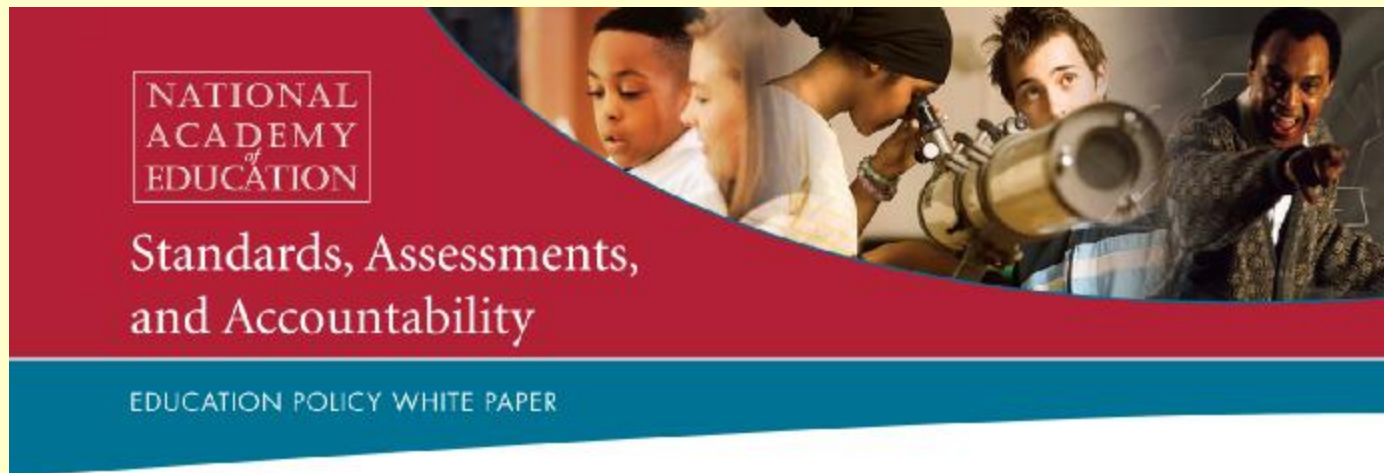


Research Priorities for Next-Generation Assessment Systems

Lorrie A. Shepard
University of Colorado at Boulder

BOTA Workshop: Best Practices in State Assessment
Washington, DC
December 11-12, 2009



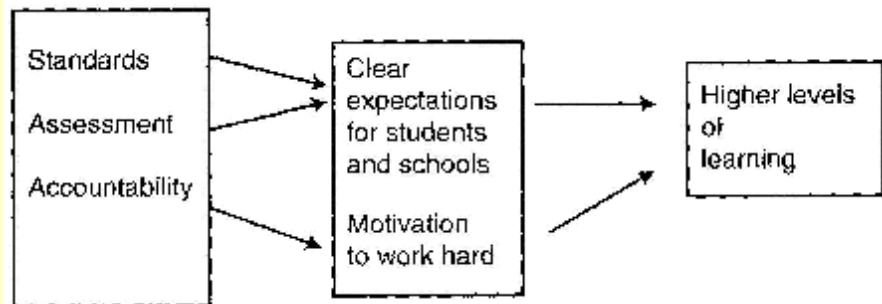
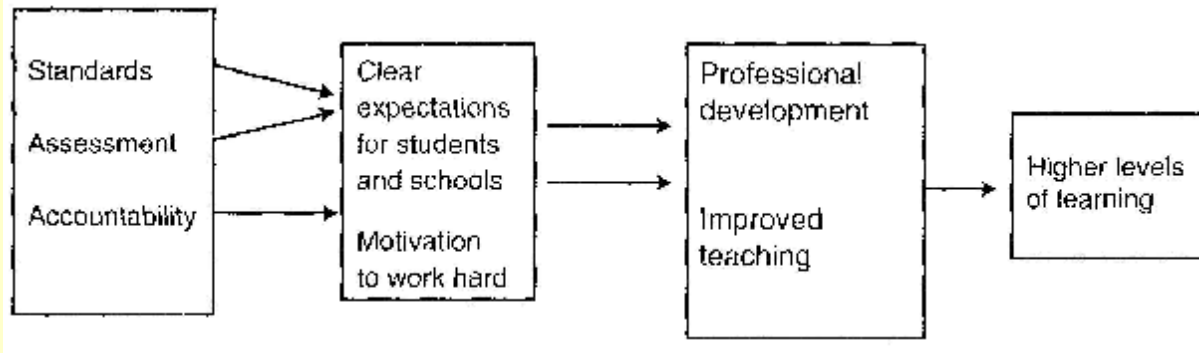
Standards, Assessments, and Accountability

Shepard, Hannaway, & Baker (2009), <http://www.naeducation.org/>

**National Academy of Education Working Group
on Standards, Assessments, and Accountability:**

Eva Baker (Co-Chair)	P. David Pearson
Jane Hannaway (Co-Chair)	Diane Ravitch
Patricia Gandara	William Schmidt
Drew Gitomer	Alan Schoenfeld
Margaret Goertz	David Stern
Helen Ladd	William Trent
Robert Linn	Mark Wilson

Bait-and-Switch Theories of Action



Elmore & Rothman (1999), *Testing, Teaching, & Learning*, NRC

1994 Goals 2000: vision of coherent capacity building

1994 ESEA set forth an incentives *theory of change* – assumed that with sufficient motivation, teachers and other school personnel would find the means to improve instruction.

Unfortunately, early implementation studies showed that many schools did not understand the changes that were needed and lacked capacity to make them happen.

Current Policy Themes

- Ø Reforming assessments with conceptually rich tasks
- Ø Integrating 21st century skills & academic content
- Ø Creating coherence between large-scale & classroom assessments
- Ø Using data to improve classroom instruction

Integrating 21st Century Skills and Academic Content

“Thinking skills are not content free!”

“Don’t save problem solving for Friday!!”

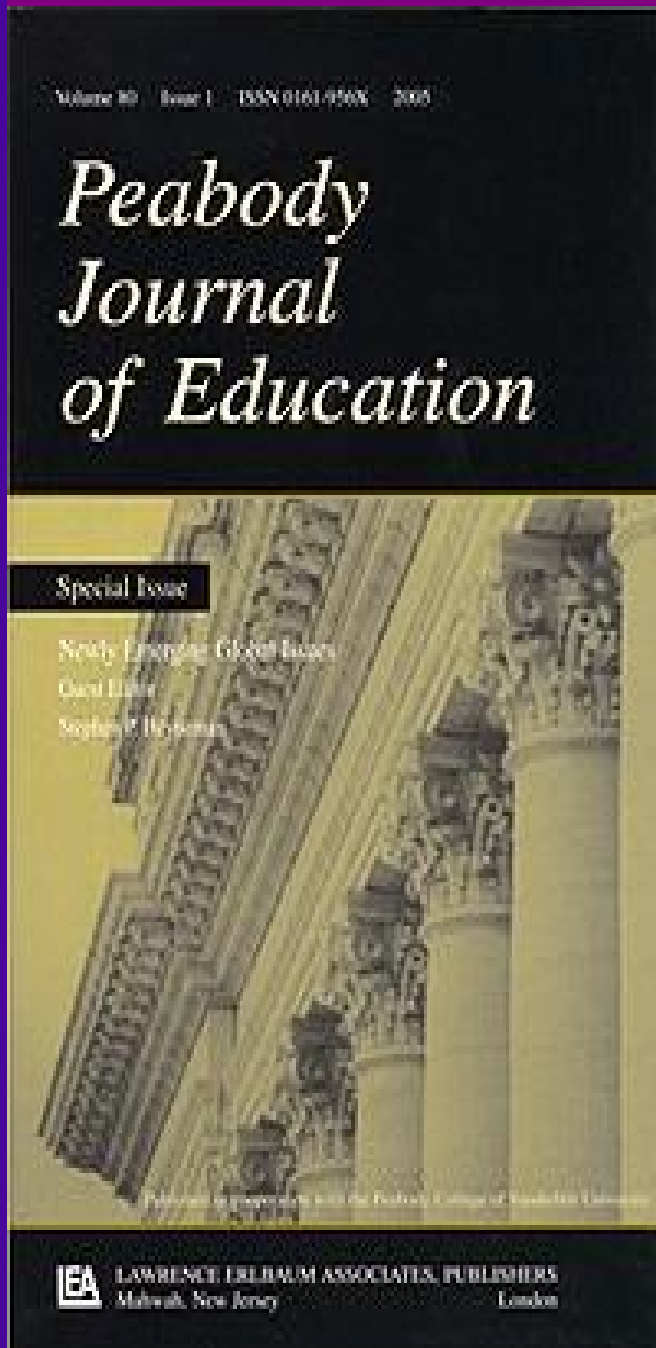
- **Both reasoning and learning are knowledge-driven. Those who are knowledge-rich reason more profoundly. They also elaborate as they study and thereby learn more effectively. (Glaser, 1984)**
- **Relevant knowledge helps people go beyond the information given and to think in problem representations, to engage in the mental work of making inferences, and to relate various kinds of information for the purpose of drawing inferences. (How People Learn, 1999)**

Without an adequate theory for teacher learning, interim assessments exacerbate the negative history of teaching the test.

- Ø Test-score inflation
- Ø Curriculum distortion

Many policy makers believe that curriculum narrowing is a necessary trade off because reading & math skills are so essential.

But policy makers are not aware of the educational consequences of test score inflation. Excessive drill and practice on worksheets resembling the test denies students opportunity to understand context and purpose that would enhance skill development.



Substantive insights are rare.

Ø Blanc et al.:

“Instructional Communities”

In one school, principal was clearly the moral and organizational star.

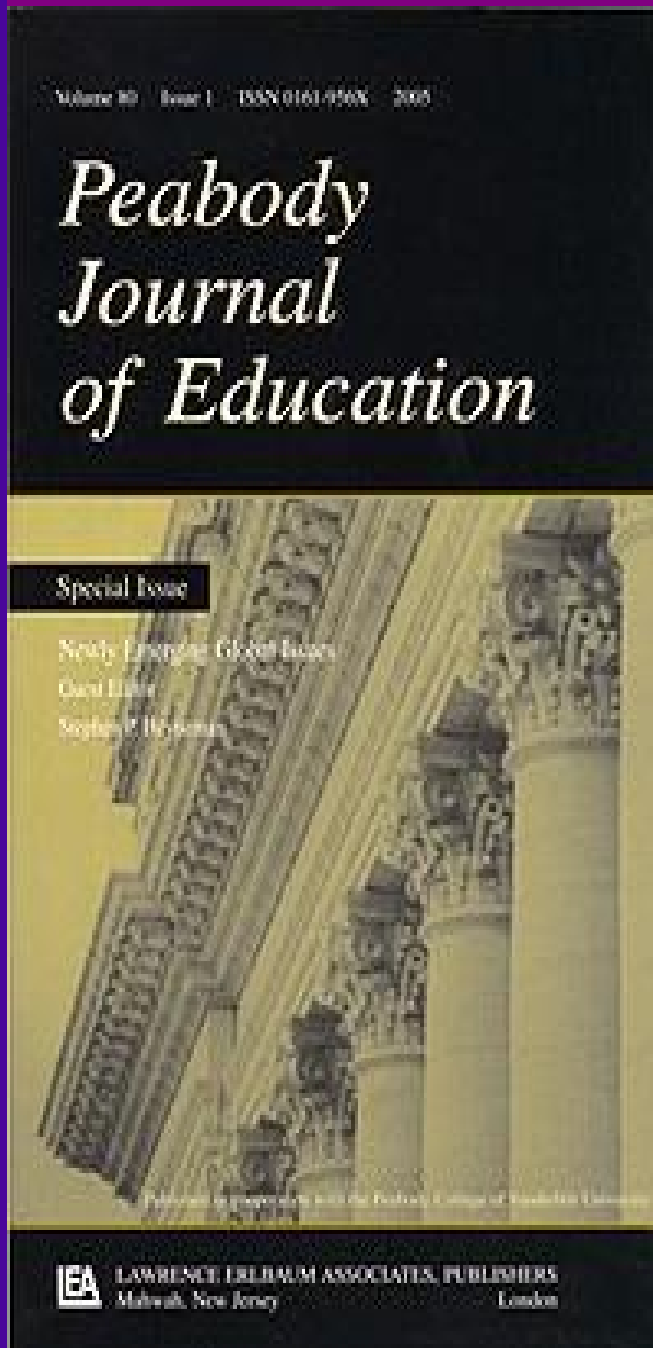
Coherent emphasis on core curriculum helped to link Benchmark results to instructional strengths and weaknesses.

Ø Olah et al.:

“Analyzing Benchmark Data”

Two of 25 teachers connected item and standards data aback to *Everyday Mathematics* curriculum.

Diagnostic insights from items, e.g. problems with regrouping, were rare.



Item-by-item teaching

Ø Bulkley et al.:

“Role of the District”

Assumed that teachers would be able to figure out what to do by looking at the data, talking with other teachers, and drawing inferences about their own needs for professional development.

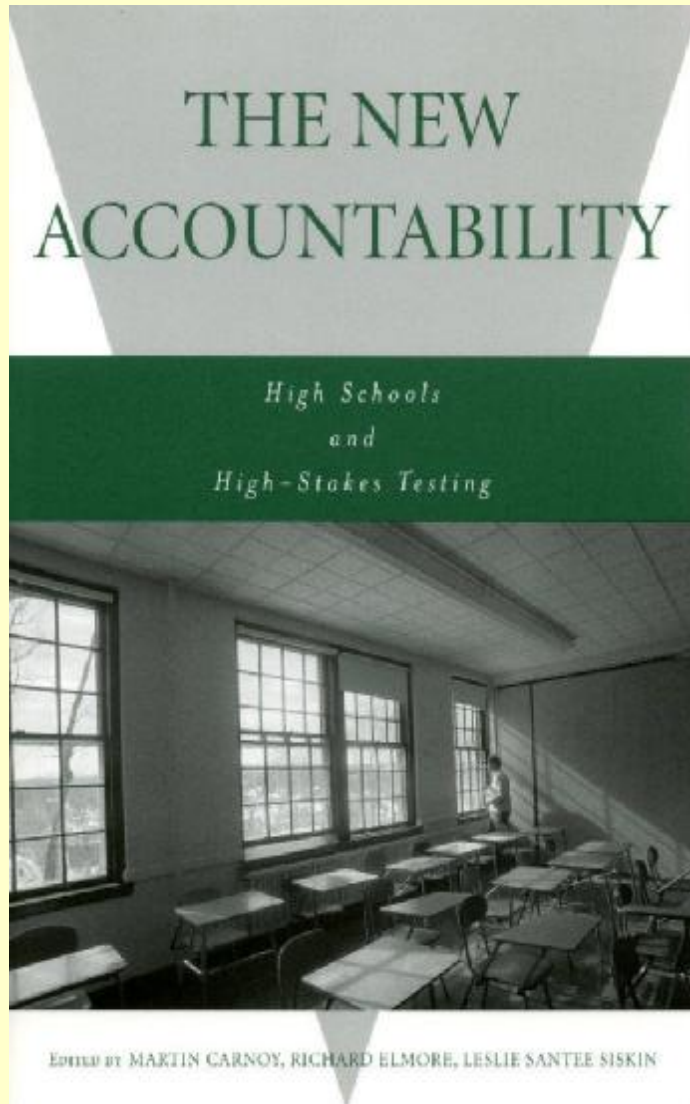
Ø Blanc et al.,

High-stakes state testing talk pervaded grade-level meetings, focus on “bubble kids,” reteach usually with the same instructional strategy, alternative strategies not informed by assessment results.

Ø Olah et al.,

Teachers invented their own thresholds to determine priorities for reteaching. Analyses of items were most often procedural or about item validity. Reteaching was focused procedurally on reteaching items missed, step by step.

Elmore (2003), “Accountability and Capacity”



- Ø Better-situated schools, serving higher SES neighborhoods, are more able to respond coherently to the demands of external accountability.
- Ø Better-positioned schools with relatively high “internal accountability” recognize that increased coherence around instructional practice requires new curriculum content and new knowledge and skills for teachers and administrators – and found ways to change the structure of the work and gain these skills.
 - We did not see evidence in any of our schools of major infusions of professional development for teachers or principals orchestrated from the school system or state level.
 - External influences tended to be in the form of increased sanctions and assumed that the purpose of accountability policy was to mobilize and focus existing knowledge, skill, and resources.



Notes about Politics and Expertise

- § In the U.S., politicians and policymakers are using remote instruments to dictate more about what goes on in schools than occurs in most other countries.**
- § When experts are convened to inform policy makers about assessment, experts are most often psychometricians and policy researchers, but rarely subject-matter experts.**
- § When subject-matter experts do participate, they are usually brought in for one segment of the process and then dismissed. They rarely oversee frameworks, item development, and interpretation of results.**

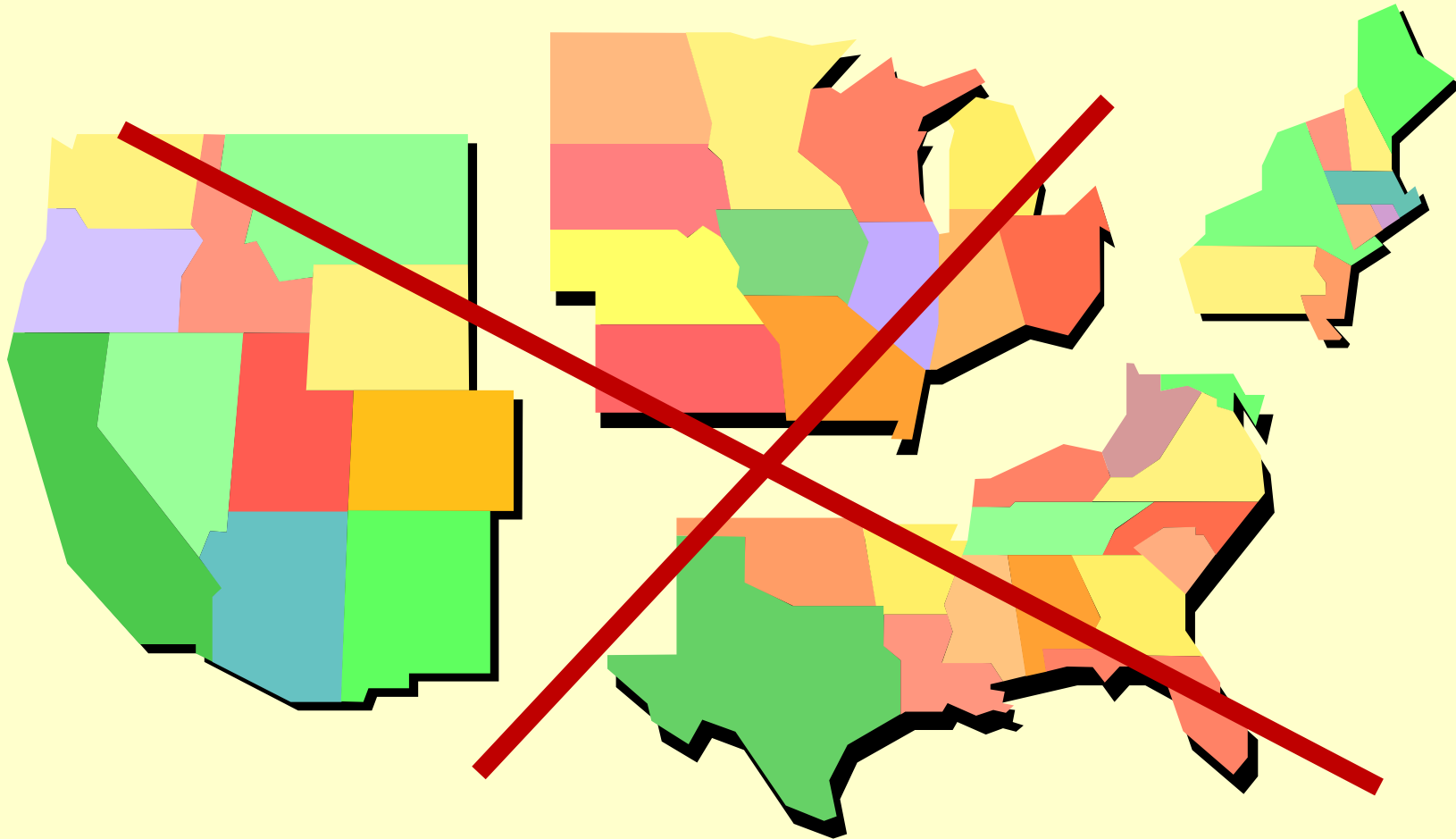
Theory of Action for Assessment Reform RFP

§ Not only do we NOT know enough to establish ONE ideal,

§ Each consortium will not be able to invent and implement ideal systems across all grades and subject areas.

It is important that DOE and grant applicants recognize political and policy obstacles as well as technical, logistical and cost challenges.

- Ø Each consortium grant award should be focused on getting at least one segment of the system to next-generation, high quality implementation (e.g. Math 4-8).**



It is unlikely that the goal of curricular coherence can be reached, if we attempt at the same time to achieve immediate participation by all 50 states.

The Importance of Curriculum

The NAEd *Standards* paper argues for the concurrent, coherent design of curriculum, assessments (both large-scale and classroom level), and teacher professional development.

- Ø Top performing countries in TIMSS have leaner, more hierarchically sequenced curricula leading to progressively more advanced topics and deeper understanding. (Schmidt et al., 2005)
- Ø Recent analysis of content standards in 14 states found they did not focus on big ideas or build from grade to grade. (Porter et al., 2009). Common Standards are a help but are still only a skeletal framework.
- Ø National control is not required for coherence (Schmidt & Prawat, 2006). Rather, *coherence leads to effective outcomes if it is achieved at whatever level of governance has authority over policy instruments.*



Curricula deepen the meaning of standards and provide teachers and students with a roadmap of how to reach proficiency.

Instead of turning Standards over to test makers for implementation, the NRC report on state science assessments calls for:

- Ø Horizontal coherence** (linking curriculum, instruction, and assessment)
- Ø Vertical coherence** (shared vision at classroom, school, district, and state levels)
- Ø Developmental coherence** (taking account of how understanding develops over time)

If RTTT \$ cannot be spent directly on curriculum development (?), DOE could nonetheless require applicants to demonstrate how they will link assessment design to new or existing curricula.

“Agreeing on Scope and Sequence Won’t Be Easy”

Every state has to attend to political processes that have led to everything-but-kitchen-sink standards and wide variation in assessment quality.

∅ Focusing on just one content area in a limited number of grades may help mitigate political difficulties, especially if considered a pilot.

∅ Skill areas such as reading, writing, and scientific inquiry may be easier to agree upon than specific content knowledge.

∅ The problem of agreeing on curriculum could be further softened by developing only a limited number of big-idea instructional units per grade.



Connected Mathematics Project

Connected Mathematics Project (CMP)

Ø Funded by NSF in 1991-1996, 2000-2006

Ø Addresses mathematical thinking and reasoning as well as mathematical concepts, skills, and procedures.

Ø Provides context and resources for teacher professional development.

NSF funding for development of new forms of summative assessments was launched independently, e.g., *Balanced Assessment of Mathematics*. Closer coordination of classroom and large-scale assessment development could directly consider generalization and transfer.

Designing Effective and Coherent Assessment Systems

∅ Formative Assessment

Curriculum-embedded, real-time processes used for feedback, to adapt instruction, and to engage students in self assessment

Substantive insight > numerical scores

∅ Benchmark/Interim Assessments??

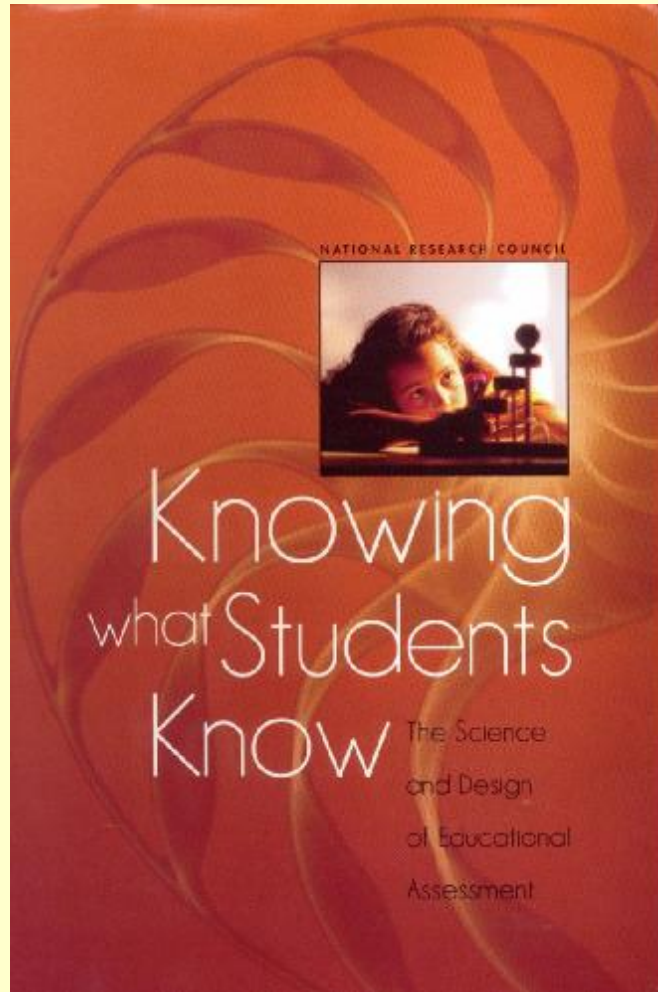
Could be culminating assessments for specific units of study or scored elements for summative accountability, but current examples are of poorer quality than current state assessments.

∅ Summative Assessments

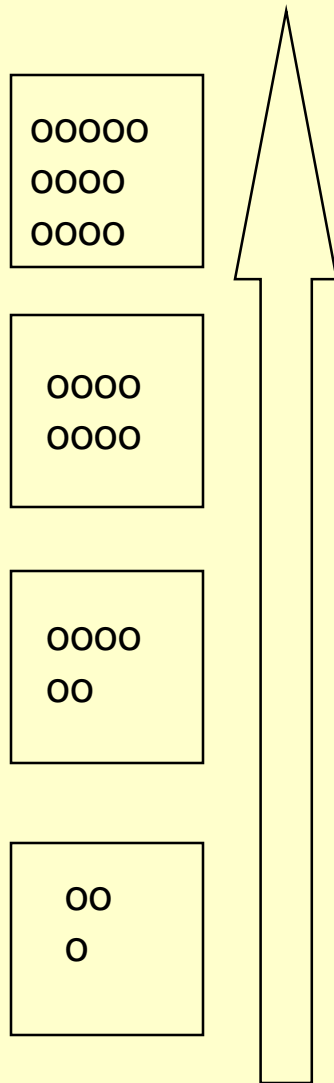
Used for program evaluation and accountability



Learning Progressions *Knowing What Students Know* (2001)

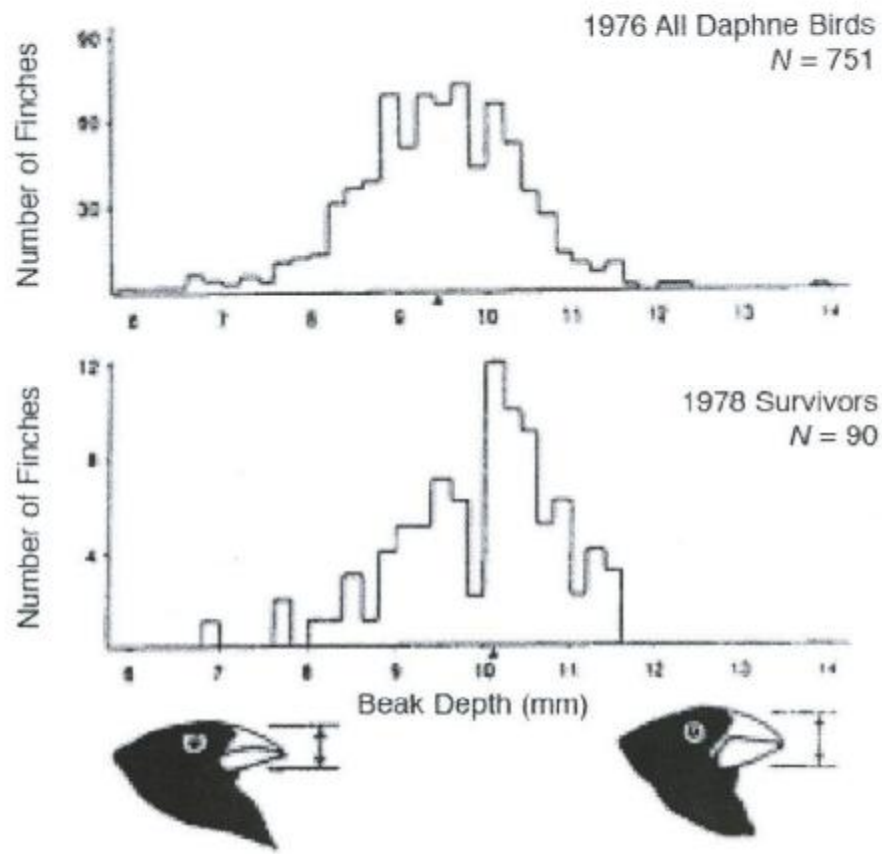


- ∅ Learning progressions or Progress Maps provide an underlying model of learning to coherently link classroom and large-scale assessments.
- ∅ “Progress maps describe skills, understandings, and knowledge in the sequence in which they typically develop: a picture of what it means to ‘improve’ in an area of learning.” (Masters & Forster, 1996).
- ∅ Learning progressions provide a substantive rather than a quantitative picture of developing competence.
- ∅ A criterion-referenced growth model.



- Ø **Substantive and statistical growth trajectories are not the same, but with care they could be brought together.**
- Ø **While much research is still needed, it is possible to start with judgmentally set sequences and then improve empirically.**
- Ø **Learning progressions are an effective tool for teacher professional development.**
 - **How learning typically unfolds helps teachers know “what next” and how to “back-up” (though we must also be aware of natural variations and departures from the typical pattern).**
 - **When embedded in units of instruction LPs can target typical misconceptions.**

Using assessment to tap common misunderstandings

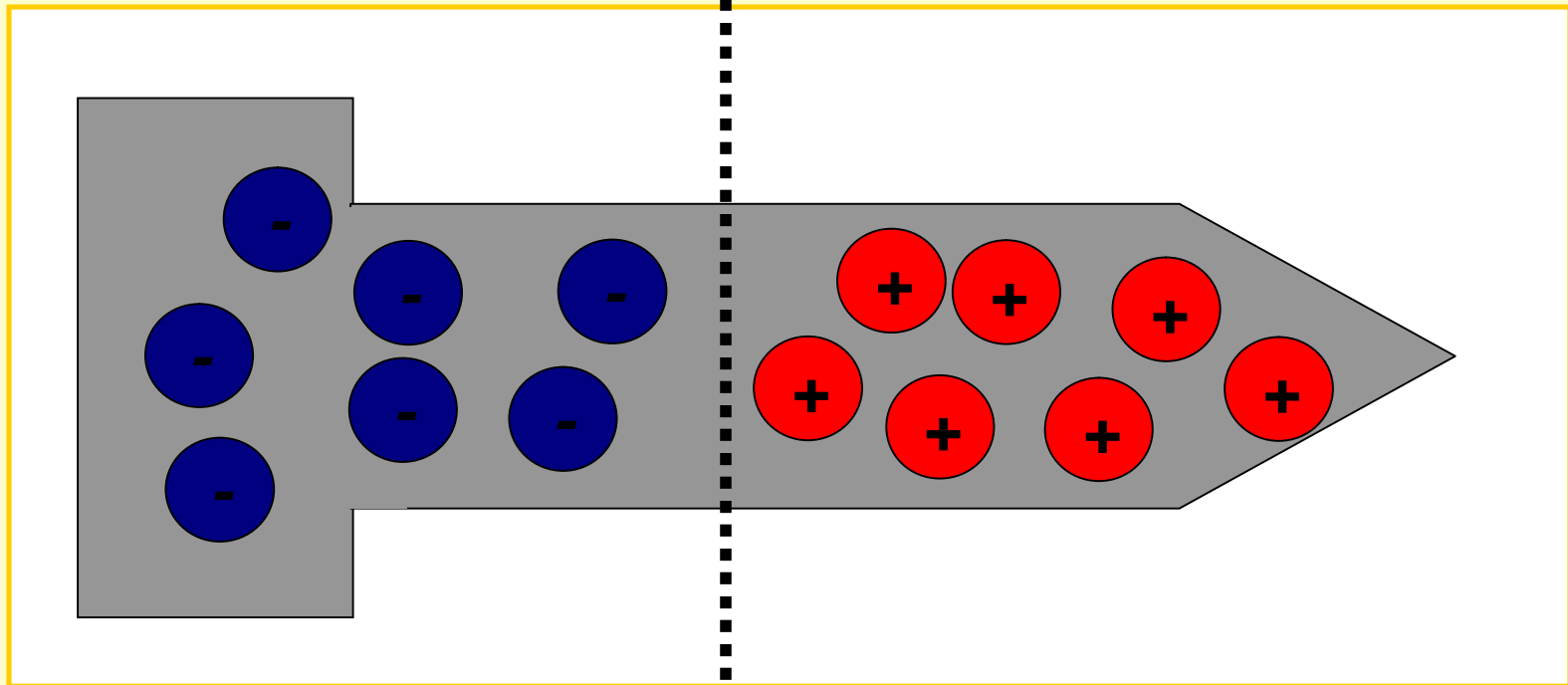


"This demonstrates natural selection because after the draught it shows birds beaks depth getting larger. Since the environment is changing the birds have to adapt to their new surroundings."

Furtak, E.M. (2009).

An example of a typical misconception and challenge question:

What happens if you cut the nail in half?



Otero, Jalovec, & HerManyHorses. (2008).

Teacher Involvement and Teacher Learning: It's Not Just About Scoring



e.g., **Research on Classroom Discourse**

If learning is a process of enculturation, then new classroom interaction patterns and social norms can be developed to make it customary for students to explain their reasoning, challenge one another's ideas, and work together to make sense of mathematical problems and solutions.



Lampert (1990). Shifting classroom participation structures to more closely resemble standards of logical argument in the mathematical community.



Cobb, Wood, & Yackel (1993). Negotiating new social norms to overcome previously constructed norms about trying to guess the teacher's solution and avoid evaluation.



Social-Constructivist and Sociocultural Learning Theories

- ∅ *Mathematical discourse* is a critical feature of math education reforms because it creates the opportunity for students to develop problem-solving and thinking abilities.
- ∅ *Scaffolding* refers to the support – guidance, coaching, hints, and encouragement – that adults provide in the *zone of proximal development* (ZPD) to enable and challenge the learner to perform at a level that she would not otherwise be able to reach.

A shared understanding and ownership of the learning goal, support in reaching the goal, internalization, and taking over of responsibility by the learner are closely parallel features of formative assessment and scaffolding.



Putnam & Borko (2000). Research on Teacher Learning and Teacher Professional Development

We know that teachers are most likely to make and sustain fundamental changes in instructional practice if given:

- time,
- conceptual and strategic support,
- and opportunities to tryout new practices in the context of their own teaching.



Lessons from Research on Teaching

- ∅ Teachers need better access to materials that model teaching for understanding – with extended instructional activities, formative assessment tasks, scoring rubrics, and summative assessments built in.
- ∅ And they need extended support while attempting to use these materials and teach in new ways.

* * * * *

- ∅ Involve expert teacher-leaders in design of assessment tasks and ensure 100% teacher participation in non-burdensome scoring, focused on capacity building and teacher learning.
- ∅ Avoid mistakes of current data systems that focus on scores and item-by-item reteaching and leave teachers to their own devices to improve.

Build in Research and Evaluation

“A national program of evidenced-based assessment development should be launched as quickly as possible. The needed development and evaluative research should be carried out in multiple laboratories and field tryout sites. However, the program as a whole should be overseen by a single agency so as to promote maximum sharing of ideas, potential solutions, and interim results that might be tried out in operational assessments.”

Research Agenda

Measurement Research Questions

- ✓ Basic

- ✓ Applied

(Evaluation research is based on a theory of action, and once sufficient experience is gained on “known side effects” or theories of what can go wrong.)

Teaching and Learning Research Questions

- ✓ Basic

- ✓ Applied



NATIONAL
ACADEMY
of
EDUCATION

Standards, Assessments, and Accountability

EDUCATION POLICY WHITE PAPER

Standards, Assessments, and Accountability **Shepard, Hannaway, & Baker (2009), NAEEd.**

Attempts to remedy some structural elements of the current accountability system most responsible for misdirecting effort:

- Ø Eliminate reporting by percent above proficiency cutscore.
- Ø Eliminate AYP and use “existence proofs” to set aspirational targets attained by at least 10% of targets.
- Ø Use fallible test scores as triggers for further investigation rather than immediate sanctions.
- Ø Hold other levels of the system accountable, not just under-resourced schools.

To truly transform learning opportunities in classrooms, in ways consistent with what we know is possible from research, we will need to remove policy structures especially low-level tests that misdirect effort, provide coherent curricula consistent with ambitious reforms, and take seriously the need for capacity building at every level of the educational system.