BRINGING PERFORMANCE-BASED ASSESSMENT INTO THE SUMMATIVE SPACE

June 28, 2017
Objectives

• To understand the foundational steps states and local education agencies must take to build capacity for using performance assessment in the summative space

• To understand the technical challenges that must be addressed in order to use performance assessments for summative uses

• To highlight examples from Ohio and other contexts about their journey to move in this direction
Our Presenters

• Dr. Raymond Pecheone, Stanford Center for Assessment, Learning and Equity (SCALE)

• Dr. Stuart Kahl, Measured Progress

• Dr. Ruth Chung Wei, SCALE

• Dr. Susan Zelman, the Ohio Department of Education
Agenda

• Principles for Designing an Innovative Assessment System FOR Learning

• Lessons Learned – What Have We Learned from Past Performance Assessment Initiatives?

• Examples of Emerging Approaches for Performance Assessment Use in the Summative Space

• A State’s Perspective: Insights from Ohio
Principles for Designing an Innovative Assessment System FOR Learning

Raymond L. Pecheone, Stanford University
Room to Innovate

Innovative assessment systems may make use of: “competency-based assessments, instructionally embedded assessments, interim assessments, performance-based assessments that combine into an annual summative determination for a student,” and “assessments that validate when students are ready to demonstrate mastery or proficiency and allow for differentiated support based on individual learning needs”

--ESSA, 2015, Section 1204(a)
Why Principles for Design?

• ESSA does not provide an underlying theory of action about the broader purposes assessments should serve (as did NCLB)

• States now have the autonomy to be the “master of their ship”

• Not all designs will lead to better assessment
Innovative Assessment FOR Learning
~Some Principles

A. Be part of a multiple measures system that is connected to curriculum and instruction
B. Support both teacher and student learning
C. Engage teachers throughout the process
D. Provide timely, useful information to teachers, students, parents, and schools
E. Be practical in the service of deeper learning
F. Support language development in diverse classrooms
A. Assessments should be part of a multiple measures system that is connected to curriculum and instruction.

Principle A in action: (An example)

A collection of *instructionally-connected* Classroom Based Assessments (CBAs) comprised of performance assessments and other appropriate item types aligned to the learning outcomes to be measured
B. Assessments should be supportive of both teacher and student learning.

Principle B in action: (Examples)

- Rigorous standards-aligned tasks that prompt student thinking and allow students to exercise voice and choice
- Teachers score some of their own students’ work so they can get results quickly and act immediately on the results
C. Assessments should engage teachers throughout the assessment process.

Principle C in action: (Examples)

- Teachers select or design Classroom Based Assessments that fit best within their curriculum,
- Teachers determine when students are ready to demonstrate mastery of the targeted competencies and administer the CBAs they have selected in any sequence across the academic year, following instruction relevant to the content addressed by the task.
BUT What about comparability?

- State-level CBA bank that includes multiple CBAs options for each measurement target per grade level, common scoring tools
- Teachers may design new tasks according to task specifications and submit for inclusion in bank in future years
- State-level assessment quality review board
Intermittent use of standardized tests in some years (an example)

<table>
<thead>
<tr>
<th>Grade</th>
<th>ELA</th>
<th>Mathematics</th>
<th>Science</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>CBAs</td>
<td>CBAs</td>
<td>CBAs</td>
</tr>
<tr>
<td>4</td>
<td>CBAs + Standardized Test</td>
<td>CBAs</td>
<td>CBAs</td>
</tr>
<tr>
<td>5</td>
<td>CBAs</td>
<td>CBAs + Standardized Test</td>
<td>CBAs + Standardized Test</td>
</tr>
<tr>
<td>6</td>
<td>CBAs</td>
<td>CBAs</td>
<td>CBAs</td>
</tr>
<tr>
<td>7</td>
<td>CBAs + Standardized Test</td>
<td>CBAs</td>
<td>CBAs</td>
</tr>
<tr>
<td>8</td>
<td>CBAs</td>
<td>CBAs + Standardized Test</td>
<td>CBAs + Standardized Test</td>
</tr>
<tr>
<td>9-12</td>
<td>CBAs + Standardized Test</td>
<td>CBAs + Standardized Test</td>
<td>CBAs + Standardized Test</td>
</tr>
</tbody>
</table>
BUT What about validity and reliability?

- **Multiple measures** – performance assessments along with other item types are used and scaled together (constructed response, selected-response) from across different assignments and tests.

- **Scoring options:**
  - Local scoring with external audit, score adjustments
  - Local scoring (10%) with external scoring (90%)
  - External moderation should include teachers to support ownership
D. Assessments should provide timely, useful information to teachers, students, parents, and schools.

Principle D in action: (Examples)

- Teachers receive periodic release days to score student work on CBAs;
- Teachers participate in local “professional moderation boards” to audit local scores;
- “Data dashboards” of student performance on all CBAs readily available to teachers and parents.
E. Assessments should be educative and practical.

Principle E in action: (Examples)

• Assessments “worth doing”
• Assessments spread across the year
• Assessments are administered like classroom assignments, not testing events
• Assessments generate actionable feedback
F. Assessments should support language development in diverse classrooms.

Principle F in action: (Examples)

• Assessments are designed and vetted with systematic attention to language development needs and opportunities.

• Teachers are involved in designing, vetting, and scoring of assessments with this design feature, which builds capacity to attend to their own students’ language development
Other Technical Considerations in the Summative Space

Validity
• Tight alignment to standards, including deeper learning outcomes and more complex cognitive demands

Comparability
• Common Architecture and Design Criteria (e.g. Cognitive complexity)
• State Assessment Quality Review Board, that includes educators
Other Technical Considerations in the Summative Space

Comparability to Standardized Assessments?

Two interpretations -

1. Strong correlation with Standardized Test scale scores
2. Proficiency level determinations are comparable
Other Technical Considerations in the Summative Space

Defensible Score Scales

• Common rubrics or scoring tools developed and accepted by state level stakeholders
• Hybrid – common state-level core rubrics with local add-on rubrics to address local priorities
Reinvention of the Summative Space

Working toward a Culture of Revision and Redemption – IS NOT “CHEATING”

• A competency based approach to measuring – rather than a “one and done” approach.
• Allows for opportunities for peer review and collaboration, feedback, “revise and resubmit”.
• Focus is on continuous improvement and growth (a “growth mindset”).
Let’s Give Performance Assessment a Chance

Presentation in session “Bringing Performance-Based Assessment into the Summative Space” at NCSA, Austin, June 28, 2017
35 years of involvement in large-scale performance assessment

I believe “real” performance assessment can play a significant role in state accountability assessments.
What have we learned from past performance assessments?

Performance Assessment Studies

- Connecticut Assessment of Educational Progress (CAEP) in Science -- 1983
- Massachusetts Educational Assessment Program (MEAP) – 1989
What have we learned from past performance assessments?

Performance Assessments that “Count” (note: all central scoring)

- CAEP Business and Office Education (BOE) – 1985
- Rhode Island Distinguished Merit Program – 1989-96
- California Golden State Examination in Chemistry – 1992-95
- Kentucky (KIRIS) Performance Events – 1992-95
- Colorado Alt – 2001-
- NECAP Science
Attributes Limiting Feasibility

- Using trained external administrators
- Providing materials
- 100% central scoring
- Time and cost!
What have we learned from state portfolio assessments?

- Vermont – early ‘90s
- Kentucky – 1992…
Problems with Portfolio Assessments

- Task quality
- Scoring consistency
- Whose work is it?

(Note: With respect to the first two, writing and math are different animals.)
Reliability of Direct Writing Assessments

![Graph showing the relationship between the number of tasks and score reliability across different studies. The graph includes data from Lindquist (1927), Coffman (1966), Swartz et al (1985), Breland et al (1987), Heironymus & Hoover (1987), and Welch (1991).]
# Analytic Scoring of Direct Writing

Average Inter-correlations of Analytic Traits

<table>
<thead>
<tr>
<th>Prompt</th>
<th>Scorer</th>
<th>Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ice Age</td>
<td>Human</td>
<td>.947</td>
</tr>
<tr>
<td></td>
<td>e-rater$^R$</td>
<td>.614</td>
</tr>
<tr>
<td></td>
<td>IEA</td>
<td>.753</td>
</tr>
<tr>
<td>Reflection</td>
<td>Human</td>
<td>.917</td>
</tr>
<tr>
<td></td>
<td>e-rater$^R$</td>
<td>.535</td>
</tr>
<tr>
<td></td>
<td>IEA</td>
<td>.648</td>
</tr>
</tbody>
</table>
Covering a Subject Area Domain

Multiple Choice

Open-Ended

Performance
Key Principles for Implementation of Innovative Accountability Assessments

- Teachers should be able to see relatively immediate instructional benefit from at least one component of a program.
- Assessment innovation should not place an additional burden on teachers.
- Transition to an innovative approach should be phased over time.

For more on these principles, see: http://www.gettingsmart.com/2016/12/innovating-large-scale-assessments/
An Optimal Program Design

<table>
<thead>
<tr>
<th>End-of-Year Summative Test</th>
<th>Performance Component</th>
</tr>
</thead>
<tbody>
<tr>
<td>Common</td>
<td>Matrix-Sampled</td>
</tr>
<tr>
<td>Field Test</td>
<td>3 Curriculum-Embedded</td>
</tr>
<tr>
<td>4-pt Constructed Response</td>
<td>Performance Assessments</td>
</tr>
<tr>
<td>25</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>60 minutes</td>
<td>Multi-Day Instructional Units</td>
</tr>
</tbody>
</table>

25 multiple-choice items.

25 2 5 1

60 minutes
Curriculum-Embedded Performance Assessment (CEPA) -- Definitions

- **A CEPA** is a multi-day instructional unit or module that consists of a sequence of instructional activities, some of which lead to student work that is used for formative purposes and some of which lead to student work that can be used for summative assessment.

- **CEPA** (Curriculum-Embedded Performance Assessment) is the general approach of using CEPAs as a component of an assessment system.

For more on CEPA go to:

A CEPA should:

- introduce or refresh foundational knowledge and skills,
- require students to demonstrate deeper learning by applying the foundational knowledge and skills,
- yield multiple, individually-produced, scorable student work products for summative use (independent measures worth a total of at least 20 points),
- involve the use of technology somewhere in the unit,
- involve small group work somewhere in the unit.
A CEPA Packet

Includes:

- Listing of relevant content standards and learning targets
- Descriptions of instructional activities and embedded performance tasks
- Scoring rubrics
- Sample student work
Teacher Input and “Ownership”

- Flexibility regarding some CEPA activities
- Development of model CEPAs
- Submission of teacher-developed CEPAS
- Teacher choice of CEPAs for accountability
- Teacher scoring

For more on teacher ownership and phasing in the innovative assessment model, see: http://www.eschoolnews.com/2017/04/24/edtech-performance-assessments/
The Two-Component Accountability Assessment Model

- addresses all the major concerns about the use of performance assessment for statewide accountability and about state accountability assessments generally,
- is consistent with the principles listed earlier,
- contributes to a balance between:
  - formative and summative assessment
  - local and state assessment
  - attention to foundational knowledge and deeper learning
and is FASTER, BETTER, CHEAPER!

For a discussion of shortcomings of other interim assessment approaches for ESSA, see:
It’s all about student learning. Period.

Thank you.
DISCUSSION – Q&A
Examples of Emerging Approaches for Performance Assessment Use in the Summative Space

Ruth Chung Wei, Stanford University
Key Criteria for Innovative Assessment System (ESSA)

- **Same assessments** must be used to measure the achievement of all participating students in the innovative assessment system.
- Must be comparable to the state academic assessment and provide for an *equally rigorous* and *statistically valid comparison* between student performance on the innovative assessment and the existing statewide assessment.
Key Criteria for Innovative Assessment System (ESSA)

• Be aligned to the challenging state academic standards and address the **depth and breadth** of such standards

• Express student results or student competencies in terms consistent with the State’s aligned academic achievement standards

• Generate results that are **valid and reliable**, and **comparable**, for all students and for each subgroup of students
Working toward comparability

Task Comparability
• Common rubrics and scoring tools
• Clearly defined content and design specifications
• Common templates or task shells

Comparability with State Tests – the challenge
• Address the same measurement targets as the standardized tests
• Address depth and breadth of the standards
Approach 1 – Mapping a Portfolio of Assessments to a State Test Blueprint

Ohio Competency Based Education (CBE) Performance Assessment Pilot

• 6th grade mathematics
• Algebra
• 8th grade science
• U.S. History
### Framing a Portfolio of Assessments to Valued Outcomes

#### Critical Abilities

<table>
<thead>
<tr>
<th>Critical Abilities</th>
<th>Curriculum Embedded</th>
<th>On-Demand</th>
<th>Constructed Response</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>Research</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Analysis of Information</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Experimentation and Evaluation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Communication in Many Forms</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Use of Technology</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interpersonal Interaction and Collaboration</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Modeling, Design, and Problem Solving</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Research** includes 3 or more Critical Abilities;
**Analysis of Information** includes 2 or more Critical Abilities;
**Experimentation and Evaluation** includes at least 1 Critical Ability;
**Communication in Many Forms** attempts to include 1 Critical Ability.

#### Math Practices

<table>
<thead>
<tr>
<th>Math Practices</th>
<th>Curriculum Embedded</th>
<th>On-Demand</th>
<th>Constructed Response</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>Make sense of problems and persevere in solving them.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reason abstractly and quantitatively.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Construct viable arguments and critique the reasoning of others.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model with mathematics.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Use appropriate tools strategically.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Attend to precision.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Look for and make use of structure.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Look for and express regularity in repeated reasoning</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Math Practices** includes 5 or more Math Practices;
**Reason abstractly and quantitatively.** includes 4 or more Math Practices;
**Construct viable arguments and critique the reasoning of others.** includes 3 or more Math Practices;
**Model with mathematics.** includes 2 or more Math Practices.
<table>
<thead>
<tr>
<th>OH Blueprint</th>
<th>Content Standard</th>
<th>Curriculum Embedded</th>
<th>On-Demand</th>
<th>Constructed Response</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>24-33%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(13-17 points)</td>
<td>6.RP.1</td>
<td>Shipping Dilemma</td>
<td>Best Buy</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>6.RP.2</td>
<td>Shipping Dilemma</td>
<td>Best Buy</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>6.RP.3</td>
<td>Shipping Dilemma</td>
<td>Vegetable Garden &amp; Best Buy</td>
<td>Car Color, The Race, &amp; Lawn Care</td>
<td></td>
</tr>
<tr>
<td>31-44%</td>
<td>6.EE.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(17-23 points)</td>
<td>6.EE.2</td>
<td>Jump a Thon</td>
<td>Dog Walking</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>6.EE.3</td>
<td>Summer Jobs</td>
<td>Dog Walking</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>6.EE.4</td>
<td>Summer Jobs</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>6.EE.5</td>
<td>Jump a Thon</td>
<td>Dog Walking</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>6.EE.6</td>
<td>Summer Jobs, Jump a Thon</td>
<td>Dog Walking</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>6.EE.7</td>
<td>Summer Jobs, Jump a Thon</td>
<td>Dog Walking</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>6.EE.8</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>6.EE.9</td>
<td>Summer Jobs</td>
<td>Snow Melt</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20-25%</td>
<td>6.NS.8</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(11-13 points)</td>
<td>6.G.1</td>
<td>Vegetable Garden</td>
<td>Lawn Care</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>6.G.2</td>
<td>Shipping Dilemma</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>6.G.3</td>
<td></td>
<td>Designing a Race</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>6.G.4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>6.SP.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>6.SP.2</td>
<td></td>
<td>Going to the Movies</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OH Blueprint</td>
<td>Content Standard</td>
<td>Curriculum Embedded</td>
<td>On-Demand</td>
<td>Constructed Response</td>
<td>Other</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------------</td>
<td>---------------------</td>
<td>------------</td>
<td>----------------------</td>
<td>-------</td>
</tr>
<tr>
<td>6.SP.3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6.SP.4</td>
<td></td>
<td></td>
<td></td>
<td>Going to the Movies &amp; Buying on-line or in-store</td>
<td></td>
</tr>
<tr>
<td>6.SP.5</td>
<td></td>
<td></td>
<td></td>
<td>Going to the Movies &amp; Buying on-line or in-store</td>
<td></td>
</tr>
<tr>
<td>20-25% (11-13 points)</td>
<td>6.NS.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>6.NS.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>6.NS.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>6.NS.7</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>6.NS.2</td>
<td></td>
<td></td>
<td>Snow Melt</td>
<td></td>
</tr>
<tr>
<td></td>
<td>6.NS.3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>6.NS.4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
**Approach 1 - Important Technical Considerations**

- “Common Task” Approach
- Common Scoring Guides – **Scoring criteria aligned to specific standards**
- Teacher designed with expert review and revision, supplemented with expert designed tasks
- Inclusion of varied item types
- Teachers calibrate and score, with external audit
- Task security
Approach 2 – Mapping a Collection of Performance Tasks to Assessment Claims

Evidence-Centered Design (Mislevy, Almond, & Lukas, 2003)

• Standards clustered into Claims, Targets, and Evidence Statements

SMARTER Balanced Claims

<table>
<thead>
<tr>
<th>ELA Claims</th>
<th>Mathematics Claims</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reading</td>
<td>Concepts &amp; Procedures</td>
</tr>
<tr>
<td>Writing</td>
<td>Problem Solving</td>
</tr>
<tr>
<td>Research</td>
<td>Communicating Reasoning</td>
</tr>
<tr>
<td>Listening</td>
<td>Modeling &amp; Data Analysis</td>
</tr>
</tbody>
</table>
Approach 2 – Mapping a Collection of Performance Tasks to Assessment Claims

Literacy Design Collaborative Modules
http://coretools.ldc.org

• Writing/Research tasks – Explanatory and Argumentative Writing or Research
• Embedded mini-tasks within each module address a range of CCLS standards
• Independent mini-tasks
Example – LDC Writing Module

CCLS Writing 5.2 Write informative/explanatory texts to examine a topic and convey ideas and information clearly

Explanatory Writing Prompt (Grade 5)
What is the theme of the poem Mother to Son? After reading Mother to Son (and an informational text on metaphors), write an essay for our class literary magazine in which you analyze how Langston Hughes' use of metaphors contributes to an understanding of the theme of this poem. Support your response with evidence from the text/s. Include several examples from the texts in your response.

Source: https://coretools.ldc.org/mods/a6c48a89-72c3-4768-98b6-e7bdaecbba98
Example – LDC Writing Module – Embedded mini-task

CCLS Reading Literature 5.4. Determine the meaning of words and phrases as they are used in a text, including figurative language such as metaphors and similes.

Prompt:
Paraphrase each stanza in the poem, demonstrating a clear understanding of the meaning of the metaphors the poet used.

Source: https://coretools.ldc.org/mods/a6c48a89-72c3-4768-98b6-e7bdaecbba98
Example – Mapping SMARTER Balanced ELA Assessment Claims to LDC Writing/Research Tasks, Mini-tasks, and other item types

<table>
<thead>
<tr>
<th>SBAC Claim 1</th>
<th>SBAC Claim 2 &amp; 4</th>
<th>SBAC Claim 3</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Reading</strong></td>
<td><strong>Writing/Research</strong></td>
<td><strong>Listening</strong></td>
</tr>
<tr>
<td>Grades 3-5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1) DRA(^1) or other on-demand reading instrument (teacher-scored, 2x/year)</td>
<td>1) LDC Writing Task (Informational or Opinion)</td>
<td>1) 20-min on-demand test of listening (computer-scored)</td>
</tr>
<tr>
<td>2) Response to Literature Mini-Tasks (Constructed Response, teacher-scored)</td>
<td>2) LDC Research Task (Informational or Opinion)</td>
<td></td>
</tr>
<tr>
<td>3) Reading Mini-tasks within LDC Writing Task</td>
<td>(1 Essay, 1 Research Report, teacher-scored)</td>
<td></td>
</tr>
<tr>
<td>4) Reading Mini-Tasks within LDC Research Task</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(^1\) DRA: Developmental Reading Assessment [http://www.pearsonschool.com/index.cfm?locator=PSZ4Z4&PMdbProgramId=23661&prognav=po](http://www.pearsonschool.com/index.cfm?locator=PSZ4Z4&PMdbProgramId=23661&prognav=po)
Approach 2 - Important Technical Considerations

- “Common Task Shell/Template” Approach
- Common Scoring Guides for each Assessment Claim - Scoring criteria aligned to specific Targets
- Teacher designed with peer review and feedback; national reviewers certify quality
- Inclusion of varied item types (essay, constructed response, on-demand performance, selected response)
- Teachers calibrate and score, with external audit
Approach 3 – Common Interim Performance Tasks (Math, Science)

Antelope Valley, California/ L.A. County Office of Education

• Teacher-developed performance tasks with expert review and feedback

• Administered as interim assessments, completed in an on-demand setting. Similar to earlier prototypes of SMARTER Balanced mathematics performance tasks.
Approach 3 - Important Technical Considerations

- **“Common Task” Approach** (Sampling topics and content domains from across a grade level course)
- Teacher designed with expert review and revision to ensure quality and comparability
- Limitations of using only performance tasks
- Task-specific scoring guides
- Teachers calibrate and score after local training, with district-level audit
- Task security
Performance Assessment – A State’s Perspective

Dr. Susan Tave Zelman – June 28, 2017
State Responsibility

If we want our education system to promote authentic problem-based learning, the state has a responsibility to provide authentic performance-based assessments in a summative system.
Moving Toward A Balanced Summative Assessment System

Examples

- Traditional Tests
- CCSS Assessments (SBAC & PARCC)
- Common Performance Tasks (Ohio, New York)
- C-PAS College Ready Assessments
- Student-Designed Projects (Envision, NY Performance Standards Consortium, Singapore, IB)

Descriptions

- Narrow Assessment
  - Standardized, multiple-choice tests of routine skills
  - Standardized tests with m-c & open-ended items + short (1-2 day) performance tasks of some applied skills
- Assessments of Deeper Learning
  - Standardized performance tasks (1-2 weeks) that include structured inquiry and demand more integrated skills, including collaboration
  - Performance tasks that require students to formulate and carry out their own inquiries, analyze & present findings, and (sometimes) revise in response to feedback
  - Longer, deeper investigations, (2-3 months) & exhibitions, including graduation portfolios, requiring students to initiate, design, conduct, analyze, revise, and present their work in multiple modalities

Paul Leather, NH

Linda Darling Hammond
Issues

**Design**
- State’s Role
- Alignment
- Integration into State Accountability
- Opportunity to Learn

**Resource**
- Money
- Time
- Capacity:
  - State
  - Vendors
  - Field
  - Technology Platform
  - Political Will

**Measurement**
- Validity
- Reliability
- Timeliness of Results
- Comparability
- Scalability
Ohio Pilot Programs

• Ohio Performance Assessment Pilot Project (Phase 1)

• Ohio Performance Assessment Pilot Project (Phase 2) See: The Task Dyad Learning System: Concept and Experiments in Technology Assisted Performance Assessment, by Moore, Monowar-Jones, Xu

• New SCALE Project (Competency Based Education Pilot)

• STEM School Performance Assessment
Join the Conversation

@OHEducation

OHEducation

@OHEducationSupt

OhioEdDept
DISCUSSION – Q&A
Thank you!