standards, curricula, and assessment the model falls apart

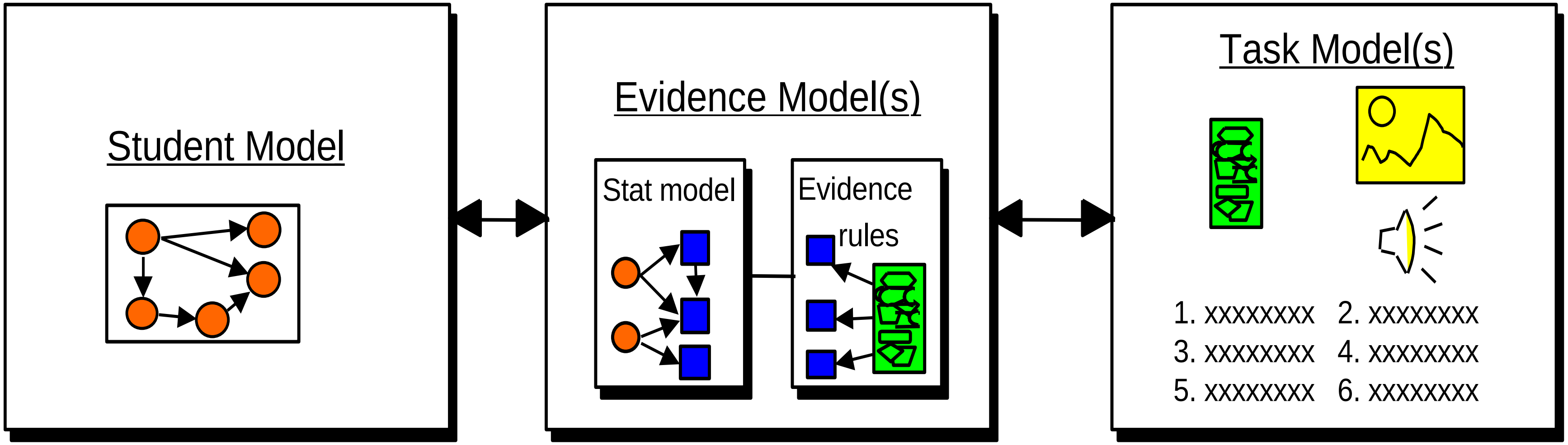
Principled-design Development Purpose

- What is a principled-design development process?
  - Guide to the development of a task that focuses the developer on the purpose of the assessment and the information required in order to design tasks that meet this purpose
- Why do we use a principled-design development process?
  - Highlight the design decisions that need to be made in the process in order to develop tasks with valid and reliable inferences
  - Articulate a **replicable and scalable design process** that states and other organizations can use to develop state summative and classroom-embedded three-dimensional science assessments
- How have we used this process?
  - Developed a **sample set of exemplary resources** to demonstrate the outcomes of the process for the development of state summative and classroom-embedded three-dimensional science assessments

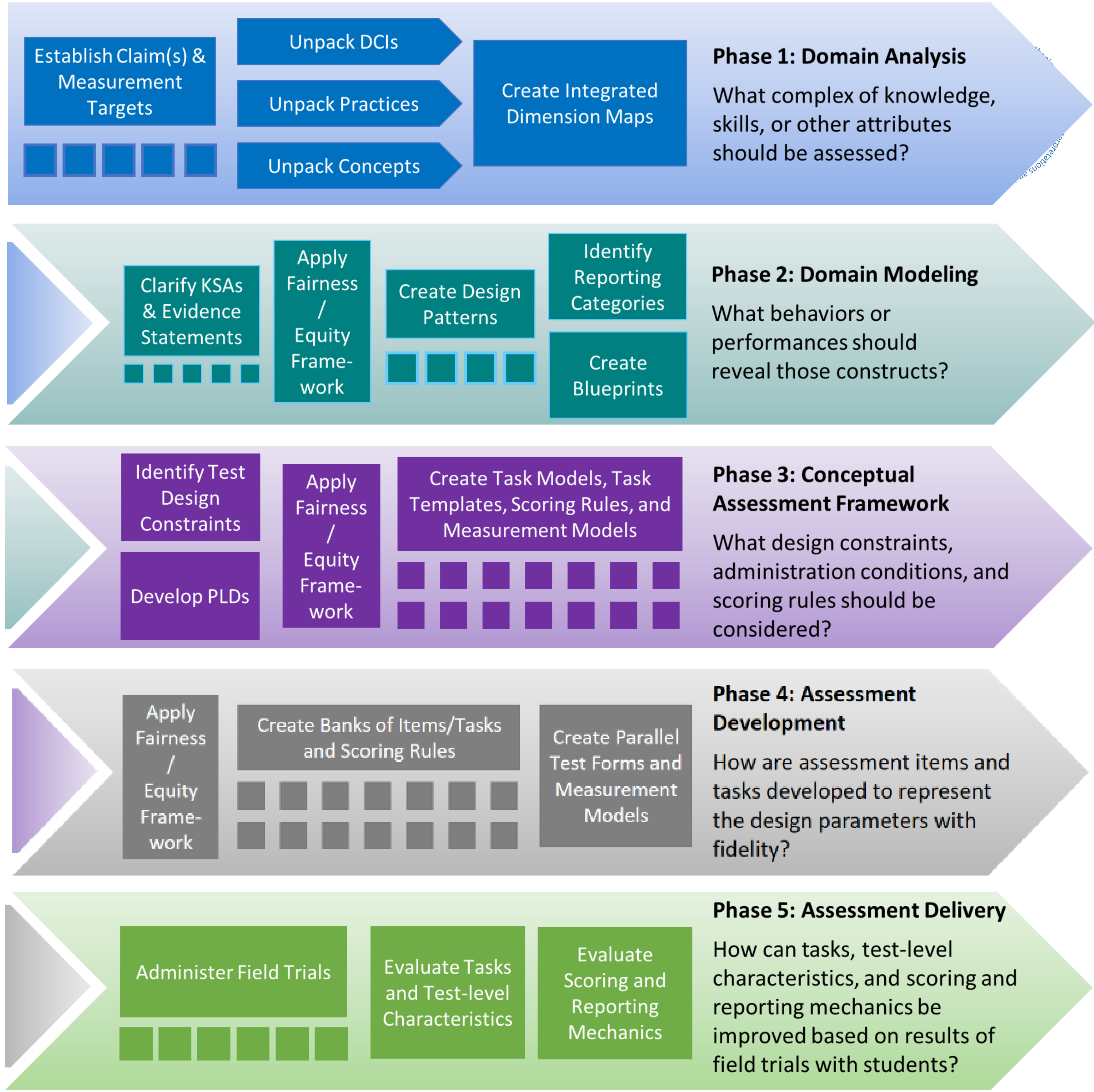
Evidence-Centered Design (ECD)

Formal, multiple-layered framework for assessment development based on Messick's (1994) guiding questions:

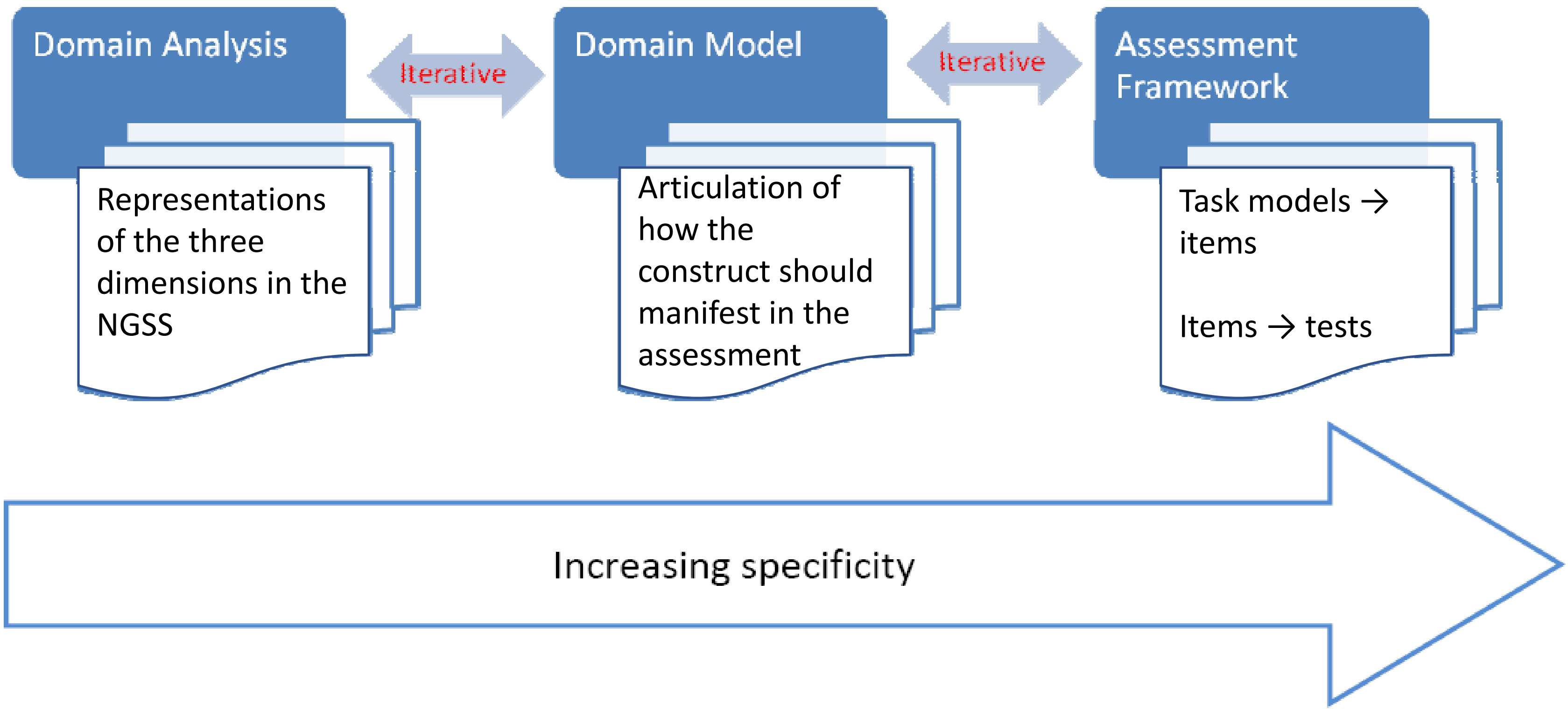
- What complex of knowledge, skills, or other attributes should be assessed?
- What behaviors or performances should reveal those constructs?
- What tasks or situations should elicit those behaviors?



Iterative Five-phase Principled-design Process



Three Critical Design Phases



Adapted from Huff, Steinberg, & Matts, 2010

Principled-design Resources for State and Local Use

- A Principled Approach to Designing **State** Three-Dimensional Science Assessment Tasks: A Process Guide
- A Guide to Develop **Classroom-based** Next Generation Science Standards Assessment Tasks: A Principled-design Approach

www.scillsspartners.org/scillss-resources/

Provide a **principled-design model and set of replicable tools,** which can be scaled to address the unique characteristics and contexts of states' assessment systems, with a particular focus on **establishing coherence among state summative assessments and classroom-based assessments** designed to be administered at a time that fits the instructional sequence in the classroom.

Benefits of Principled-Design for Large-Scale Assessment

- Principled articulation and alignment of design components
- Articulation of a clear interpretation and use argument and population of a strong validity argument
- Reuse of extensive libraries of design templates
- For accountability
  - Clear warrants for claims about what students know and can do
  - Build accessibility into design of tasks (not retrofitted into tasks)
  - Cost versus scale

Benefits of Principled Design for Classroom-Based Assessment

- Highlights the intended outcomes of classroom-based assessment
- Points to the connections among curriculum, instruction, and assessment, which are linked in a coherent system
- Provides tools to accomplish the development of classroom-based assessment tasks and rubrics

Developing State Summative and Classroom-based Science Assessment Tasks Using a Principled-design Approach

Objectives

Establish a shared understanding of:

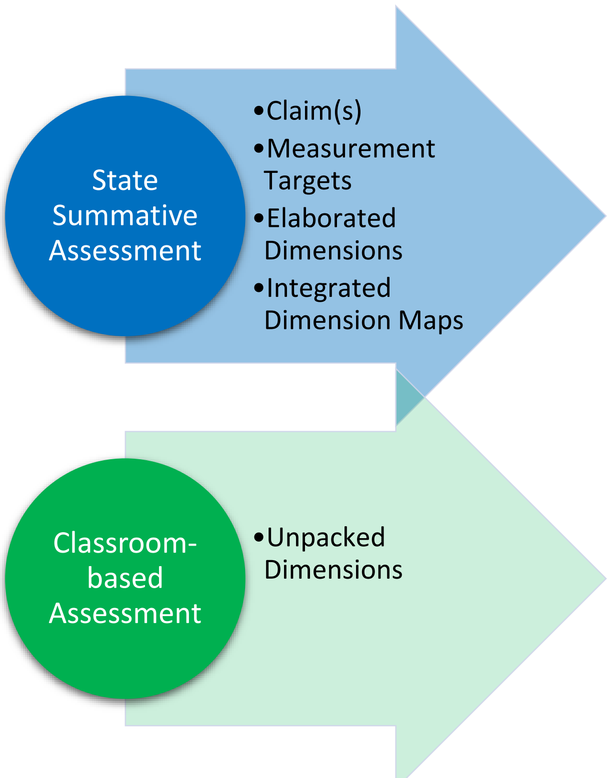
- Principled-design and its application in the context of SCILLSS; and
- The SCILLSS principled-design state summative and classroom-based assessment resources.

Principled-design State Summative and Classroom Phases and Elements

Phase	State Summative Elements (Grades 5, 8, 11)	Classroom Elements (Grades 5, 8, 11)
Domain Analysis	• Overall Claim • Measurement Targets • Elaborated (Unpacked) Dimensions • Integrated Dimension Maps	• Unpacking Tool
Domain Modeling	• Design Patterns	• Task Specifications Tool
Task Conceptualization	• Task Templates • Task Specifications • Item Specifications	
Assessment Development	• Tasks • Scoring Rubrics/Scoring Notes	• Tasks • Scoring Rubrics/Scoring Notes

Phase 1: Domain Analysis

Resources:



Goal:

- To obtain a deep understanding of the performance expectation (PE) and its components
- To provide information on how students engage with the different components
- To provide information on the boundaries of student performance

Claim

- Relates to expected student learning
- Represents and supports an assessment argument
- Links to forms of evidence
- Explores the question:

“What *warrants* the claim?”

State Summative Assessment

Measurement Targets

- Statements that provide descriptions of the performance defined in the claim
- Measurement targets are grade- and bundle-specific.
- Contribute to consistent learning targets, coherent results, consistent judgments of competence, and curriculum, instruction, and assessment alignment
- For SCILLSS, the *NGSS Example Bundles* were utilized as a way of organizing the standards for the development of the measurement targets.

State Summative Assessment

Elaborated Dimensions

- Elaboration of the NGSS dimensions is completed during domain analysis in which:
  - Substantive information is gathered about the domain of interest that will have implications for assessment; and
  - The construction of learning performances are informed to describe the knowledge that students need to demonstrate as they progress toward achieving the measurement target expectations.
- Elaborations articulate clear expectations, appropriate assessment boundaries, required background knowledge, and student challenges and misconceptions.

State Summative Assessment

Integrated Dimension Maps

- Integrated Dimension Maps are visual representations of the DCIs, SEPs, and CCCs.
  - Highlight how the different dimensions are integrated with each other
  - Highlight what pieces should be assessed together, and what pieces can be assessed separately

State Summative Assessment

Unpacking the Dimensions

- Provides a clear focus for what is to be measured and helps educators to plan for assessment
- Ensures educators who are designing NGSS-aligned tasks have a clear and deep understanding of each of the dimensions represented in a PE prior to beginning task development

Classroom-based Assessment

Unpacking the Dimensions of a Performance Expectation Tool

- Provides guidance for unpacking a PE
- Template for documenting unpacking

Grade:
NGSS Performance Expectation:

	Science and Engineering Practices (SEP)	Disciplinary Core Ideas (DCI)	Crosscutting Concepts (CCC)	
Foundations	**SEP:**	**DCI:**	**CCC:**	
Key Aspects	•	•	•	
Prior Knowledge	•	•	**Relationships to SEPs**	•

Classroom-based Assessment

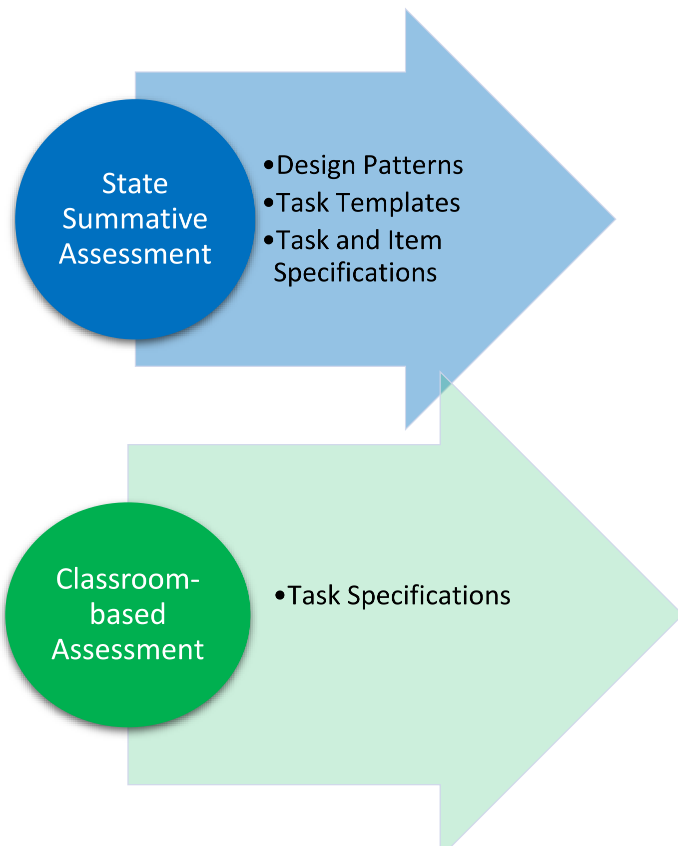
Components of the Unpacking Tool

- *Key aspects* are the underlying concepts that support each dimension of the PE and represent knowledge necessary for understanding or investigating more complex ideas and solving problems.
- *Prior knowledge* refers to the background knowledge that is expected of students to develop an understanding of the SEP and DCI.
- *Relationships between the CCC and the SEP* is included since when students are performing a SEP, they are often addressing one of the CCCs.

Classroom-based Assessment

Phase 2: Domain Modeling and Phase 3: Conceptual Assessment Framework

Resources:



Goal:

– To clearly lay out the assessment argument
  - What will be covered?
  - What will not be covered?
  - How will students demonstrate their knowledge?
  - What do tasks look like?

Design Patterns

- Before developing assessment tasks, a design pattern must be specified (Mislevy & Haertel, 2006) for each learning performance.
- The design patterns serve to complete the documentation of the assessment argument connecting task designs to performance expectations.
- Identify:
  - Focal Knowledge, skills, and abilities (fKSAs)
  - Observations (i.e., evidence) to support inference
  - Features of task situations that elicit target KSAs
- Guide planning for the key elements of the task models in the conceptual assessment framework

State Summative Assessment

Task Template

- The task template is a tool to support writing families of tasks that includes specific details of materials and task settings in the assessment implementation phase.
- The contents:
  - Are informed by the ECD framework;
  - Suit the needs and processes of the project; and
  - Therefore, can vary with respect to specificity and detail.
- Allows for multiple items or tasks to be developed based on the template

State Summative Assessment

Task Specifications

- The integration of Phase 2 and Phase 3 provides the rationale for the formulation and content of task specifications.
- The task specifications define for task developers the key components of the task needed to ensure that the evidence of student learning collected and evaluated is consistent with the fKSAs represented by the PEs.
- Identifies for a selected fKSA:
  - The decisions needed to be made to elicit evidence of student competency;
  - Variable features that inform design decisions to evoke that evidence;
  - Aspects of the assessment situation that may be varied;
  - The responses or artifacts the students will produce that, subsequently, will be used in the evaluation (scoring) procedures; and
  - The task context (i.e., phenomena, design problems).

State Summative Assessment

Task Specifications

Task Specification Component	Description
Target fKSA[1]	The Focal Knowledge, Skills, and Abilities to which the task is written
Additional Knowledge, Skills, and Abilities (aKSAs)[1]	Other knowledge/skills/abilities that may be required by tasks designed to measure the fKSAs
Potential Observations[1]	Aspects of the work product that would reflect on students' fKSAs
Characteristic Features[1]	Aspects of the assessment situation that are needed to evoke the desired evidence
Variable Features[1]	Aspects of the assessment situation that may be varied (often to shift the difficulty or the focus of the task)
Task Model[2]	Description of the environment in which examinees will say, do, or make something, to provide the data or evidence about what they know or can do as broadly conceived
Task Model Variables[2]	Variables for features of tasks (e.g., reading level, use of graphics, symbols, equations, etc.) that indicate the design decisions needed with regard to specific tasks and items
Work Product Summary[2]	Description of the responses or artifacts the students will produce that, subsequently, will be used in the evaluation (scoring) procedures
Evidence of High Level of Performance[3]	Defines behaviors that you would expect to see if a high-performance student was engaging with the SEP or CCC
Task Context[2]	The possible types of phenomena or design problems that will be represented in the task
Assessment Boundary[3]	Grade specific boundary conditions of the task
Universal Test Design Considerations	Strategies implemented to maximize accessibility and fairness

State Summative Assessment

Item Specifications

- Utilization of the task and item specifications leads to a determination of the item-response formats required to elicit necessary evidence of student competency of the targeted fKSA.
- The item specifications provide information to create an item(s) that will provide some of the necessary evidence with respect to a selected fKSA
- Identifies:
  - A rationale of what the student will do to demonstrate competency of a targeted fKSA;
  - Construct-relevant vocabulary;
  - Allowable stimulus materials (e.g., data tables, animation), item type, and "model" stem; and
  - The nature of the response options (e.g., Distractors may include…).

State Summative Assessment

Item Specifications

Item Specification Component	Description
Target PE	**5-PS1-2** Measure and graph quantities to provide evidence that regardless of the type of change that occurs when heating, cooling, or mixing substances, the total weight of matter is conserved.
fKSA	**5.1b** Students are able to investigate or create an explanation around conservation of matter using measurements when substances are mixed, or undergo a change in form, properties, or state
Rationale:	Students will describe that the total weights of the substances did not change, regardless of the reaction or changes in properties that were observed.
Construct-relevant Vocabulary	conclude, data, experiment, hypothesis, investigation, model, predict, solution, variable, matter
Allowable Stimulus Material	animation, data tables, graphics, graphs, simulation, text
Item Type:	SR (multiple-choice and multi-select)
Model Stem	Describe how the measurements or graph serve as evidence to support a statement/conclusion about conservation of matter.
Correct Answer:	Part a - D; Part b – A, D
Response Options	Distractors may include graphs or measurements that would show variation in weight or volume and predictions that do not reflect the conservation of matter.
Item Notes/Reference Source:	• In all cases, the unit grams will be used. • Whenever the term “grams” is used, the term “amount” is preferred instead of “weight.” However, when clarity is needed, the term “weight” will be used. • Although students are not to be assessed on the term “closed system,” examples of closed systems (jar covered with lid, etc.) should be a part of the stimulus.

State Summative Assessment

Identifying Assessment Task Specifications

- Allows educators to translate the PE-specific unpacking of the three dimensions into assessment tasks
- Allows educators to determine what counts as evidence for student learning
- Helps educators develop assessment tasks that allow students opportunities to call upon, transfer, and apply learning that has occurred during instruction to new challenges, much the way a scientist or engineer would, in an assessment situation

Classroom-based Assessment

Assessment Task Specifications Tool

- Identifies key elements needed to be addressed by task developers to develop meaningful and interpretable assessment tasks
- Template for documenting task specifications

Classroom-based Assessment

Assessment Task Specifications Tool

Element	Description
Performance Expectation	• Indicate the PE from the instructional sequence to be assessed.
Knowledge, Skills & Abilities (KSAs)	• Develop statements, which specify what is expected of students to demonstrate (i.e., knowledge, skills, and abilities) to provide evidence that they have learned one or more aspects of a PE.
Student Demonstration of Learning	• List what students should be able to do to demonstrate that they have met the KSA(s). • Define qualities of student performance that constitute student evidence.
Work Product	• Determine the "vehicles" (i.e., work products) that are intended to contain observable evidence (e.g., a model, an argument, a description, a graph, a chart).

Classroom-based Assessment

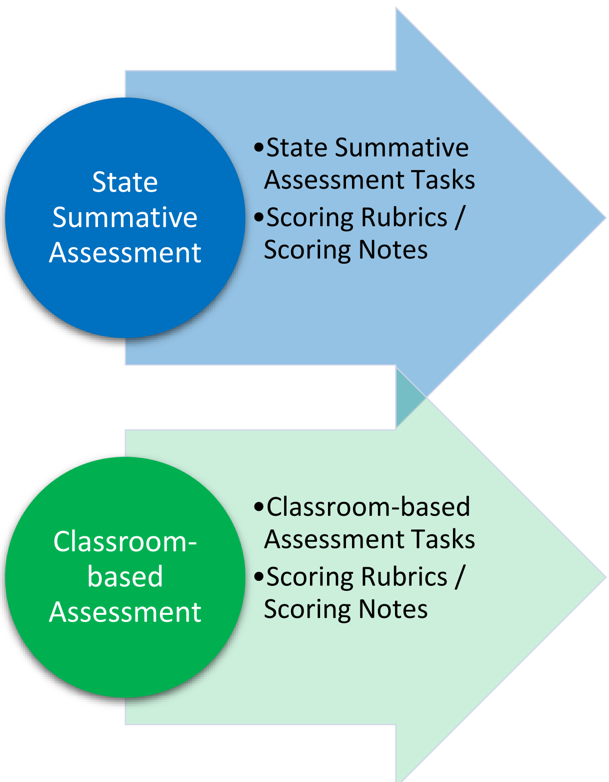
Assessment Task Specifications Tool

Task Features	• List the task features from which the task writer selects to develop an assessment task. • Reference the "Clarification Statement" in the NGSS for the PE as appropriate. • Note: A single question/task may not represent all the features listed.
Aspects of an assessment task that <u>can be varied</u> to shift complexity or focus	• Allows for a range of tasks to be developed of varying complexity. • Allows for development of tasks that focus on various skills related to the PE. • Allows the task developer to match features of the task with the characteristics of students such as their interests, familiarity, and provided instruction.
Assessment Boundaries	• List information that is NOT assessed (i.e., related above grade-level ideas and skills). • Reference the "Assessment Boundary" in the NGSS for the PE as appropriate.

Classroom-based Assessment

Phase 4: Assessment Development

Resources:



Goal:

– To develop tasks and rubrics that are aligned to the assessment argument

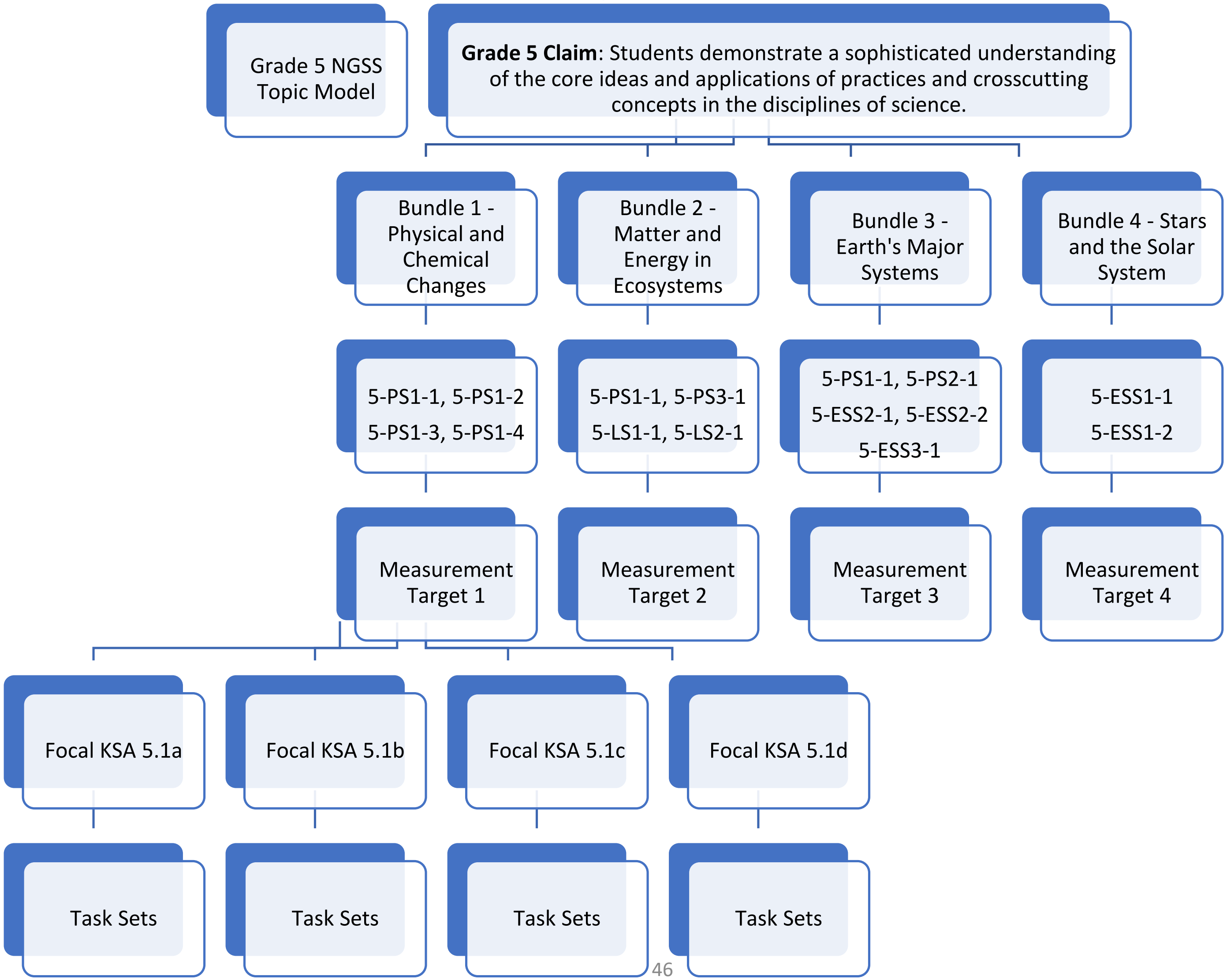
– To describe the evidence of student learning to be elicited by the tasks

State Summative Assessment Tasks

- A SCILLSS task is envisioned as a set of three or more items of varying types linked with a common stimulus.
- A task stimulus consists of passages, graphs, models, figures, diagrams, data tables, etc.
- The number of items associated with a task is dependent on the number and nature of the fKSAs and PEs it is written to measure.
- The number of dimensions addressed by each item is also variable.
- Tasks are designed to assess students along a range of proficiency and across an appropriate range of cognitive complexity.

State Summative Assessment

Grade 5 SCILLSS Conceptual Assessment Framework Hierarchy



State Summative Assessment

Sample Task Map

Focal Knowledge, Skills, and Abilities (FKSAs)

- ○ 5.1a
- ✖ 5.1b
- ○ 5.1c
- ○ 5.1d

Performance Expectation

- ○ 5-PS1-1
- ✖ 5-PS1-2
- ○ 5-PS1-3
- ○ 5-PS1-4

Task Context and Items

Phenomena related to the conservation and particulate nature of matter

Task

Task Context and Items	Dimension Representation		
Item 1	SEP ○	DCI ○	CCC ○
Item 2	SEP ○	DCI ○	CCC ○
Item 3	SEP ○	DCI ○	CCC ○
Item X	SEP ○	DCI ○	CCC ○

State Summative Assessment

Sample Item and Scoring Notes

- Written for 5-PS1-2 and fKSA 5.1b *(Students are able to investigate or create an explanation around conservation of matter using measurements when substances are mixed, or undergo a change in form, properties, or state)*
- Fits into a task comprised of multiple items and structured around a student investigation related to changing matter by melting, evaporating, and dissolving substances
- Includes a task context and narrative describing the different aspects of the item that elicit knowledge and skills
- Includes suggestions for other items that could be in the tasks, scoring notes, and alignment remark

State Summative Assessment

Classroom-based Assessment Tasks

- Enable educators to get their fingers on the pulse of individual students, groups of students, and/or the entire class as to where they are in their science learning and collect evidence to ultimately inform instruction
- Must elicit evidence related to students' integration of knowledge of DCIs, engagement with SEPs, and facility with building connections across ideas
- Provide an indication of the student's current understanding of the selected KSAs as set forth in the Task Specifications Tool
- May include multiple parts, questions, or prompts connected to a phenomenon or problem-solving context or event

Classroom-based Assessment

Example Classroom-based Task

This task is about the particles of matter. Be sure to answer question 1 and question 2.

Anticipatory set
Reminder to student

1. Jose cleaned his salt water fish tank. The water in the tank looked clear. His friend Carl visits and asks, "Why can't I see the salt in the water?" Jose creates a model to show Carl what happens to salt when stirred into water.

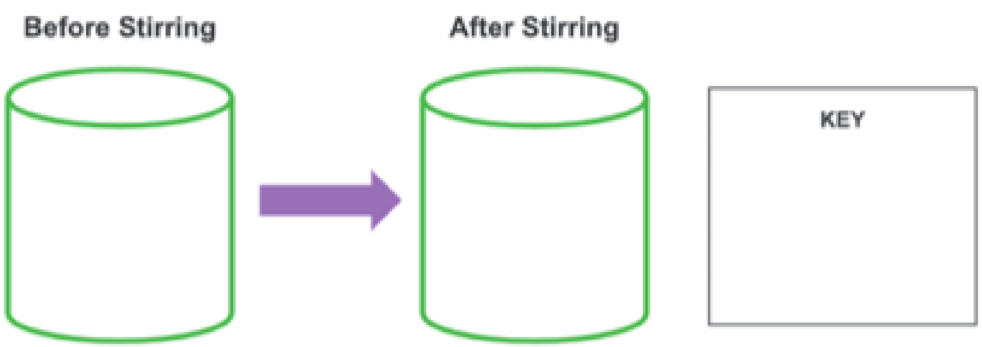
Task context/stimulus

Complete the model below to show:

- the salt particles and water particles **before** stirring the mixture
- the salt particles and water particles **after** stirring the mixture

Be sure to complete the key to show the salt particles and water particles in both conditions of your model.

Prompt/question and student directions



Provided model and key templates

2. Describe the change to the salt particles after being stirred in the water. Be sure to use information from your model to support your explanation.

Prompt/question and student directions

Classroom-based Assessment

Rubrics

- Define the criteria that educators use to interpret and evaluate student evidence of learning
- Include descriptors for each question or prompt in the assessment task that describe the full range of student understanding from low to high levels of competency
- The type of evidence gathered may vary from situation to situation.

Classroom-based Assessment

Example Classroom-based Task Rubric

Student Expectations of Learning	1	2	3
Student uses the model to describe how matter composed of tiny particles is too small to be seen.	Student develops a model that does **not** account for observable phenomena. Description is incorrect or is not provided.	Student develops a model that shows: • a flawed connection between bulk matter and particles too small to be seen (e.g., salt particles are represented at the same size under both conditions; all salt particles are clustered in one area of the container under both conditions; key is incorrect, etc.) Description is partially correct (e.g., does not refer to the scale of the dissolved salt particles).	Student develops a model that shows: • two representations each with two different bulk matter and matter too small to see (particles) • the representations correctly position and scale the particles relative to each other • the key is correct Description is correct.

Rating scale

Student demonstration of knowledge, skill, and/or ability

Evaluative criteria of student's model

Evaluative criteria of student's explanation of the model

Statement of Student Expectation of Learning

Classroom-based Assessment

Student Exemplars

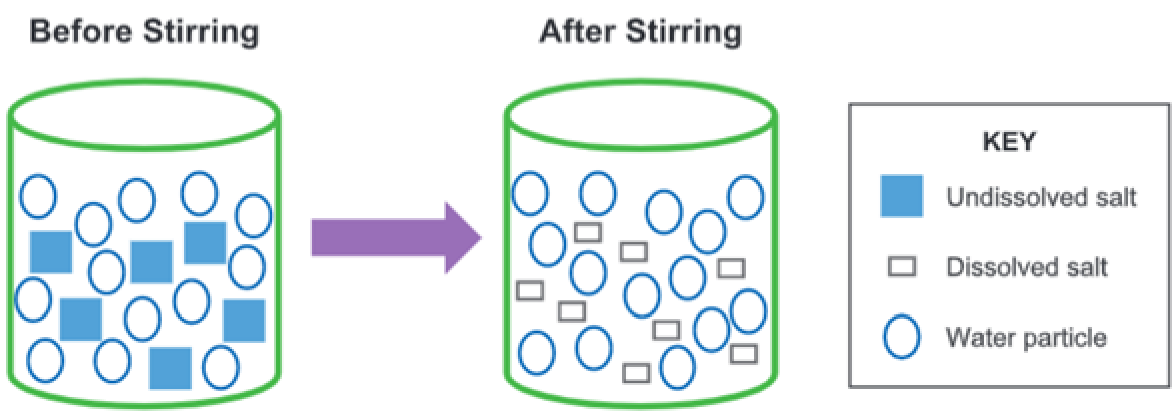
- A high-level response is scientifically accurate, complete and coherent, and consistent with the type of student evidence expected.
- A low-level response may include misconceptions, is incomplete, and is not consistent with the type of evidence expected.
- Student responses should yield accurate inferences about students' KSAs that inform educator actions either to:
  - Continue with the instructional sequence as planned; or
  - Adjust the design, delivery, and sequence of instruction.

Classroom-based Assessment

Student Exemplar: Model and Key

Selects and identifies relevant aspects of the context in the model

Shows relationships between water and dissolved salt particles, which vary in size



Includes relevant components and labels to represent understanding of the system

Classroom-based Assessment

Student Exemplar: Explanation

"The model shows that the salt particles dissolve. They break into smaller pieces after they are stirred into water. The salt particles are still in the water, but you can't see them. That's because they got so small."

Uses the model to describe how matter composed of tiny particles too small to be seen can account for observable phenomena (e.g., salt dissolving into water).

Classroom-based Assessment

Concluding Remarks

- Utilizing our principled-design process, item writers (or classroom teachers) can see the tasks they create as addressing the same underlying science, in terms of common fKSAs, Characteristics Features of Tasks, and Potential Observable features of students' performances.
- Different choices about Additional KSAs, Variable Features of Tasks, and Work Products are required in order to meet the varying constraints and purposes of different assessment contexts.
- Having common and explicit design patterns and task templates enhances the instructional validity of assessment as well as the evidentiary value of tasks.

Developing and Evaluating Classroom Tasks and Student Artifacts with Local Educators

SCILLSS Classroom Science Assessment Development Workshops

- Five workshops total in SCILLSS partner states (NE, MT, WY)
  - June 6-7, 2019: Lincoln, Nebraska
  - July 23-24, 2019: Laramie, Wyoming
  - August 14-15, 2019: Bozeman, Montana
  - January 13, 2020: Helena, Montana
  - January 24-25, 2020: Kearney, Nebraska
- Range of 12 to 30 to educators per session
  - Small development teams (3-5 educators) at grade 5, middle school, and high school
- Educators collaborate to develop classroom-based assessment tasks and rubrics using a principled-design approach and provide valuable feedback about the feasibility, clarity, and utility of the SCILLSS design process and resources and tools.

Workshop Goals

For the project team . . .

- To increase educators' knowledge of a principled-approach for developing three-dimensional tasks aligned to NGSS-like standards for use within classrooms
- To pilot a principled-design process for developing three-dimensional classroom science assessment tasks aligned to the performance expectations in the Next Generation Science Standards
- To pilot a set of professional learning resources that states and districts can use to scale-up the work and build educator capacity to design quality assessment tasks using a principled-design approach
- To gather feedback from educators about:
  - The design process and its utility
  - The clarity and utility of the professional learning tools and resources

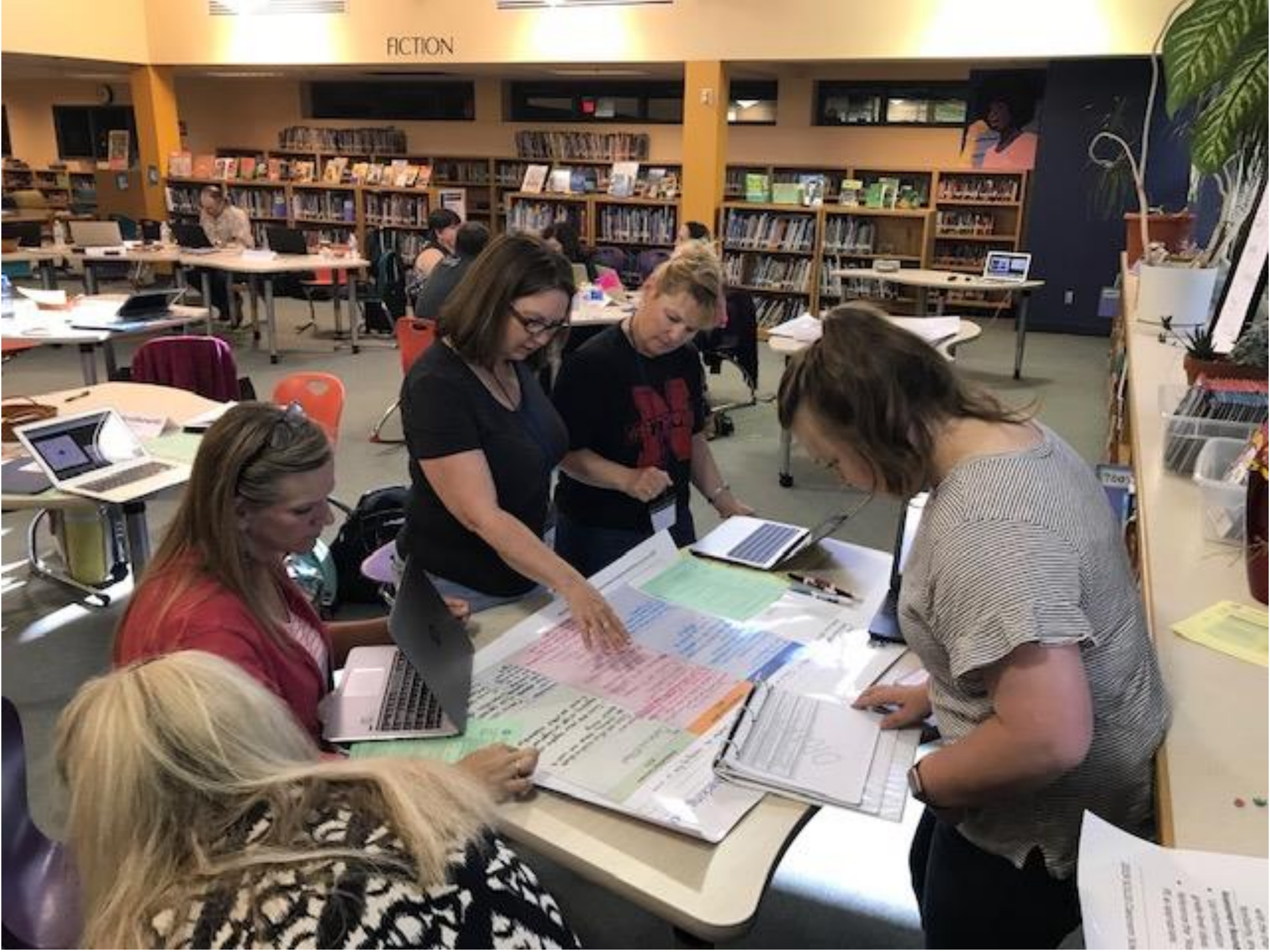
Workshop Goals

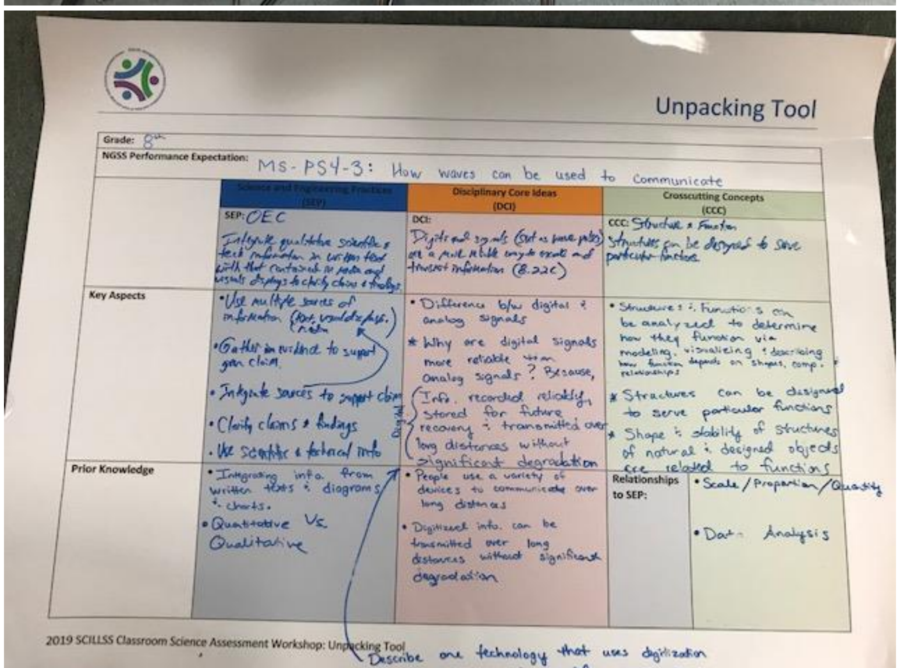
For the participating educators . . .

- To develop an understanding of classroom-based science assessments, their relationship to other forms of assessment, and their purposes and uses in a standards-based system of curriculum, instruction, and assessment
- To develop an understanding of a principled-approach for developing three-dimensional tasks aligned to NGSS-like standards for use within classrooms
- To collaborate to develop a classroom science assessment task and rubric(s) for their assigned grade or domain to support instruction

Workshop Outcomes

- Educators discussed and selected a performance expectation (PE) to work with in their assigned groups
- For their selected PE, the groups:
  - Used a tool to unpack the PE and its corresponding dimensions in order to highlight the important aspects of the PE
  - Used a tool to develop guidance for the development of tasks and rubrics
  - Developed tasks and rubrics that measured the target PE
  - Reviewed and revised these tasks and rubrics
- At the end of this meeting each group produced:
  - A set of design documents
  - A rough draft of a task and scoring rubric(s)


FICTION

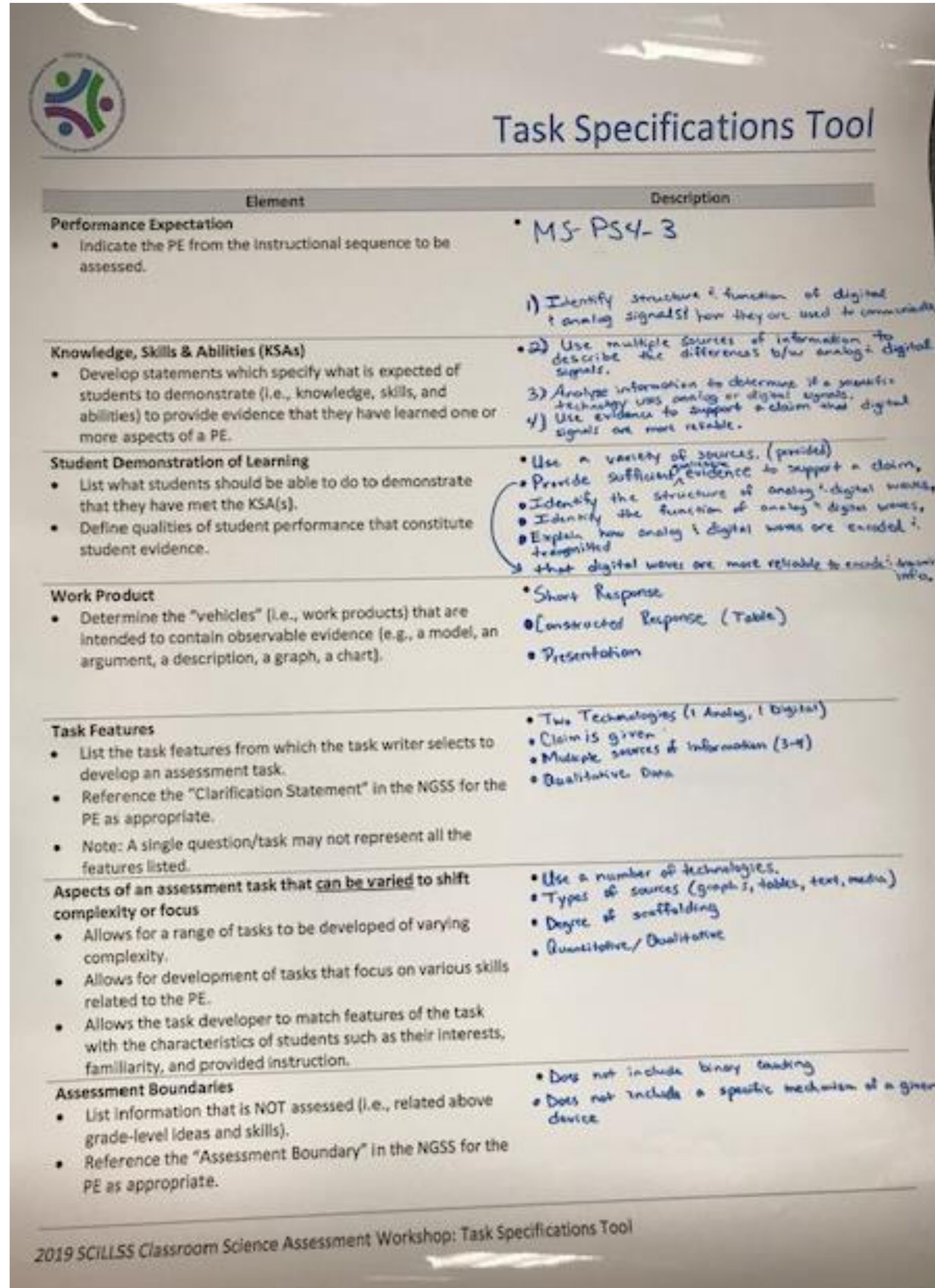


Unpacking Tool

Grade: 8th			
NGSS Performance Expectation: MS-PS4-3: How waves can be used to communicate			
	Science and Engineering Practices (SEP)	**Disciplinary Core Ideas (DCI)**	**Crosscutting Concepts (CCC)**
	SEP: OEC Integrate qualitative scientific & tech information in written text with that contained in media and visuals displays to clarify claims & findings	DCI: Digitized signals (sent as wave pulses) are a more reliable way to encode and transmit information (8.22C)	CCC: Structure & Function Structures can be designed to serve particular functions
Key Aspects	• Use multiple sources of information (text, visuals, graphs, media) • Gather evidence to support given claim • Integrate sources to support claim • Clarify claims & findings • Use scientific & technical info	• Difference b/w digital & analog signals * Why are digital signals more reliable than analog signals? Because, Info. recorded reliably, stored for future recovery & transmitted over long distances without significant degradation	• Structures & Functions can be analyzed to determine how they function via modeling, visualizing & describing how function depends on shapes, comp. & relationships * Structures can be designed to serve particular functions * Shape & stability of structures of natural & designed objects are related to functions
Prior Knowledge	• Integrating info. from written texts & diagrams & charts. • Quantitative Vs. Qualitative	• People use a variety of devices to communicate over long distances • Digitized info. can be transmitted over long distances without significant degradation	**Relationships to SEP:** • Scale / Proportion / Quantity • Data Analysis

2019 SCILLSS Classroom Science Assessment Workshop: Unpacking Tool

Describe one technology that uses digitization to advance science & scientific investigations



Task Specifications Tool

Element	Description
Performance Expectation • Indicate the PE from the instructional sequence to be assessed.	• MS-PS4-3 1) Identify structure & function of digital & analog signals & how they are used to communicate.
Knowledge, Skills & Abilities (KSAs) • Develop statements which specify what is expected of students to demonstrate (i.e., knowledge, skills, and abilities) to provide evidence that they have learned one or more aspects of a PE.	• 2) Use multiple sources of information to describe the differences b/w analog & digital signals. 3) Analyze information to determine if a specific technology uses analog or digital signals. 4) Use evidence to support a claim that digital signals are more reliable.
Student Demonstration of Learning • List what students should be able to do to demonstrate that they have met the KSA(s). • Define qualities of student performance that constitute student evidence.	• Use a variety of sources. (provided) • Provide sufficient evidence to support a claim, that digital waves are more reliable to encode & transmit info. • Identify the structure of analog & digital waves, • Identify the function of analog & digital waves, • Explain how analog & digital waves are encoded & transmitted
Work Product • Determine the "vehicles" (i.e., work products) that are intended to contain observable evidence (e.g., a model, an argument, a description, a graph, a chart).	• Short Response • Constructed Response (Table) • Presentation
Task Features • List the task features from which the task writer selects to develop an assessment task. • Reference the "Clarification Statement" in the NGSS for the PE as appropriate. • Note: A single question/task may not represent all the features listed.	• Two Technologies (1 Analog, 1 Digital) • Claim is given • Multiple sources of information (3-4) • Qualitative Data
Aspects of an assessment task that can be varied to shift complexity or focus • Allows for a range of tasks to be developed of varying complexity. • Allows for development of tasks that focus on various skills related to the PE. • Allows the task developer to match features of the task with the characteristics of students such as their interests, familiarity, and provided instruction.	• Use a number of technologies. • Types of sources (graphs, tables, text, media) • Degree of scaffolding • Quantitative / Qualitative
Assessment Boundaries • List information that is NOT assessed (i.e., related above grade-level ideas and skills). • Reference the "Assessment Boundary" in the NGSS for the PE as appropriate.	• Does not include binary counting • Does not include a specific mechanism of a given device

2019 SCILLSS Classroom Science Assessment Workshop: Task Specifications Tool

Post-Workshop Survey Questions

1. Did the use of a principled-design process and assessment development tools help you when designing your classroom-based science assessment task?
2. What aspects of the principled-design process and assessment development tools are clear? Please provide suggestions as to how other aspects could be improved.
3. What is one message that you took away from this training?
4. Do you have an interest in or feel the need for further professional development in designing classroom-based assessments? Please provide reasons to explain your response.

Survey Results

1. Did the use of a principled-design process and assessment development tools help you when designing your classroom-based science assessment task?

96% of respondents (n=48) answered "yes"

- The tools guided me in the process and helped me focus on what was needed in the task.
- The tools were very helpful and made me think about each and every part of the assessment.
- The process and tools made me more aware of what I need to think about when designing new assessments.
- They allowed us to more carefully and closely consider different components of 3D teaching and standards/PEs.
- It was meaningful to be able to unpack the indicator so that we could better understand exactly what is expected of students.
- The tools helped to keep us on track. We had a great idea, but it didn't fit what we were trying to get to in the end. It helped bring us back to center.
- I was able to refer back to see if what we were asking students to do was congruent with our intentions.
- It helped me think about how prompts would elicit evidence.
- They were good guides as to what to do step by step.

Survey Results

2. What aspects of the principled-design process and assessment development tools are clear?

- The part that made the tools clear was listening to Howard talk about the research being done about assessment. This helped explain the “why” behind the tools.
- Pulling apart everything first and developing the task last was a switch. It was clear how/why the process really helps develop more valid assessments.
- The unwrapping of the standards to build a task and create a rubric.
- The unpacking and task specification tools were clear. The instructions and process were easy to understand.
- The tools were clear because examples were provided and descriptions were given on the side.

Survey Results

Please provide suggestions as to how other aspects could be improved.

- I guess I can not communicate each of the tools by name in the process a few days later. The tools were helpful in designing the tasks and unpacking the standards, but I do not know names for each tool.
- Some of the vocabulary is unfamiliar, so you are asking teachers to learn new vocabulary, unpack a standard, and then design an assessment. It is a lot to take in in two days. I found the KSAs, student demonstration of learning, work product and task features to be confusing. Maybe teachers need to have more examples of assessments to get to an understanding of what is expected.
- Wasn't sure how a task doesn't need to assess each of the KSA's AND still have the spec tool be a valid way to create several tasks. If this is true, various tasks would be made over different KSA's.
- The process was clear but making sure we were following all of the 3D expectations and the learning expectations is not easy.
- I thought all of it was clear once explained. My struggle would be to actually write KSAs. A formula to do that would be helpful.
- Slow the process down and walk the participants through more examples before creating their own.
- I just need practice…

Survey Results

3. What is one message that you took away from this training?

- It is good to think of the end in mind.
- That I CAN develop really great assessments within my classroom and I don't have to look to someone else who is an "expert" to do it for me. I also made so much progress on understanding how to assess 3D standards.
- That designing assessments are difficult to do, but not impossible.
- The importance of really taking components of standards and 3D apart in order to create tasks that will produce information about students' learning that successfully gives us the information we really want.
- Science educators need to unpack standards before doing anything else with lesson design and assessment.
- NGSS assessment is 3D and not just about using models, but so much more in-depth.
- Making sure that we are considering why we are assessing students and learning should drive the types of assessments we write.
- Large scale changes will start by evidence in the classroom.
- This process helps us to think inside out which increases the validity of an assessment.
- The time and effort put into mindful organized tasks will produce valuable products.

Survey Results

4. Do you have an interest in or feel the need for further professional development in designing classroom-based assessments? Please provide reasons to explain your response.

- I would love to work more on designing assessments and then piloting them in my classroom with rubrics. I really think it would be nice to have a basis of lessons for teachers in my state to pull from. Working with other teachers from around the state also makes the experience richer as well as the product.
- Yes! I just feel assessment done right is such a power tool in teaching and I want to learn everything there is to know about how to design assessments.
- Yes! More please! I need to deepen my understanding and ability to transfer the information to others.
- Yes, I would have liked to have the opportunity to make more tasks. It was great to produce one, however, I wish there had been a lot more in attendance so each team could have made one and shared. (Maybe leaving with an entire unit or one task per unit of standards.)
- I believe that this process takes time and doing it with other people is beneficial. I also would use a video.

Survey Results

4. Do you have an interest in or feel the need for further professional development in designing classroom-based assessments? Please provide reasons to explain your response.

- YES! I could use several days more training on these same processes to feel comfortable enough with this design process that I can create tasks as well as become comfortable enough to assist other teachers with the process (informal settings and in settings such as Nebraska Association for Teachers of Science conference).
- Yes, because I feel like we skimmed the top of this process and am not sure I'd be confident that we did it correctly.
- Yes, but only so I can get faster at it.
- Yes! Mostly I feel the need to have more practice with guidance.
- It would be amazing to have a repository of the unpacking and task specification tools completed. I understand there is great value in building them, but time is always an issue.
- I think other teachers in my science district would benefit from learning this process. The big hurdle is finding the time to work on creating the assessments.

Principled Design from a State and Local Perspective

State Perspective – Nebraska

Nebraska is utilizing the principled-design approach to develop science assessments in our system.

- NE theory of action calls for curriculum, instruction, and assessment designed for Nebraska College and Career Ready Standards to be implemented systemically and systematically.

Benefits:

✓provides coherence in science assessment system
✓leads to instructional shifts by teachers
✓grows assessment literacy of teachers
✓results in valid high-quality tasks with meaningful scores

Professional Learning in Nebraska: Focus on Principled Design

Summer 2020

- Shift in focus from summative to formative classroom science assessment development due to pandemic
- Teachers use principled-design approach, tools, and templates
- Teachers develop 24 classroom tasks for both grade 5 and grade 8 that span the breadth of the NE science standards

Summer 2021

- Summative task development
- Teachers utilize principled design
- Teachers are familiar with the process, tools, and templates from classroom formative workshops

Local Support for Principled-Design

Local science teacher associations support and champion the process, tools, and templates.

- Nebraska Association of Teachers of Science
  - Two-day pop-up workshop January 2019
  - 50 educators and professional developers
  - Classroom tasks developed at grade 5, 8 and HS Life and Earth Sciences

- Nebraska Educational Service Units lead regional professional learning

74

Classroom Pilot Tasks

Teachers' comments about the experience

- The assessments were of great quality.
- Writing a constructed response was hard for many of my students, so I need to provide more instruction and opportunities for my students to write in science.
- I plan to use these results as a way to reteach the misconceptions I saw.
- Time to learn this process is a factor.
- I will be more deliberate about teaching students how to use claims/evidence/reasoning model of communication.
- I will be more explicit about teaching 5th grade students how to respond to these type of prompts.
- The writing was overwhelming to some students, so they gave up.

References

Almond, R.G., Steinberg, L.S., & Mislevy, R.J. (2002). Enhancing the design and delivery of assessment systems: A four-process architecture. *Journal of Technology, Learning, and Assessment, 1*(5).

American Educational Research Association (AERA), the American Psychological Association (APA), and the National Council on Measurement in Education (NCME) Joint Committee on Standards for Educational and Psychological Testing. (2014). *Standards for educational and psychological testing*. Washington DC: AERA.

Forte, E. (2013a). *Re-conceptualizing alignment in the evidence-centered design context*. Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, CA.

Forte, E. (2013b). *Evaluating alignment for assessments developed using evidence- centered design.* Paper presented at the Annual Meeting of the National Council on Measurement in Education, San Francisco, CA.

Huff, K., Steinberg, L., & Matts, T. (2010). The promises and challenges of implementing evidence-centered design in large-scale assessment. *Applied Measurement in Education, 23*(4), 310-324.

Messick, S. (1994). The interplay of evidence and consequences in the validation of performance assessments. *Educational Researcher, 23*(2), 13-23.

References

Mislevy R. (2007). Validity by design. *Educational Researcher*, 36(8), 463-469.

Mislevy, R. J., & Haertel, G. (2006). Implications of evidence-centered design for educational assessment. *Educational Measurement: Issues and Practice, 25,* 6-20.

Mislevy, R.J., Steinberg, L.S., & Almond, R.G. (2002). On the structure of educational assessments. *Measurement: Interdisciplinary Research and Perspectives, 1,* 3-67.

NGSS Lead States. (2013). Appendix E – Progressions Within the Next Generation Science Standards. *Next Generation Science Standards: For States, By States.* Washington, DC: The National Academies Press.

National Research Council. (2001). *Knowing what students know: The science and design of educational assessment*. Committee on the Foundations of Assessment. Pellegrino, J., Chudowsky, N., and Glaser, R., editors. Board on Testing and Assessment, Center for Education, Division of Behavioral and Social Sciences and Education. Washington DC: National Academy Press.

National Research Council. (2006). *Systems for State Science Assessment.* Committee on Test Design for K–12 Science Achievement. M.R. Wilson and M.W. Bertenthal, eds. Board on Testing and Assessment, Center for Education, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.

National Research Council. (2012). *A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas.* Committee on a Conceptual Framework for New K-12 Science Education Standards. Board on Science Education, Division of Behavioral and Social Sciences and Education, Washington, DC: The National Academies Press.

References

National Research Council. (2014*). Developing Assessments for the Next generations o Science Standards.* Committee on a Conceptual Framework for New K-12 Science Education Standards. Board on Science Education, Division of Behavioral and Social Sciences and Education, Washington, DC: The National Academies Press.

Pellegrino, J. W. (2018). *Assessing Science Learning: Some of the Complexities & Challenges* [PowerPoint Slides].

Steinberg, L. S., Mislevy, R. J., Almond, R. G., Baird, A. B., Cahallan, C., DiBello, L. V., et al. (2003). *Introduction to the Biomass project: An illustration of evidence-centered assessment design and delivery capability (CSE Technical Report 609)*. Los Angeles: National Center for Research on Evaluation, Standards, and Student Testing (CRESST), Center for the Study of Evaluation, UCLA.