PERFORMANCE-BASED ASSESSMENTS FOR TEACHING AND LEARNING

FRIDAY, APRIL 13, 2018
3:00 PM ET/ 2:00 PM CT/ 1:00 PM MT/ NOON PT
NCSL Student-Centered Learning Commission

About the Commission

- The NCSL Student-Centered Learning Commission is a bipartisan group of state legislators studying legislative policy options, obstacles and recommendations to help state legislators move forward with systems that support student-centered learning opportunities.

Examples of Student-Centered Learning Principles

- Learning is personalized
- Learning is competency-based
- Learning takes place anytime, anywhere
- Students have ownership over their learning
NCSL Student-Centered Learning Commission Webinar Series

- Webinar 1: Overview of K-12 Competency-based Learning
- Webinar 2: ESSA and Personalized Learning: A Nationwide Look
- Today’s Webinar: Performance-based Assessments for Teaching and Learning
Today’s Presenter

Scott Marion
Executive Director,
Center for Assessment
April 13, 2018
Scott Marion, Center for Assessment

Innovative Assessments for Teaching and Learning
NCSL Student-Centered Learning Commission
Presentation Outline

• Current context
• Performance assessment basics
• Deeper learning and performance assessment
• The New Hampshire example
• Ongoing challenges
Concerns About Current Testing

✓ We’ve over-promised what our tests can do
✓ We’re over-testing because of an incoherent Babel of tests
✓ We’ve under-delivered meaningful and useful information to teachers and students
✓ Many of our test are irrelevant for students
✓ We are not capitalizing on some key tech advances
✓ Lack of assessment literacy
Back to the Future

We were doing really exciting things in the decades prior to NCLB, just to name a few:

- Maryland School Performance Assessment Program
- Kentucky Reform Act
- California Learning Assessment System
- Oregon’s CIM & CAM
- Vermont Portfolio System
- Wyoming Body of Evidence System

And those were just the state-level reforms!
It Wasn’t Perfect

• The rush to accountability in certain states led to concerns—many misplaced—about technical quality
• But we felt like we were learning how to do some really interesting things
• This NRC report documents many of those efforts

Why are we here?

The business of schools is to invent tasks, activities, and assignments that the students find engaging and that bring them into profound interactions with content and processes they will need to master to be judged well educated.

Performance-based assessments appear to be the common denominator!

Performance assessments are generally multi-step activities ranging from quite unstructured to fairly structured. The key feature of such assessments is that students are asked to produce a product or carry out a performance (e.g., a musical performance) that is scored according to pre-specified criteria, typically contained in a scoring guide or rubric (Marion & Buckley, 2016, p. 51).
Why Performance Assessment?

• It costs more, takes more time, and is hard to do well, so why do it?
  – The only way to measure the intended construct (i.e., theoretical trait or constellation of knowledge and skills in a domain),
  – A better way to measure the intended construct,
  – To produce instructional information in addition to accountability information,
  – To provide both a learning and assessment opportunity for students, and
  – To signal the types of instructional tasks many would like to see in classrooms.
### Characteristics of High-Quality Tasks
*(Shepard & Marion, 1997)*

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Essential</strong></td>
<td>Represents the big ideas and skills of the domain</td>
</tr>
<tr>
<td><strong>Complex</strong></td>
<td>Requires students to engage with the content in deep and meaningful ways</td>
</tr>
<tr>
<td><strong>Authentic</strong></td>
<td>Not contrived, represent real-world activities</td>
</tr>
<tr>
<td><strong>Equitable</strong></td>
<td>Not biased, allow diverse students to show what they know</td>
</tr>
<tr>
<td><strong>Instructional</strong></td>
<td>Coherent with instruction and should provide students an opportunity to learn</td>
</tr>
<tr>
<td><strong>Rich</strong></td>
<td>Opportunities to develop extensions beyond task</td>
</tr>
<tr>
<td><strong>Engaging</strong></td>
<td>Thought-provoking and interesting problems</td>
</tr>
<tr>
<td><strong>Active</strong></td>
<td>Students construct meaning with other people and/or resources</td>
</tr>
<tr>
<td><strong>Accessible</strong></td>
<td>Students of differing ability levels can work productively on the task</td>
</tr>
<tr>
<td><strong>Feasible</strong></td>
<td>Can be completed within bounds of time and cost, locally appropriate</td>
</tr>
</tbody>
</table>
Note that good assessment tasks are interchangeable with good instructional tasks.

We also have evidence that “teaching to” problem types like these improves learning.
What is deeper learning?

- **Deeper learning**, as a term, has been popularized by a major Hewlett Foundation initiative.
- Modern theories of learning focus on developing deep understanding to facilitate transfer.
  - Why do we care about transfer?

Students cannot develop deep understanding unless they are provided opportunities on both learning and assessment tasks.
Assessments must be based on research-based models of learning in order to ensure appropriate interpretations and instructional actions.

Adherence to outdated, naïve, and/or implicit notions of learning are an impediment to assessment literacy and assessment reform.

Generalization

• **Generalizability** is really the **assessment** analogue of **transfer in learning**

• Do the results of one assessment or a small set of assessments provide **credible evidence** that the student **really knows (is competent)** what is being claimed based on the assessment scores?
  – In other words, is it a “one and done?”

• A **concern** for **CBE** systems designed to declare that students have demonstrated mastery of the competency and can “move on” (e.g., receive credit)
Cognitive Rigor

• Cognitive rigor is one of the ways that we operationalize deeper thinking

• The cognitive rigor of an assessment task is:
  – The type and level of thinking required of students to successfully engage with and solve a task
  – Characterized by the way in which students interact with content
Increasing cognitive complexity (using a math example)

Fill in the given table with the results of your bouncy ball experiment.

Organize the results of your bouncy ball experiment in a table.

Create a mathematical model that best reflects the results of your bouncy ball experiment and justify your decision for the chosen model.
"I'm saying that, if you'd been a little more proficient in math, this would never have happened!"
Why Innovate?

- Need to find ways to support multiple users in the system
- Need to “rebalance” the system
- Need to support increases in student and educator learning
  - We need to capitalize on the affordances offered by technology
  - Need to better capture thinking processes as well as products
- Need to manage costs
New Hampshire’s Innovative Model

• The New Hampshire Department of Education (NH DOE) was granted a series of waivers from NCLB and ESSA to implement the **Performance Assessment of Competency Education** (PACE) as a pilot assessment and accountability system for a limited number of school districts.

• We just submitted an application for the ESSA Innovative Assessment Demonstration Authority.

• Led by the NH DOE in close partnership with the district leads and the Center for Assessment.
PACE as a “re-Balanced” Assessment System

• The emphasis on local assessments and collaboratively-created “common tasks” along with the limited use of the state assessment helps to rebalance the system.

• Such a system supports multiple stakeholders:
  – Teachers
  – Leaders
  – Policy Makers
  – Parents
  – Students!
The PACE Assessment System

Competency 1

Competency 2

Competency 3

Competency 4

Local performance assessments

Local performance assessments

Local performance assessments

Local performance assessments

District-Level Competency Scores

PACE Common Performance Task

PACE Comparable Annual Determinations

State summative assessment in select grades

Marion. Center for Assessment. NCSL April 13, 2018
Supporting Deeper Learning for Students

The assessments used to evaluate student mastery of the PACE competencies are designed to embody rich learning goals.

- Remember our discussion about deep understanding, transfer, and generalizability
- Students need multiple and varied opportunities with both learning and assessment tasks
The Problem: Your town’s population is predicted to increase over the next 3 years. As one of the town planners, you are asked to address this issue in terms of the town’s water supply. In order to meet the future needs of the town, you need to make a proposal to add a water tower somewhere on town property that will be capable of holding 45,000 ± 2,000 cubic feet of water. The town is looking for a water tower to contain the most amount of water while using the least amount of construction material.

Student Task: Your job is to prepare a proposal that can be submitted to the town planning committee. Using your calculations of surface area and volume for two different designs, describe and analyze the characteristics that lead you to a final recommendation.
Essential Question: How is energy transferred between places and converted between types?

• You are working for a company that wants to find affordable and environmentally-friendly ways to reduce the need for wood and charcoal when cooking.
• You have been tasked to create a device that uses renewable energy.
• You and a group will research, design, build, and test a solar cooker, applying everything you have learned about energy this past quarter.
• Your final goal is to change the temperature of a cup of water.
Nice Tasks, but...

• We learned how to create rich and engaging tasks in the past

• Principled assessment design approaches (e.g., ECD) helps to develop task models to avoid reinventing the wheel

We’ve also learned a lot about improving the reliability and accuracy of scoring tasks and comparably evaluating student work across schools & districts
But we still have issues and challenges...

- **Scaffolding**—what is the role of the teacher or other adult in guiding the student toward competency?
- **Whose work is it (group)?** In many extended projects and tasks, having students work in groups for all or part of the task is both efficient and educationally appropriate.
- **Generalizability**—How much evidence does it take for you to believe that the student truly is competent? In other words, what amount and what type of evidence are **sufficient** to support your claims?
- **Comparability**—when students have choice in how they demonstrate competency, how do we know that the multiple ways are comparable (enough)?
- **Assessment literacy**...
Building Assessment Literacy

• Developing **assessment literacy** among educators and leaders has been an unmet challenge for more than 50 years

• Now we are raising the stakes with assessment literacy requirements for **competency** and other **deeper learning** systems

• We can draw from what we’ve learned about **improving classroom assessment practices**...
What We’ve Learned from Research on Teaching

• Teachers are most likely to make and sustain fundamental changes in instructional (and assessment) practices if provided:
  – Time
  – Conceptual and strategic support
  – Opportunities to try new practices in the context of their own teaching

• **Principle #6:** *We learn to do the work by doing the work. Not by telling other people to do the work, not by having done the work at some time in the past, and not by hiring.*

• We are employing CBE and personalized learning because we are trying to engage students in more authentic and meaningful learning.

• Why wouldn’t we do the same for educators and leaders?
Authentic Learning

• A common thread of among highly effective professional learning was that the work “counted”
  – Wyoming Body of Evidence
  – NH’s PACE pilot

• In other words, the work of these professional learning opportunities was authentically used in assessment and accountability systems almost immediately
Many continue to talk about “train-the-trainers” even though there is little evidence of effectiveness.

A deeper, research-based approach is based on Lave and Wenger’s (1991) study of how “apprentices” become “masters”

– We’ve used this framework to support “cadres of experts” who then helped sustain and lead the work.
Finally, I’d be remiss if I didn’t mention the need for real innovation in our technological supports for the type of program we’re trying to implement in NH:

✓ Collaborative task authoring
✓ Asynchronous scoring consistency evaluations
✓ Cross-school or district calibration
✓ Uploading and archiving student work
✓ Uploading archiving other data (e.g., school grades, demographics)
Stay Connected

- An archived version will be available shortly, please visit:

- Learn more about NCSL’s Student-Centered Learning Commission:

- Connect with NCSL Staff:
  Sunny Deye, Program Director, Education Program
  sunny.deye@ncsl.org / 303-856-1469