Assessment of Science Learning: Living in Interesting Times

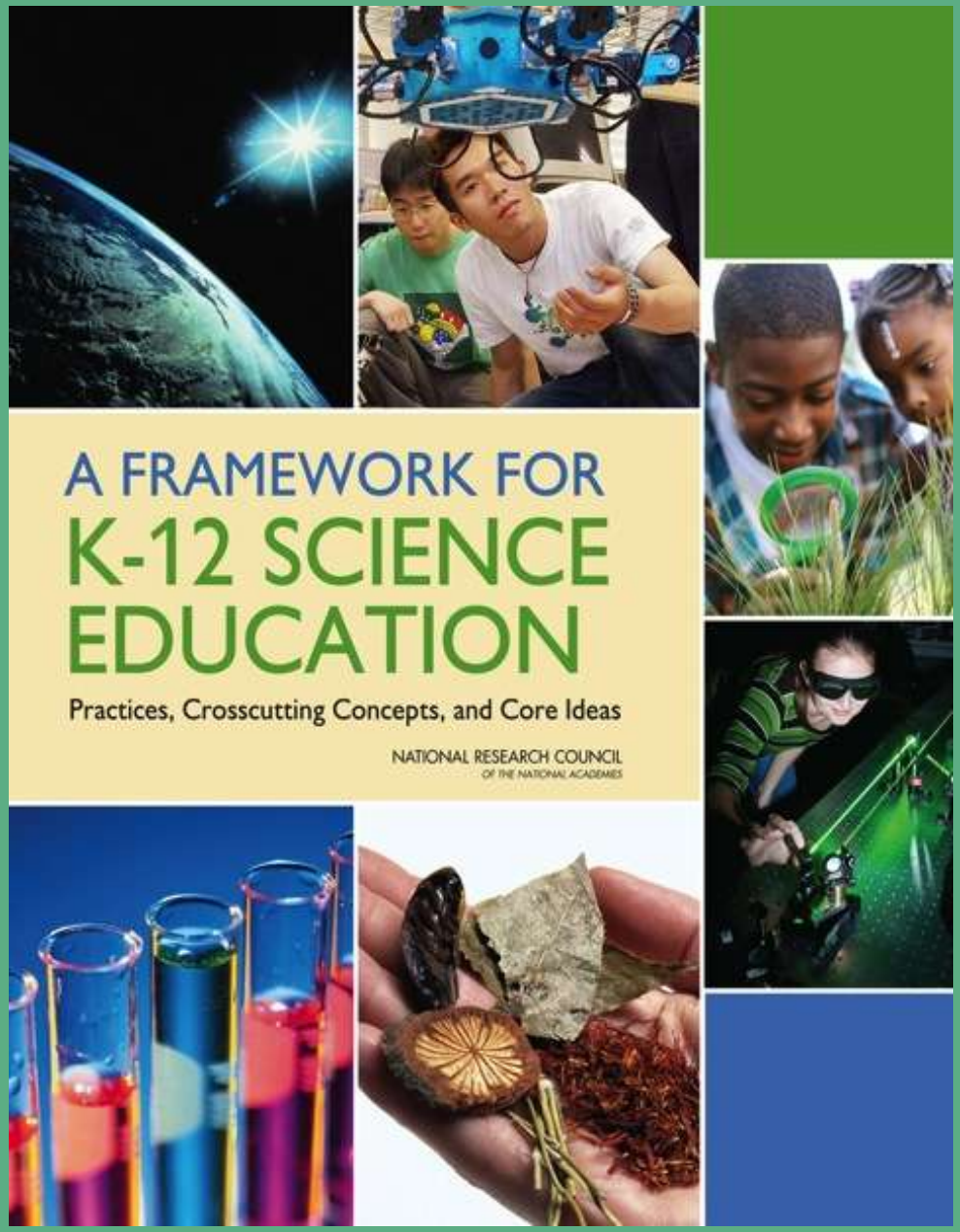
Jim Pellegrino

University of Illinois at Chicago
Learning Sciences Research Institute

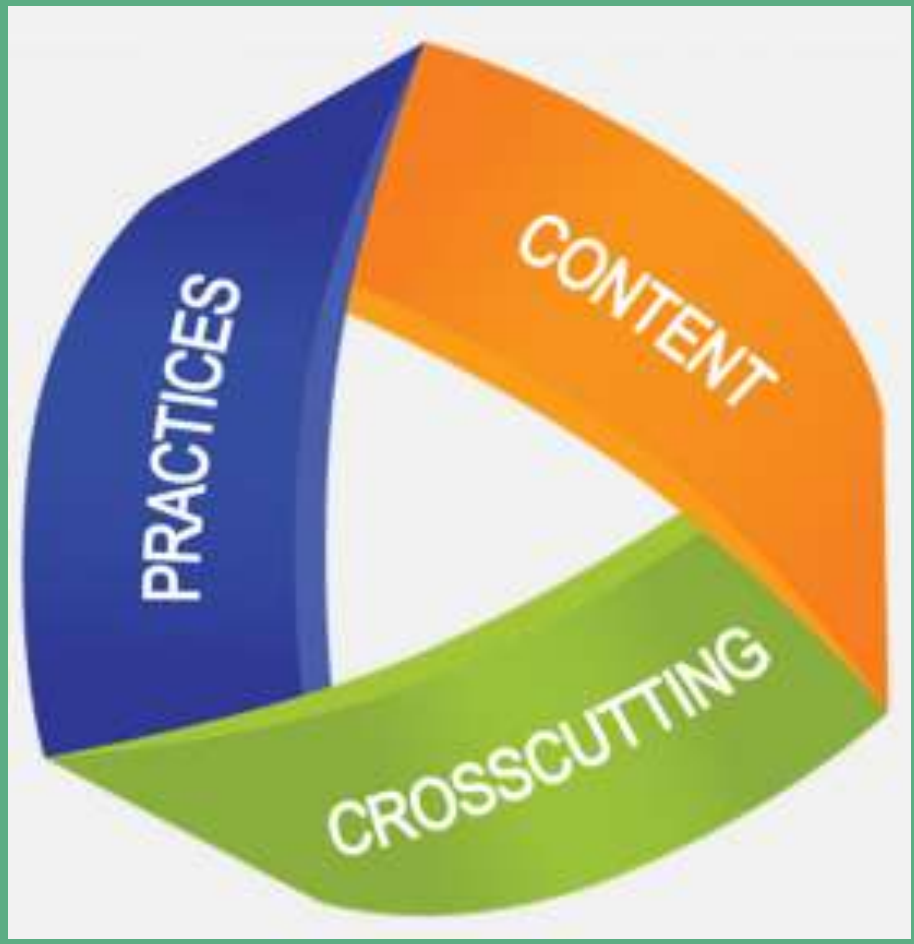
Chinese Curse?

- *There is a Chinese curse which says 'May he live in interesting times.' Like it or not we live in interesting times. They are times of danger and uncertainty; but they are also more open to the creative energy of men than any other time in history*.

  – Robert Kennedy, 1966.



NEXT GENERATION SCIENCE STANDARDS

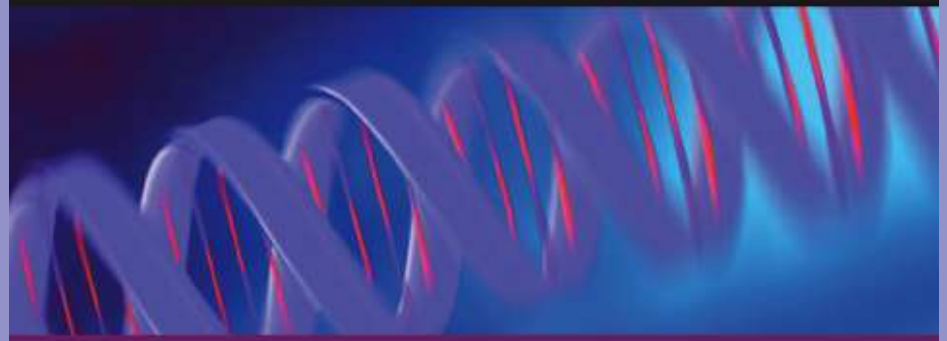


CollegeBoard
inspiring minds

AP® BIOLOGY

Curriculum Framework

2012–2013



CollegeBoard

AP® CHEMISTRY

Curriculum Framework

2013–2014

MCAT Medical College Admission Test | AAMC

MCAT 2015

A Better Test for Tomorrow's Doctors

Preview Guide for the MCAT 2015 Exam (Second Edition)

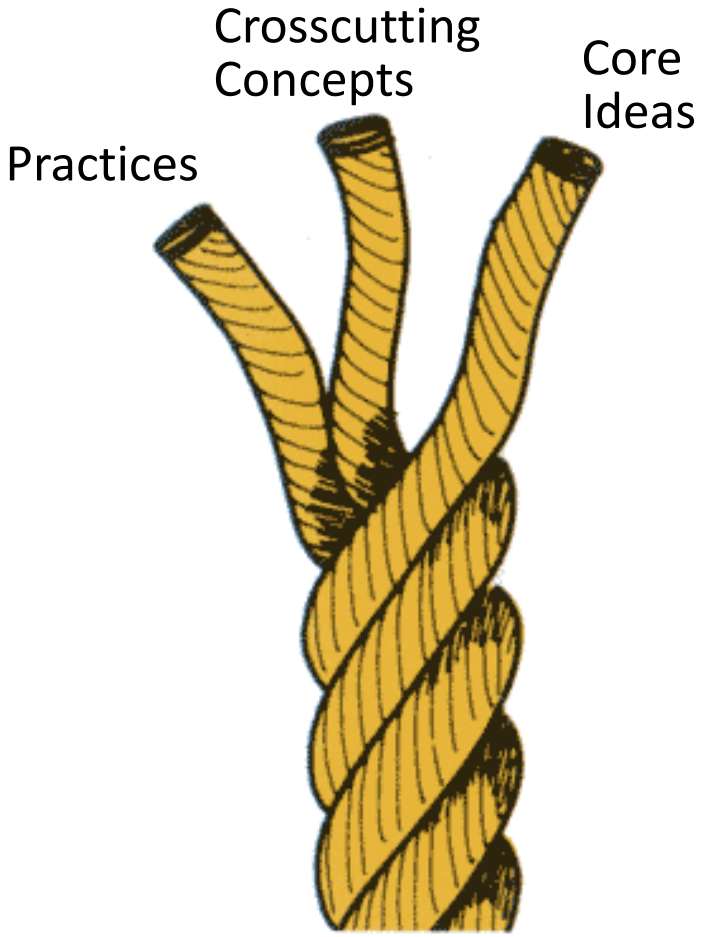
September 2012

New Definition of Competence

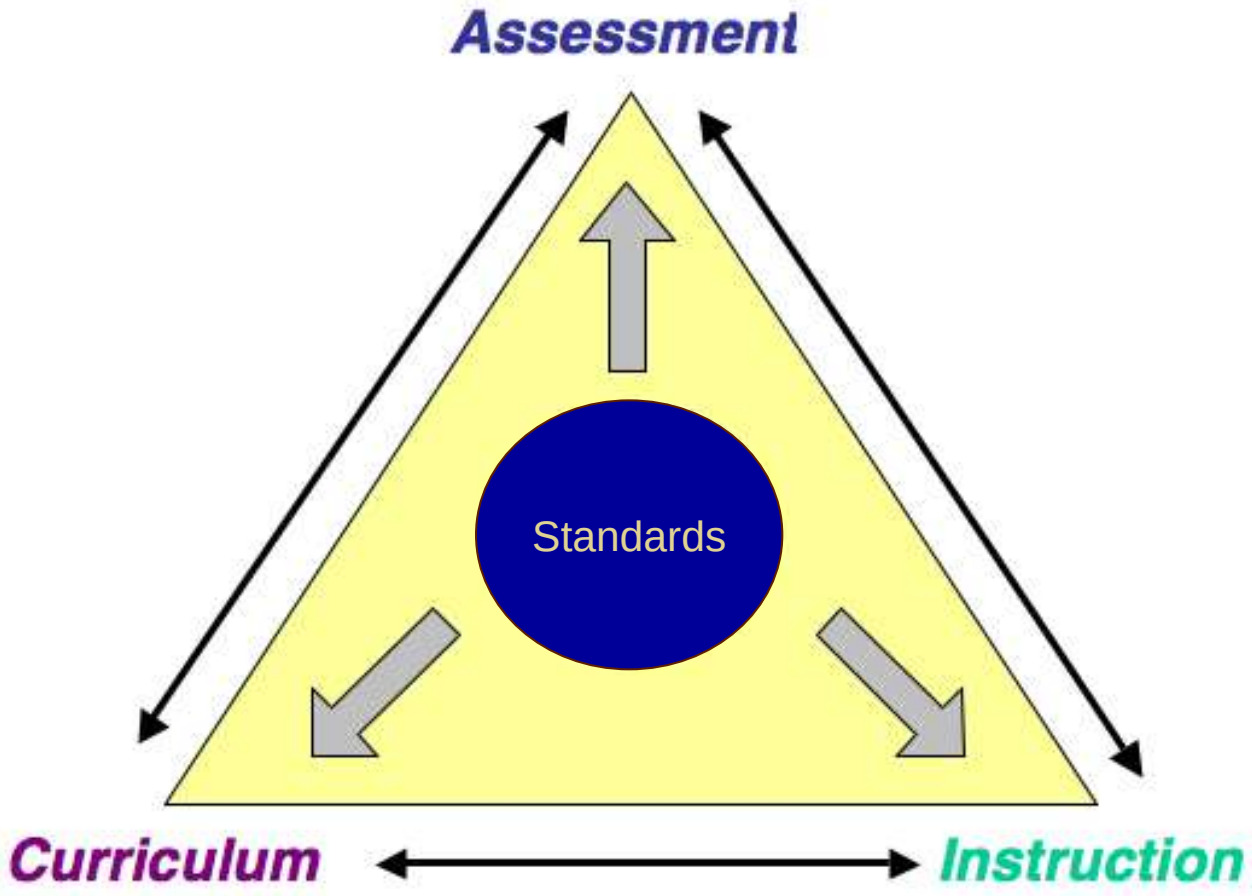
- The NRC Science Framework has proposed descriptions of student competence as being the intersection of knowledge involving:
  - **important disciplinary practices**
  - **core disciplinary ideas,**
  - **and crosscutting concepts** with
  - **performance expectations** representing the intersection of the three.
- It views competence as something that develops over time & increases in sophistication and power as the product of coherent curriculum & instruction

Goals for Teaching & Learning

- Coherent investigations of core ideas across multiple years of schooling
- More seamless blending of practices with core ideas
- Performance expectations that require reasoning with core disciplinary ideas
  - explain, justify, predict, model, describe, prove, solve, illustrate, argue, etc.



Aligning Curriculum, Instruction & Assessment



7 Things for Us to Think About
(+ or – 2)

1. assessment contexts, purposes and uses,
2. the nature of assessment and the importance of research on learning,
3. assessment design processes,
4. validity arguments,
5. measurement and statistical inference,
6. affordances of technology, and
7. systems of assessment

Contexts and Purposes

- Educational assessment typically occurs in multiple contexts:
  - Small scale: individual classrooms
  - Intermediate-scale: districts
  - Large-scale: states, nations
- Within and across contexts it can be used to accomplish differing purposes:
  - Assist learning (formative)
  - Measure individual achievement (summative)
  - Evaluate programs (accountability)

A Multiplicity of Actors & Needs

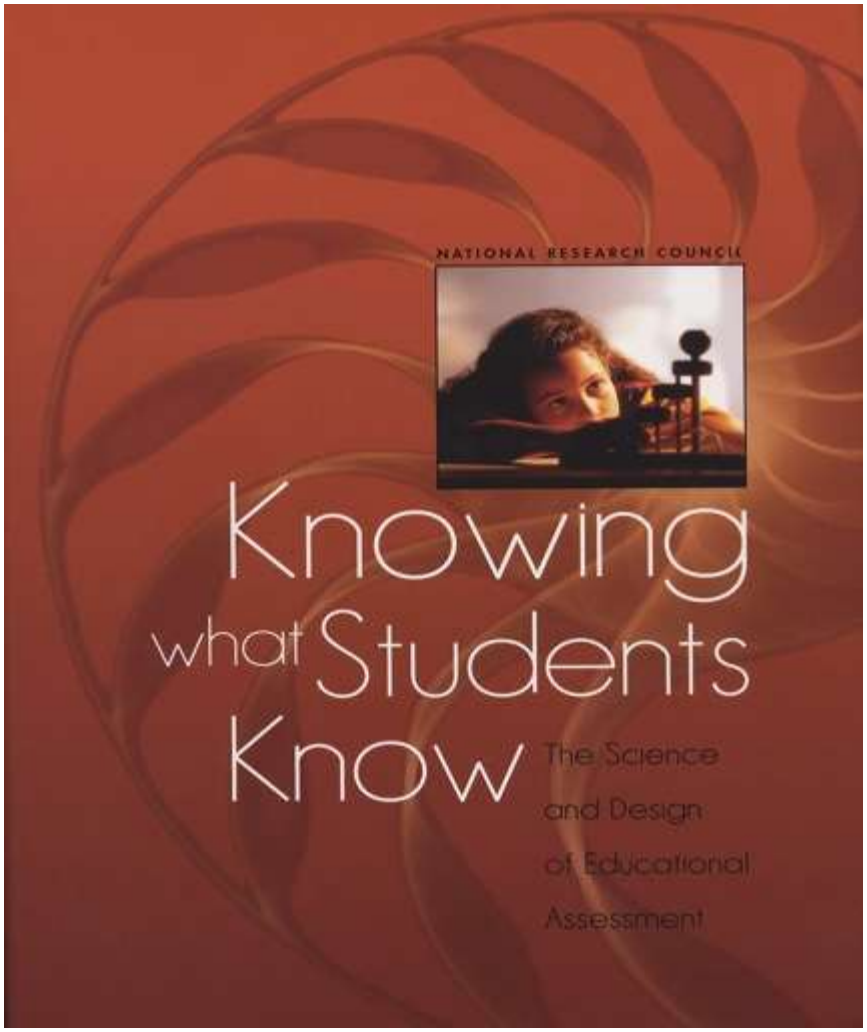
The reason we have so many different forms and types of assessment is that "***One size does not fit all***"

- Educators at different levels of the system need different information at different time scales
- They have differing priorities, they operate under different constraints, & there are tradeoffs in terms of time, money, and type of information needed

7 Things for Us to Think About
(+ or – 2)

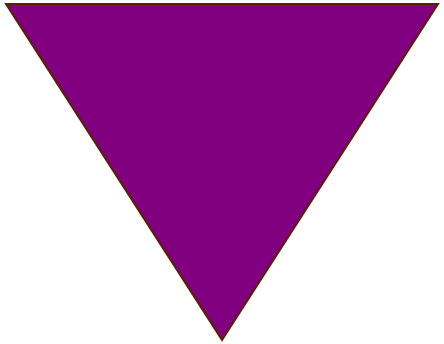
1. assessment contexts, purposes and uses,
2. the nature of assessment and the importance of research on learning,

Assessment as a Process of Reasoning from Evidence



observation

interpretation



cognition

Must be coordinated!

Why Models of Development of Domain Knowledge are Critical

- Tell us what are the important aspects of knowledge that we should be assessing.
- Give us strong clues as to how such knowledge can be assessed
- Can lead to assessments that yield more instructionally useful information
  - diagnostic & prescriptive
- Can guide the development of systems of assessments intended to cohere
  - across grades & contexts of use

Learning Progressions in Science

An Evidence-based Approach to Reform

Prepared by
Tom Corcoran
Frederic A. Mosher
Aaron Rogat

Center on Continuous Instructional Improvement
Teachers College–Columbia University

Learning Progressions in Science

Current Challenges and Future Directions

Alicia C. Alonzo and
Amelia Wenk Gotwals (Eds.)

LeaPS Learning Progressions in Science

SensePublishers

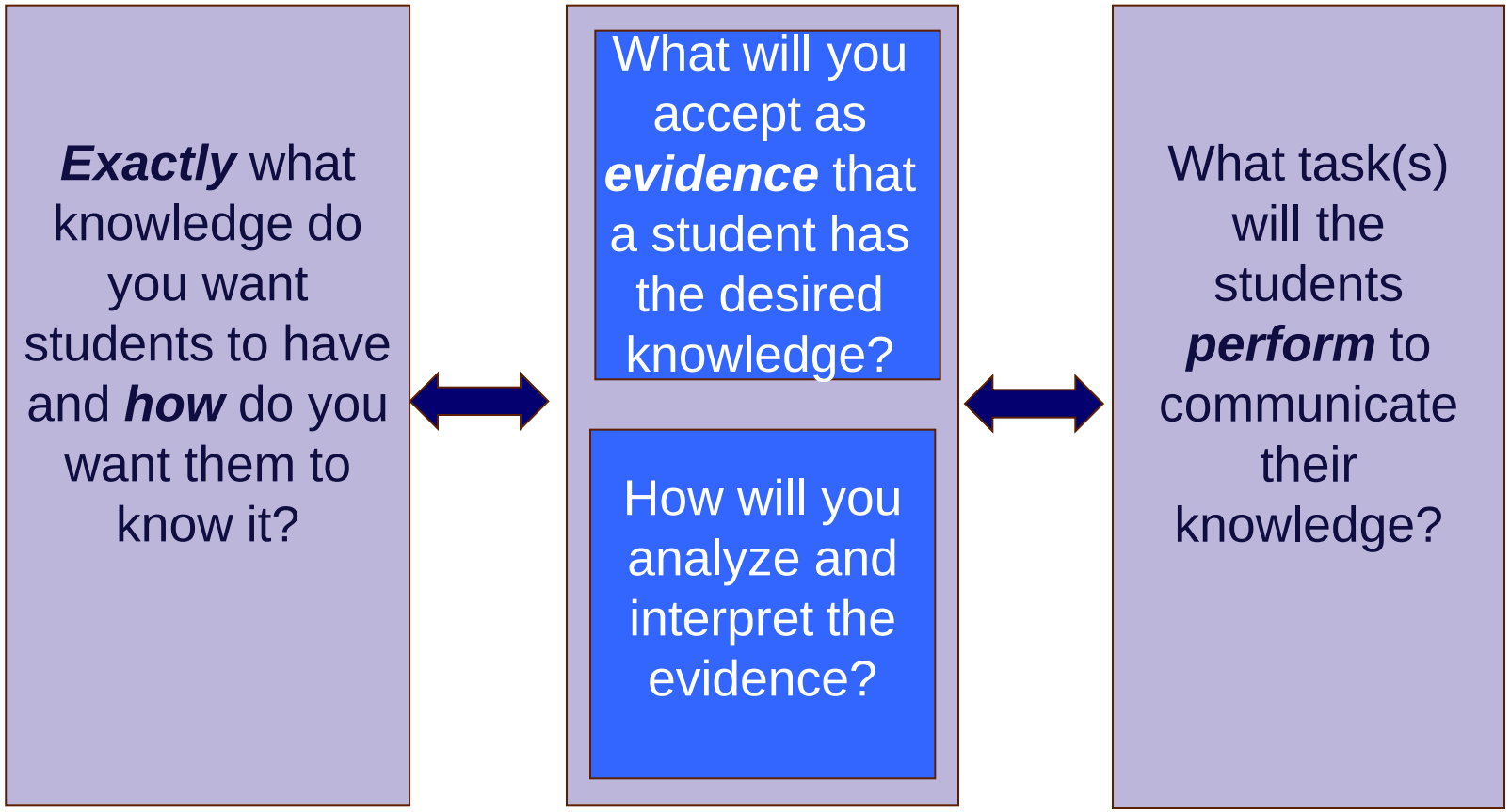
7 Things for Us to Think About (+ or – 2)

1. assessment contexts, purposes and uses,
2. the nature of assessment and the importance of research on learning,
3. assessment design processes,

Issues of Assessment Design & Development

- Assessment design spaces vary tremendously & involve multiple dimensions
  - Type of knowledge and skill and levels of sophistication
  - Time period over which knowledge is acquired
  - Intended use and users of the information
  - Availability of detailed theories & data in the domain
  - Distance from instruction and assessment purpose
- Need a principled process that can help structure going from theory, data and/or speculation to an operational assessment
  - ***Evidence Centered Design***

Evidence-Centered Design



Need to consider what this might mean when it comes to performance expectations integrating Disciplinary Core Ideas & Practices

7 Things for Us to Think About
(+ or – 2)

1. assessment contexts, purposes and uses,
2. the nature of assessment and the importance of research on learning,
3. assessment design processes,
4. validity arguments,

Messick's Perspective on Validity

- Validity is an integrated evaluative judgment of the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of inferences and actions based on test scores or other modes of assessment. (Messick, 1989)

Kane's Perspective on Validity

- Mike Kane's 2006 chapter on Validity distinguishes the ***interpretive argument*** versus the ***validity argument***.
- The ***interpretive argument*** is a linked set of propositions that forms the rationale or basis for the proposed test use/interpretation
- The ***validity argument*** consists of evaluative studies in support of the propositions in the interpretive argument.

Three Aspects of Validity

- **Cognitive:** The extent to which an assessment taps important forms of domain and disciplinary knowledge and skill and does so in ways that are not confounded with other aspects of cognition such as language or memory load.
- **Instructional:** The extent to which it supports teaching practice and provides valuable and timely instructional information.
- **Inferential:** The extent to which it reliably yields model-based information about student knowledge and skills for diagnostic or other purposes.

7 Things for Us to Think About
(+ or – 2)

1. assessment contexts, purposes and uses,
2. the nature of assessment and the importance of research on learning,
3. assessment design processes,
4. validity arguments,
5. measurement and statistical inference,

Advances in Measurement: Beyond Models of General Proficiency

- Three general sets of measurement issues that can be accommodated by various models
  - continua vs classes
  - single vs multiple attributes
  - status vs change
- There exists a progression of models and methods of varying complexity and contexts of application
  - Types of understanding instead of rankings
  - Multiple aspects of proficiency rather than single score
  - Change and growth over time
  - Diagnostic indices
  - Families of models adaptable to broad range of uses
    - e.g., Bayes Nets which are basis for intelligent tutoring systems

Connecting Learning Theory & Measurement Theory

- At a conceptual level we need to explore the fit between particular statistical models and methods and varying descriptions of competence and learning
- Models we need will likely vary with the timescale of learning, the "grain size" of analysis, and with the intended purpose and use of the inferences we wish to make
- Dialogue and Collaboration is needed among domain experts, psychometricians, and learning scientists

7 Things for Us to Think About
(+ or – 2)

1. assessment contexts, purposes and uses,
2. the nature of assessment and the importance of research on learning,
3. assessment design processes,
4. validity arguments,
5. measurement and statistical inference,
6. affordances of technology

A Claim We May Need to Embrace

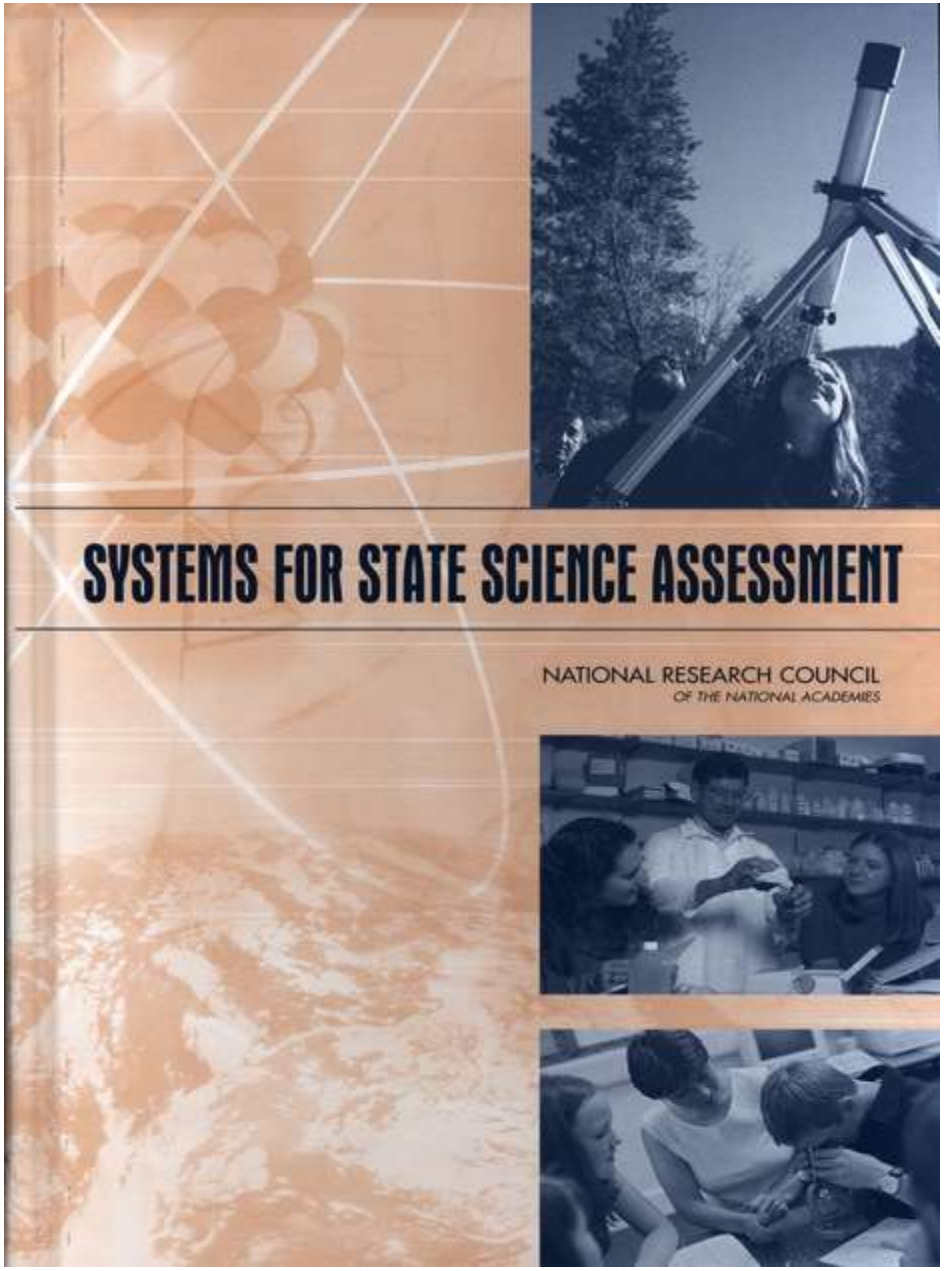
- Much of what is ***new, different, and important*** in the NRC Science Framework and NGSS cannot be adequately assessed by conventional methods, items, and measurement models
  - The capacity to engage in the practices as connected to important domain-specific ideas and understandings
  - We are interested in the processes of thinking as well as the products of those processes

Advantages of Technology for Science Assessment

- Present authentic, rich, dynamic environments
- Present phenomena difficult or impossible to observe and manipulate in classrooms
- Represent temporal, causal, dynamic relationships "in action"
- Allow multiple representations of stimuli and their simultaneous interactions (e.g., data generated during a process)
- Allow overlays of representations, symbols
- Allow student manipulations/investigations, multiple trials
- Allow student control of pacing, replay, reiterate
- Capture student responses during research, design, problem solving

7 Things for Us to Think About
(+ or – 2)

1. assessment contexts, purposes and uses,
2. the nature of assessment and the importance of research on learning,
3. assessment design processes,
4. validity arguments,
5. measurement and statistical inference,
6. affordances of technology, and
7. systems of assessment



- **Desired end product is a multilevel system**
  - Each level fullfills a clear set of functions and has a clear set of intended users of the assessment information
  - The assessment tools are designed to serve the intended purpose
    - Formative, summative or accountability
    - Design is optimized for function served
- **The levels are articulated and conceptually coherent**
  - They share the same underlying concept of what the targets of learning are at a given grade level and what the evidence of attainment should be.
  - They provide information at a "grain size" and on the "time scale" appropriate for translation into action.

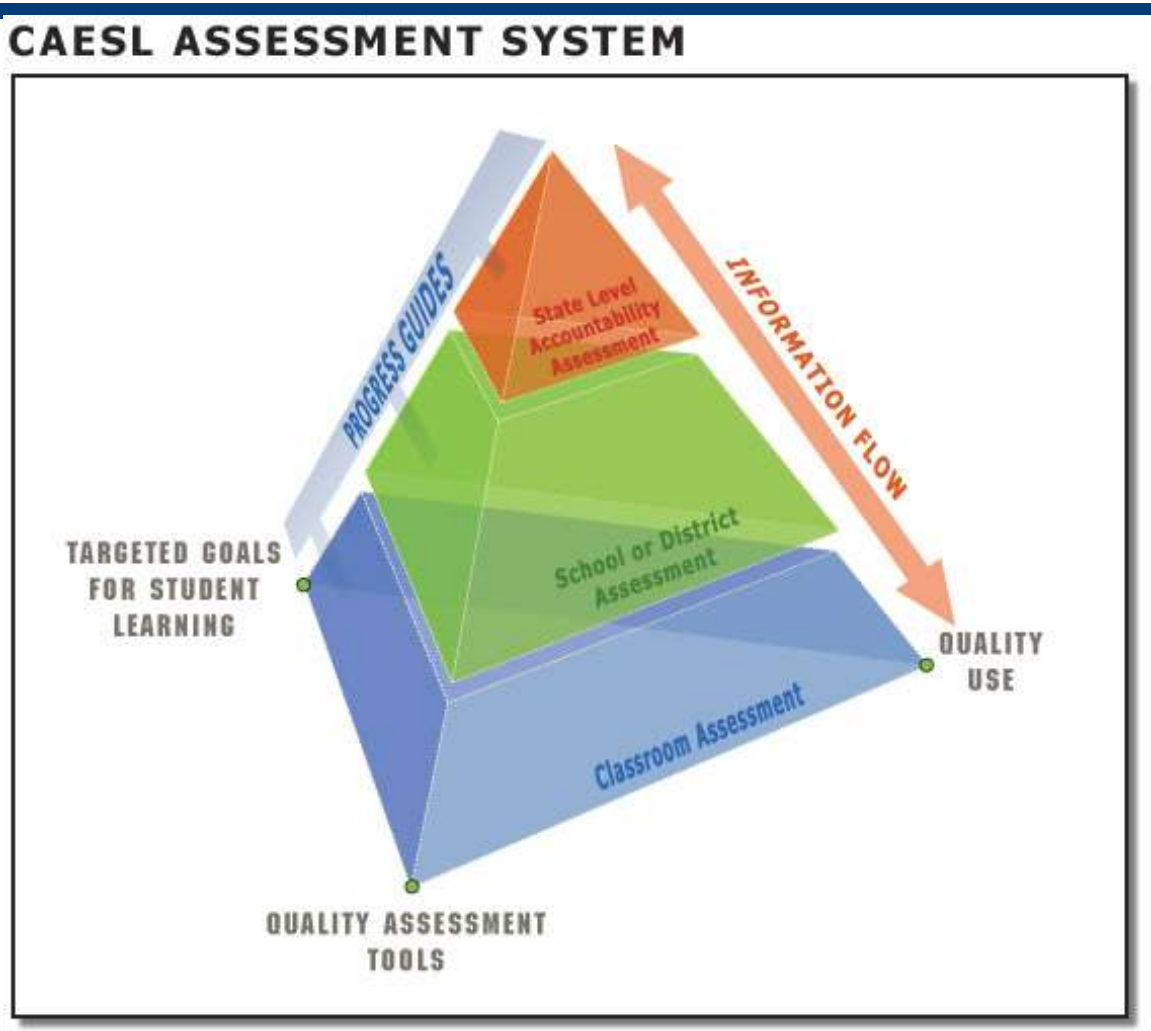
What Such a System Might Look Like

An Integrated System

Coordinated across levels

Unified by common learning goals

Synchronized by unifying progress variables



Multilevel Assessment System

The Key Design Elements of Such a Comprehensive System

- The system is designed to track progress over time
  - At the individual student level
  - At the aggregate group level

- The system uses tasks, tools, and technologies appropriate to the desired inferences about student achievement
  - Doesn't force everything into a fixed testing/task model
  - Uses a range of tasks: performances, portfolios, projects, fixed- and open-response tasks as needed

Assessment Should not be the "*Tail Wagging the STEM Education Dog*"