

Next Generation Assessment Systems

Scott Marion
Center for Assessment

July 2017 New York Regents Retreat



July 17, 2017

Concerns About Current Testing

- ✓ We've over-promised what our tests can do
- ✓ We're over-testing because of an incoherent Babel of state and local 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



Focus of Discussion

- Stakeholders, purposes, and uses
- Systems of assessment
- Innovative assessments
- A process for moving forward

Purposes, Uses, and Users

Purposes/Uses

- Accountability
- Monitoring Equity
- Instruction/learning
- Grading
- Program/curricular evaluation

Context and users

- State policy leaders
- District leaders
- District CIA leaders
- Principals
- Teachers
- Students
- Parents

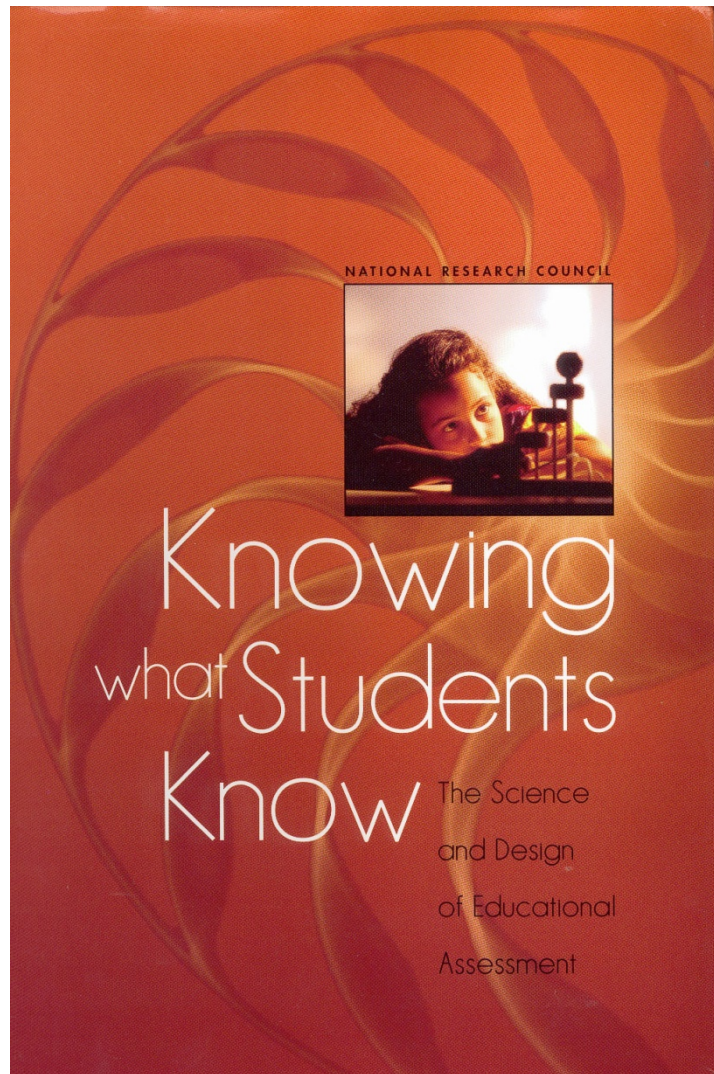
Assessments must be designed to support well-defined **purposes** and intended **uses**.

Assessment Design Involves Tradeoffs

A key trade-off in current assessment design: **Accountability** versus **instructional** support and improvement for individual students

“Ironically, the questions that are of most use to the state officer are of the least use to the teacher.” Pellegrino, Chudowsky, & Glaser (2001)

Why? Timing, grain size, connection to taught curriculum...



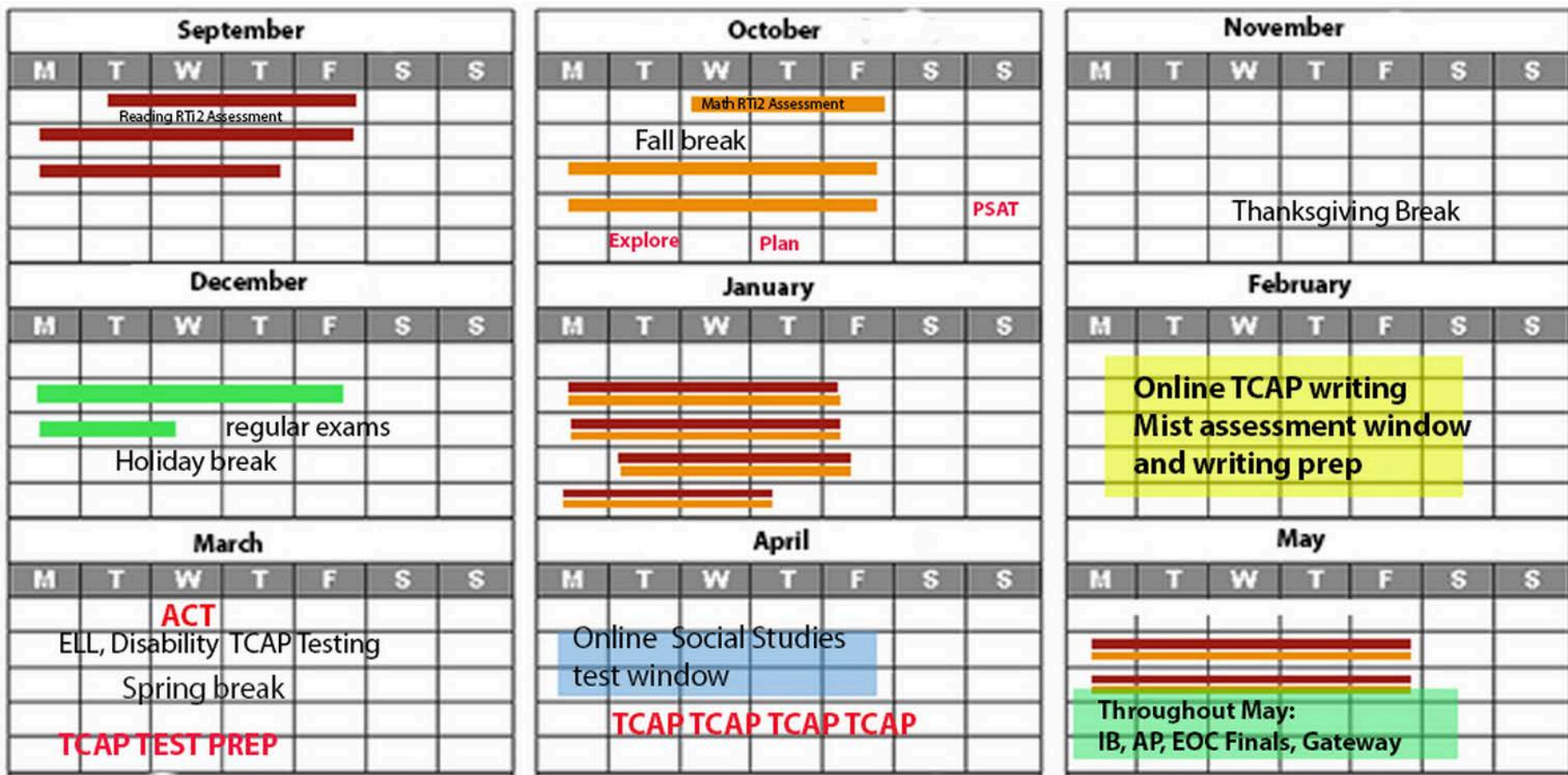
A Call for Assessment Systems...

- The differing purposes and intended uses of large-scale and classroom level assessments make clear that **different assessments** are needed
 - standardized vs. dynamic/flexible
 - uniform vs. variable dates
 - independent vs. assisted (scaffolded) performance
 - delayed vs. immediate feedback
 - stringent requirements for technical accuracy vs. less stringent requirements

How do we keep these multiple assessments from becoming incoherent and inefficient?

Uncoordinated and Incoherent Assessments

Why? Different users, different purposes, lack of common learning model...



Balanced Assessment Systems to Serve Multiple Purposes



- Since *Knowing What Students Know* (Pellegrino, et al., 2001), we've seen increasing calls for Balanced Assessment Systems
 - Coherent
 - Comprehensive
 - Continuous
- Assessment systems designed to serve multiple purposes require thoughtful planning about which data will be privileged at each level (Chattergoon & Marion, 2016).

Who's Responsible for Achieving Balance?

Microcosm:
Classroom-Level Assessments



Macrocosm:
System of Assessments



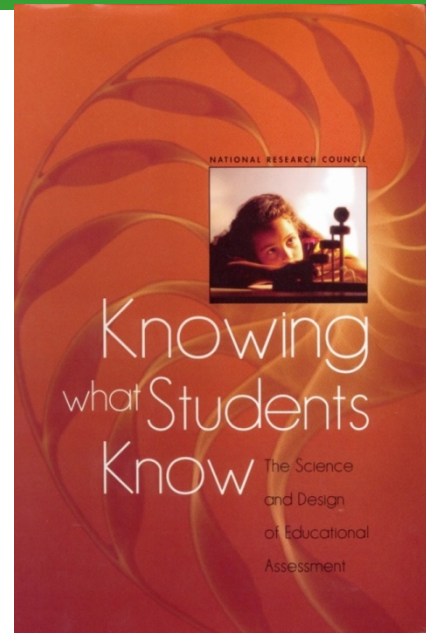
What's the Glue?

Building assessments on an *assessment triangle* requires:

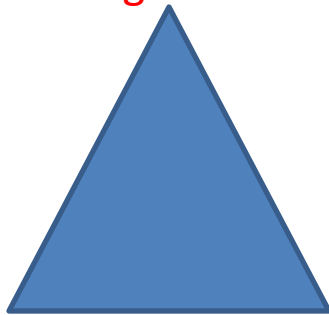
- ***A model of student cognition and ways of developing competence in a domain,***
- ***tasks for eliciting/observing,***
- ***& interpretation processes.***

To support learning, assessment systems must be ***coherent: Vertically*** between classroom and large-scale, and ***horizontally among curriculum, instruction and assessment.***

Models for instructional guidance must be much more fine-grained than for accountability tests.



Cognition



Interpretation

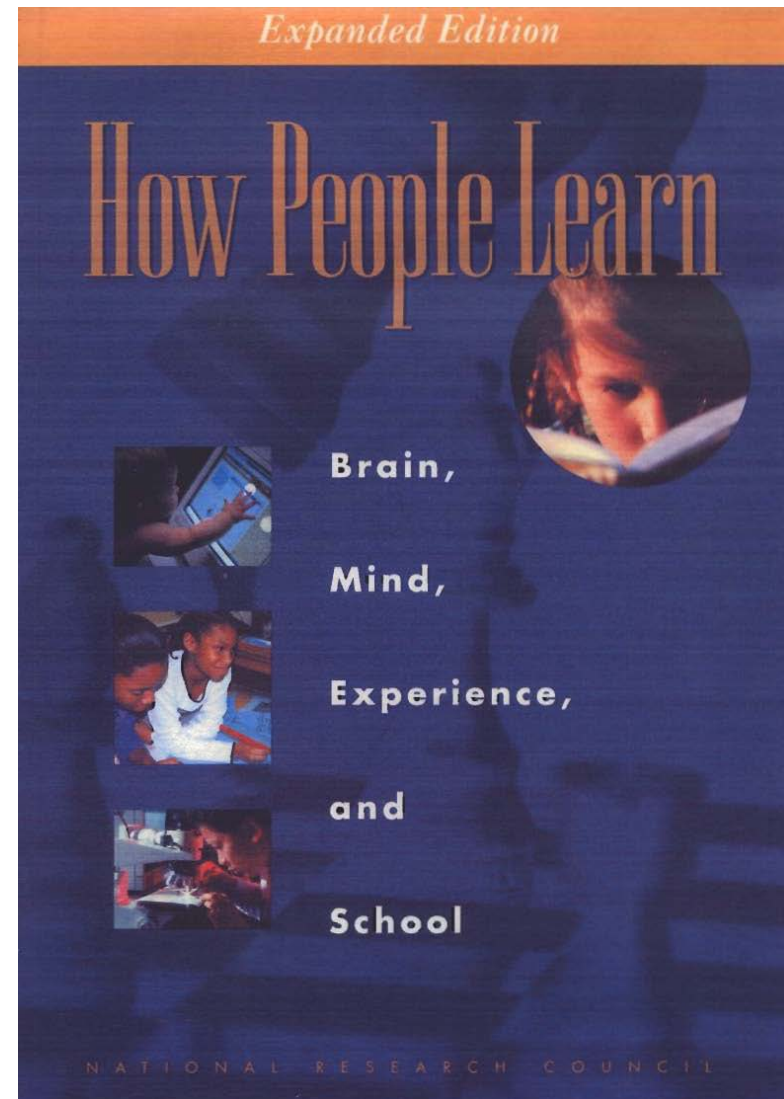
Observation

Not Just Any Model of Learning

Assessments and assessment systems must be based on **research-based models of learning**

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

Bransford, Brown, Cocking (Eds.). (1999). [How People Learn: Brain, Mind, Experience, and School](#). National Research Council (in the process of being updated).



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 by the US Department of Education (USED) 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.
 - Four NH districts in Year 1, 9 in Year 2, 32 in Year 3
- Led by the NH DOE in close partnership with the district leads and the Center for Assessment

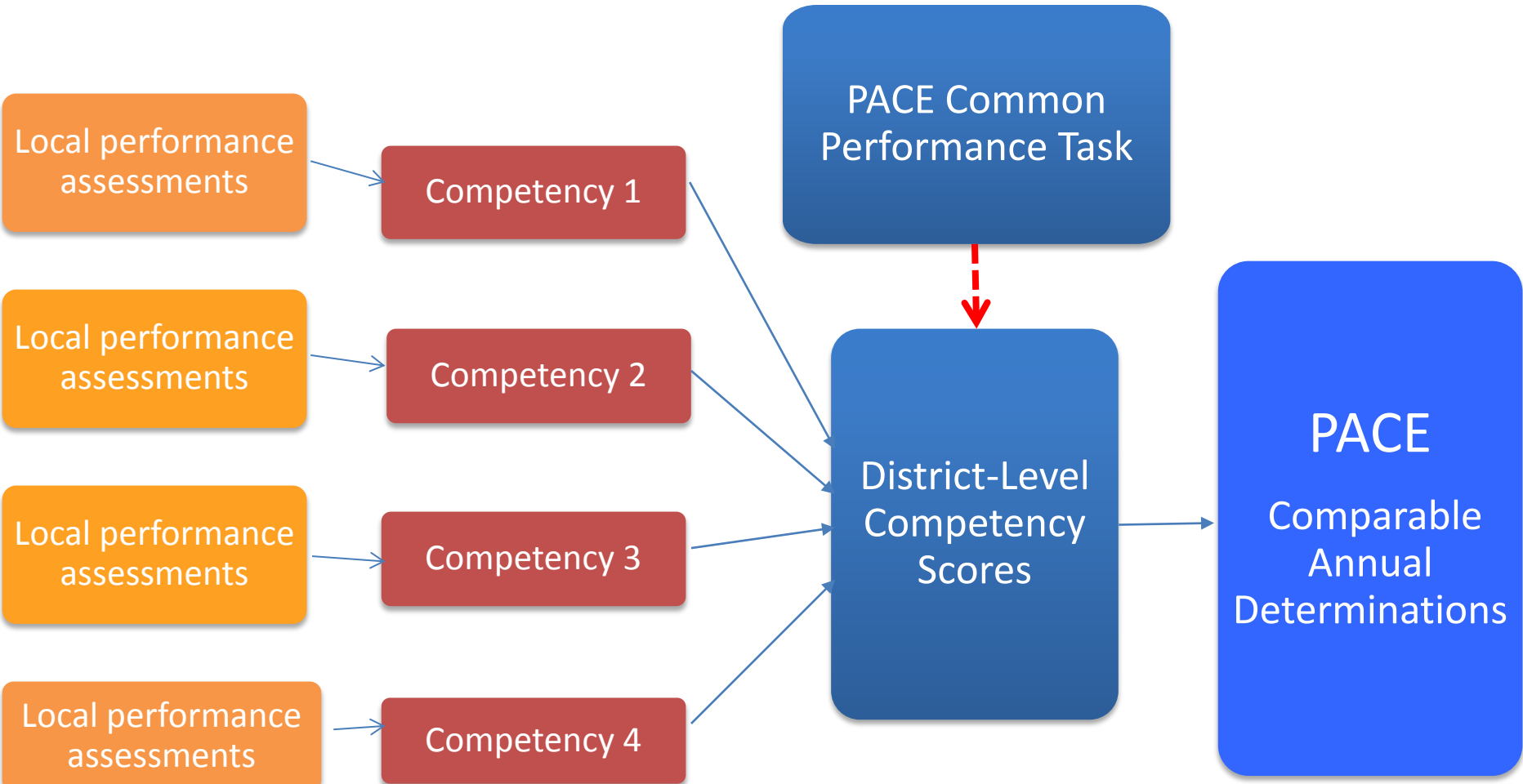
PACE as a “re-Balanced” Assessment System



- 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
- Requires additional resources and intense capacity building

The PACE Assessment System

State summative assessment in select grades



Supporting Deeper Learning for Students

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

- **Modern theories of learning** make clear that developing **deep understanding** is necessary to facilitate **transfer**.
- Students cannot develop deep understanding unless they are provided **multiple and varied opportunities** with both **learning** and **assessment tasks**.



PACE Example – Water Tower Proposal

HS Geometry PACE Common Task

- **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 $45,000 \pm 2,000$ 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.



PACE Example – Middle School Solar Cooker

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.

How to move forward to a plan...

- Assessment is highly **political** and **visible**
- Broad-based surveys help gather stakeholder opinions, but it is often necessary to turn to a deliberative body to wrestle with the difficult choices (optimization under constraints)
- Many states have turned to ad hoc committees (e.g., Assessment Task Force) to advise policy makers
 - Includes various types of educators from different types of school systems, higher education, business, politics, parents, and others
 - For example, see this [report](#) from Wyoming that was used to guide the recent RFP.



NY Regents Retreat

July 17, 2017

Building Systems of Assessments that
Support Deeper Learning

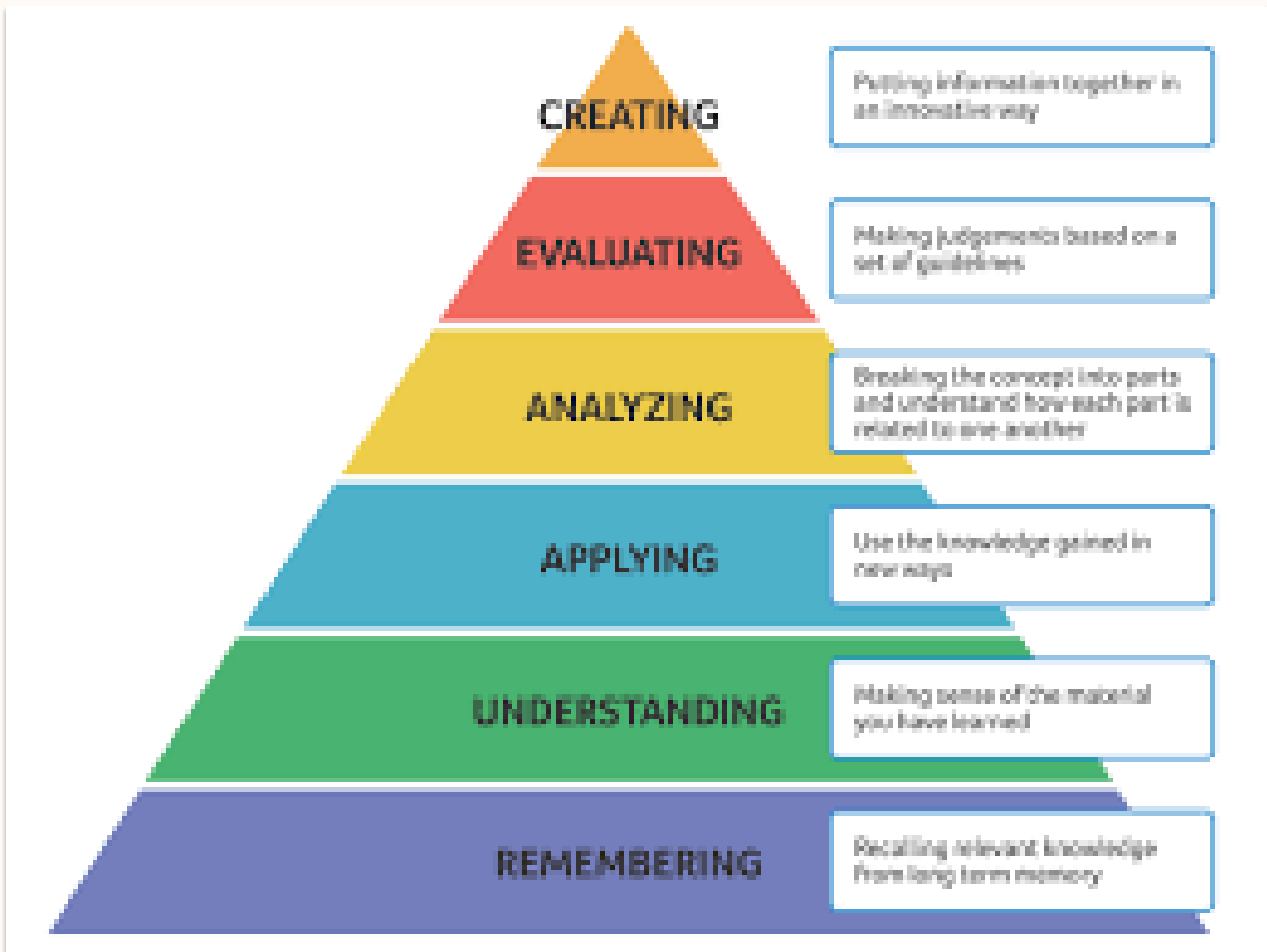
Goal: Assessment of, as, and for Learning

Assessment tools and systems are designed to continuously improve teaching and learning.



ESSA (2015) Testing Changes

- Tests must include “multiple up to date measures of student academic achievement, including measures that assess higher order thinking skills and understanding, which may include measures of student academic growth and may be partially delivered in the form of portfolios, projects, or extended performance tasks”
- Tests may be a single summative assessment or may be “multiple statewide interim assessments that result in a single summative score”
- States may apply for innovative assessment pilots



Bloom's Taxonomy

Assessment Continuum

(All Choices Carry Various Tradeoffs)

Examples

Traditional Tests

CCSS Assessments (SBAC & PARCC)

Performance Based Items & Tasks (MARS, BAM)

Extended Performance Tasks (SCALE, EPIC, ILN)

Student-Designed Projects (Envision, NY Performance Standards Consortium, Singapore, IB)

Assessments of Deeper Learning

Descriptions

Standardized, multiple-choice tests of routine skills

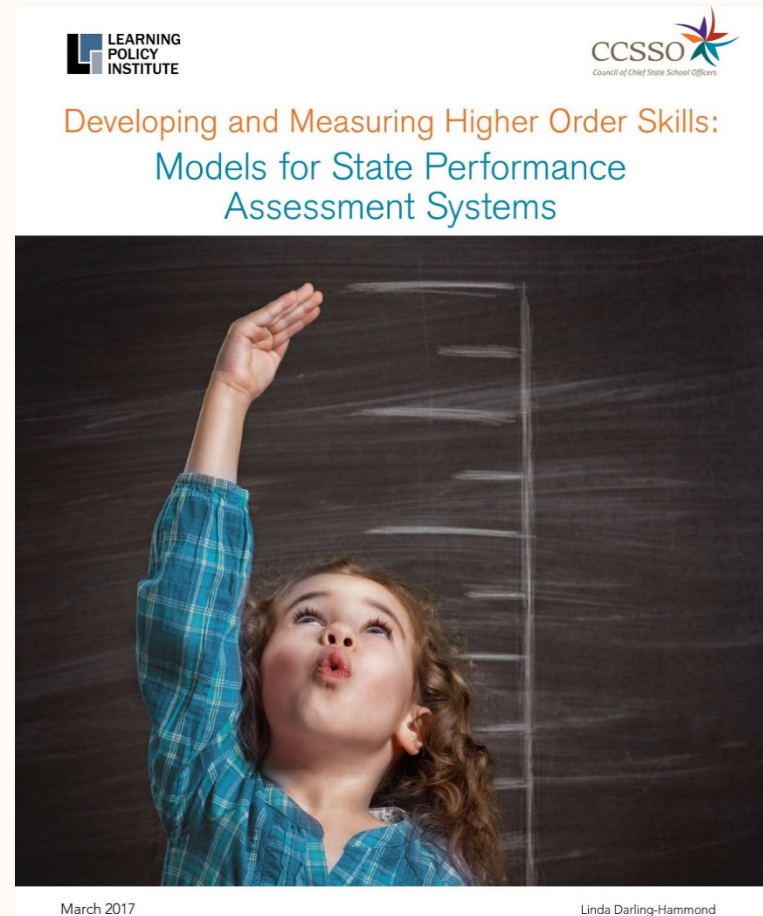
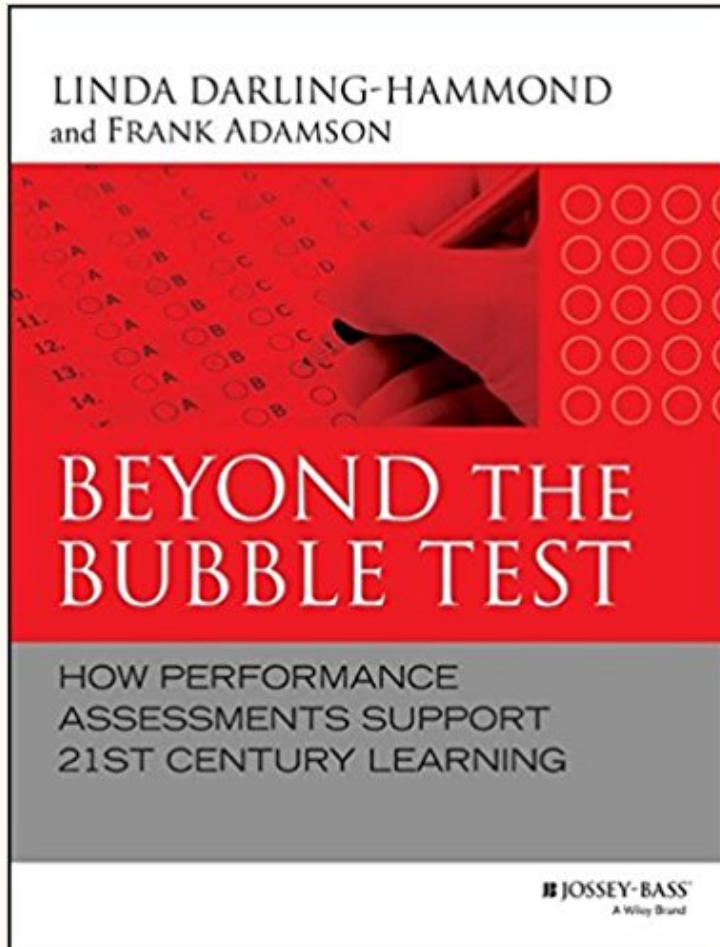
Standardized tests with m-c & open-ended items + short (1-2 day) performance tasks of some applied skills

Systems of standardized performance items and tasks (1 day to 1 week) that measure key concepts in thought-provoking items that require extended problem solving

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

Building on What We've Learned



Potential Assessment Design Options

A comprehensive system that incorporates standardized tests, local and common tasks, plus exhibitions in an integrated system.

- ***Performance items or tasks*** as part of traditional 'sit-down' tests.
- ***Curriculum-embedded tasks*** that take place in the classroom over days or weeks.
- ***Portfolios*** that collect multiple tasks demonstrating skills in one or more subjects.

1. Performance Items on Tests

- Essays
- Document-Based Questions
- Simulations
- Problem Solutions
- Research Tasks

NY Regents Exams are examples

2. Curriculum-Embedded Tasks

- Implemented in the classroom during school year.
- May be common tasks or locally developed
- May produce scores or be combined with test results to produce a summative score.
- Common Tasks + End of Year Tests are used in many countries + IB and, now, AP. Performance tasks = 20-60% of total summative score
- NY 35% option for Regents tests during early '90s was a local example. [See pp. 16-24 for others.]

Singapore GCE Examinations

Time-based Written Papers

- 3 hour duration
- Open-ended essays, structured questions, case studies, source-based questions
- Externally set and marked by SEAB/CIE

School- based Coursework

- Longer duration, weeks or months
- Product (e.g. Science investigation; artwork; or design task), Oral Presentation, Independent Study
- Tasks set by SEAB/CIE, internally marked by teachers, externally moderated by SEAB/CIE)

SCHOOL-BASED SCIENCE PRACTICAL ASSESSMENT

To Assess Experimental Skills, Students...

- Identify a problem, design and plan an investigation, evaluate their methods and techniques
- Follow instructions and use techniques, apparatus and materials safely and effectively
- Make and record observations, measurements, methods, and techniques with precision and accuracy
- Interpret and evaluate observations and experimental data



An Assessment Plan for Science

Sample Science Assessment Plan		
Grades K-2		Locally-selected/designed performance tasks
Grade 3		Locally-selected/designed performance tasks
Grade 4	Innovative Science Test	Common curriculum-embedded science inquiry
Grade 5		Locally-selected/designed performance tasks
Grade 6		Locally-selected/designed performance tasks
Grade 7	Innovative Science Test	Common curriculum-embedded science inquiry
Grade 8		Locally-selected/designed performance tasks
Grade 9		Locally-selected/designed performance tasks
Grade 10	Innovative Science Test	Common curriculum-embedded science inquiry
Grades 11-12		Capstone science investigation (local)


The CCSSO / SCALE Performance Assessment Task Bank

- By Networks
 - ILN Task Bank
- By Category
 - Research and analysis
 - Experimentation and evaluation
 - Writing
 - Oral communication
 - Use of technology
 - Collaboration
 - Modeling & design

All Network Content 1 2 >

All (75) **Learning Modules (31)** Tasks (31) Scoring Guides (13)


Maximum Volume



Learning Module
Mathematics
Rating: ★★★★★

Innovation Lab Network


Becoming the Bard: Rewriting a Scene fro...



Learning Module
Language Arts
Rating: ★★★★★

Innovation Lab Network


Planning a Dash



Learning Module
Language Arts
Rating: ★★★★★

Innovation Lab Network

Your Earthquake and Volcano Risk



Learning Module
Sciences
Rating: ★★★★★

Innovation Lab Network

Pilot Teacher Feedback

“Students enjoy completing performance tasks much more than taking a multiple choice test. They can show their thinking and see what other classmates produce. They enjoy being challenged and want those opportunities.”

Washington State Civics Classroom-Based Assessment

High School Recommended for 11th Grade - Constitutional Issues CBA

Citizens in a democracy have the right and responsibility to make informed decisions. You will make **an informed decision on a public issue** after researching and discussing different perspectives on this issue.

Directions to Students: In a cohesive paper or presentation, you will:

State a position on an issue that considers the interaction between individual rights and the common good AND includes an analysis of how to advocate for your position.

- Provide reason(s) for your position that include:
 - **An analysis of how the Constitution promotes a specific ideal or principle logically connected to your position on the issue.**
 - **An evaluation of how well the Constitution was upheld by a court case OR a government policy related to your position on the issue.**
 - **A fair interpretation of a position on the issue that contrasts with your own.**
- Make explicit references within the paper or presentation to three or more credible sources that provide relevant information AND cite sources within the paper, presentation, or bibliography.

3. Portfolios / Collections of Evidence

Single Subject

- Writing (KY, VT, England GCSE)
- AP Art, Technology, Research, Seminar

Multiple Subject

- Graduation Portfolios (RI, NH, WA, NY Performance Standards Consortium)

Kentucky Writing Portfolio

As part of KY's reform in the 1990's...

Specific tasks w/ common rubrics to measure:

- Reflective Writing
- Expressive Writing/Literary Writing
- Transactive Writing



England's General Certification of Secondary Education (English)

Unit and Assessment	Tasks
<p><i>Reading literacy texts</i> Controlled assessment (coursework) 40 marks</p>	<p>Responses to three texts from choice of tasks and texts. Candidates must show an understanding of texts in their social, cultural and historical context</p>
<p><i>Imaginative Writing</i> Controlled assessment (coursework) 40 marks</p>	<p>Two linked continuous writing responses from a choice of Text Development or Media</p>
<p><i>Speaking and Listening</i> Controlled assessment (coursework) 40 marks</p>	<p>Three activities: a drama-focused activity; a group activity; an individual extended contribution. One activity must be a real-life context in and beyond the classroom</p>
<p><i>Information and Ideas</i> Written exam 80 marks (40 per section)</p>	<p>Non-Fiction and Media: Responses to unseen authentic passages</p> <p>Writing information and Ideas: One continuous writing response – choice from 2 options</p>

Summary: transcript, GPA, test scores, statement of goals, distinctive accomplishments or "badges," short essay, 2-minute video clip from portfolio presentation, table of contents

Science &
Math
Inquiry



Investigation of climate change trends in a local community (science and mathematics), includes paper, data set, and PowerPoint

Social
Science
Inquiry



What social and political forces influenced the passage of the 14th Amendment to the Constitution? (historical inquiry)

Literary
Analysis



The American Dream in 20th century literature (literary analysis), includes videotaped presentation to panel

World
Language
Exhibition



Demonstration of competence in world language : Tamil (audiotaped conversation and paper)

Graduation Portfolio

4. A Comprehensive System (NH)

Grade	ELA	MATH	SCIENCE
K-2	Local PBAs	Local PBAs	Local PBAs
3	Smarter Balance	Common PACE PBA	Local PBA
4	Common PACE PBA	Smarter Balance	Common PACE PBA
5	Common PACE PBA	Common PACE PBA	Local PBA
6	Common PACE PBA	Common PACE PBA	Local PBA
7	Common PACE PBA	Common PACE PBA	Local PBA
8	Smarter Balance	Smarter Balance	Common PACE PBA
9	Common PACE PBA	Common PACE PBA	Common PACE PBA
10	Common PACE PBA	Common PACE PBA	Common PACE PBA
11	Smarter Balance → SAT	Smarter Balance → SAT	Common PACE PBA
12	Capstone Project / Portfolio that is Exhibited		

Standardized Tests (with Performance Components)

Performance-Based Assessments / Portfolios

Used to validate local assessment results

Used to enrich test results and inform teaching

Interactive Elements of a Comprehensive Assessment System



Considerations

- What skills will students develop?
- What skills will teachers develop and need?
- How can tasks and rubrics be designed with quality to support validity & reliability?
- How can assessments be administered to accommodate diverse students and support common inferences about learning?
- How can NY build on its long and varied experiences with performance assessment?