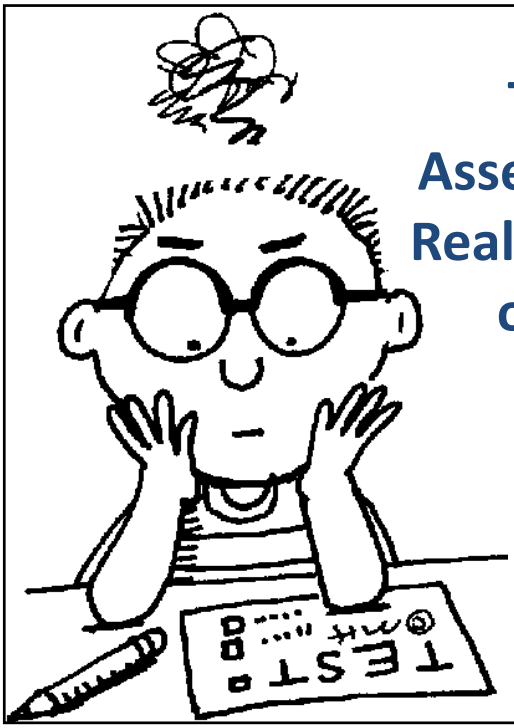


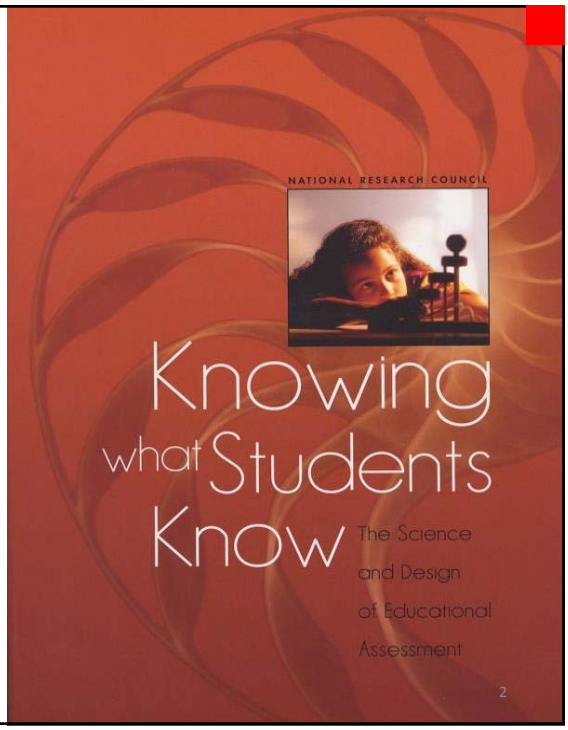
The Future of Assessment in Texas: Realizing the Promise of Educational Assessment

Jim Pellegrino
Learning Sciences
Research Institute
University of Illinois
at Chicago

Background

Based on ideas drawn from the National Research Council report:

Knowing What Students Know: The Science and Design of Educational Assessment.



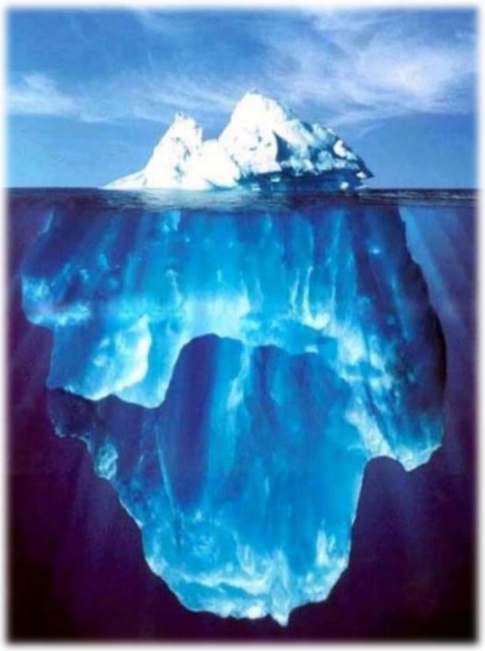
Discussion Topics

- The Nature of Educational Assessment
- Federal Law, College Readiness Standards, and High Quality Assessments
- Considering a Balanced/Comprehensive Assessment System as Texas and Its Districts Move Forward

3

- Defining formative, interim, and summative assessment
- Characteristics, uses, and examples of formative, interim, and summative assessment

THE NATURE OF EDUCATIONAL ASSESSMENT

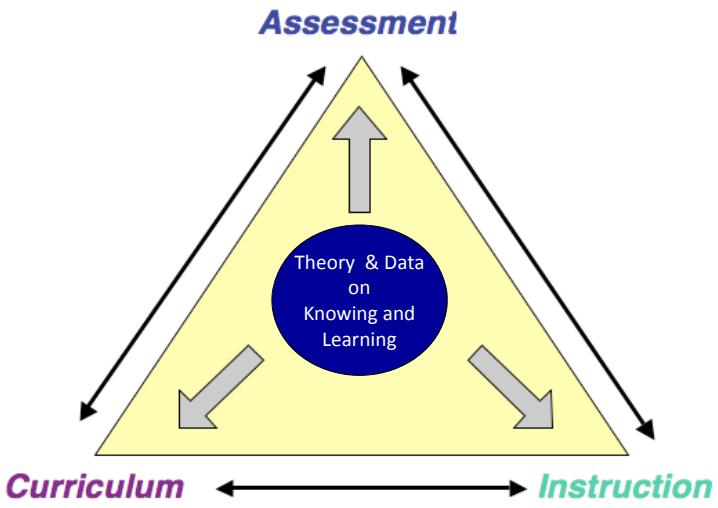


4

What is educational assessment? What is its primary purpose?

- Assessment is a process of gathering information for the purpose of making judgments about a current state of affairs.
- In educational assessment, the information collected is *designed* to help teachers, administrators, policy makers, and the public infer what students know and how well they know it, presumably for the purpose of enhancing future outcomes.
- Some of these outcomes are more immediate such as the use of assessment in the classroom to improve student learning and others are more delayed such as the use of assessment for program evaluation.

Where Does Assessment Fit in the Educational System?



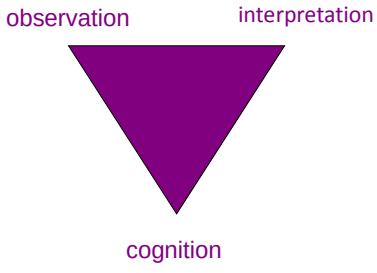
What functions and purposes is assessment supposed to serve?

- Educational assessment typically occurs in multiple contexts:
  - Small scale: individual classrooms
  - Intermediate-scale: districts
  - Large-scale: states, nations, international
- Within and across contexts it can be used by different stakeholders to accomplish differing purposes:
  - Assist learning (formative)
  - Measure individual (or group) achievement (interim/summative)
  - Evaluate programs (Interim/summative)
- Both the purpose of assessment and the context in which it occurs influence the design.

Why does assessment of student learning seem to be such a major challenge?

You Can Never Really Know What a Student Knows:
Assessment is a Process of Reasoning from Evidence

- **cognition**
  - Theories, models & data about how students represent knowledge & develop competence in the domain
- **observations**
  - tasks or situations that allow one to observe students' performance
- **interpretation**
  - method for making sense of the data



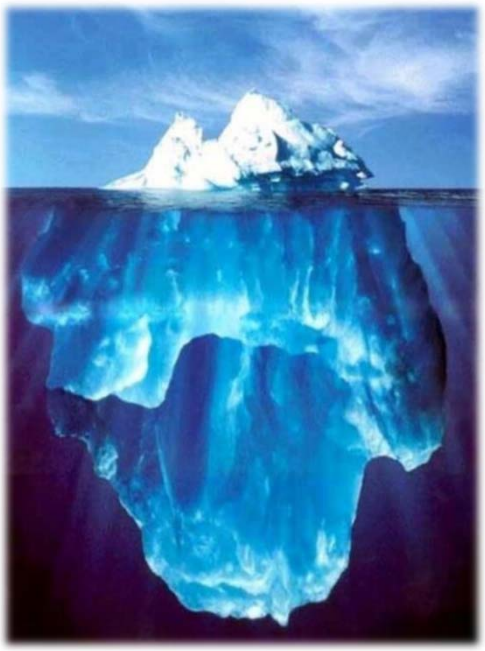
Must be coordinated!

Why do we seem to assess so much? Can we get the job done with just one test?

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

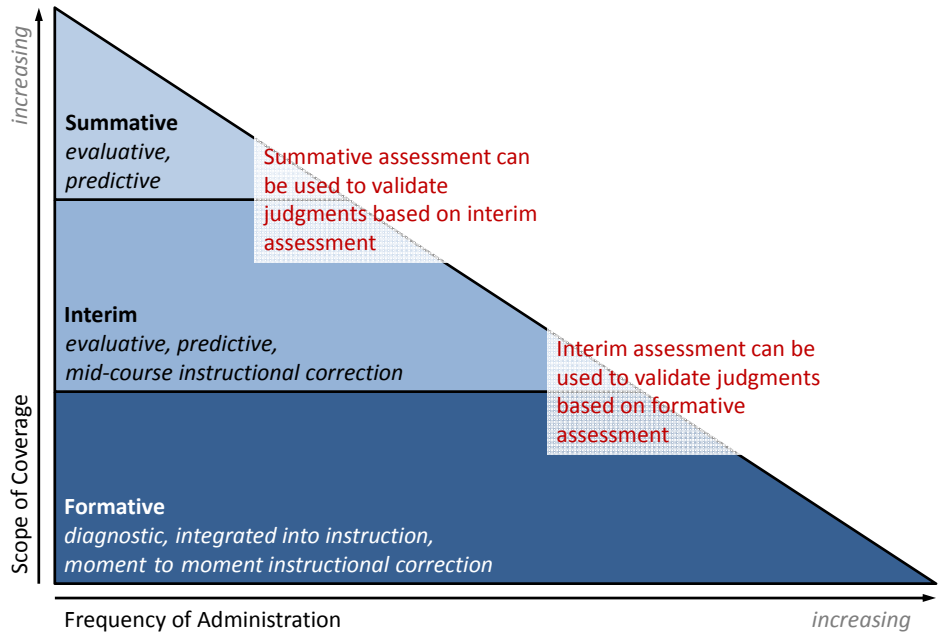
- Educators at different levels of the system need different information at different times and in different forms
- They have differing priorities, they operate under different constraints, & there are tradeoffs in terms of time, money, and type of information needed.
- Assessments must be designed, developed and reported with the intended user and use in mind.

DEFINING FORMATIVE, INTERIM, AND SUMMATIVE ASSESSMENT



10

Three "Tiers of Assessment"



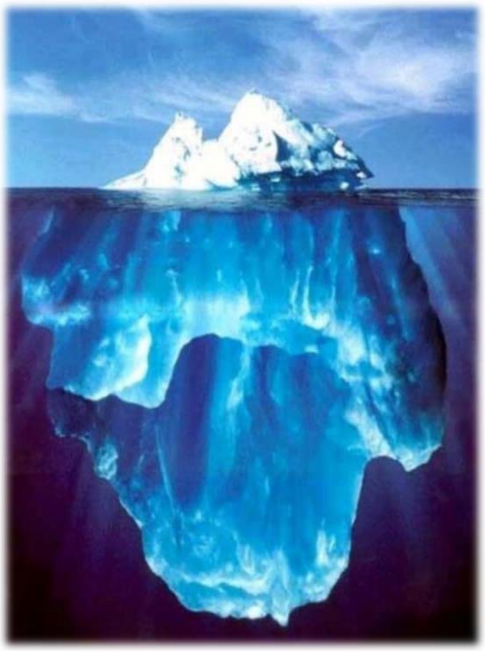
Defining an Assessment System

"A collection of assessments does not entail a system any more than a pile of bricks constitutes a house" (Coladarci, 2002).

The system must be composed of elements that cohere and work together in terms of the intended functions and interpretive uses.

Owner	Tier/Type of Assessment: Formative	Interim	Summative
Teacher	• Strategically planned mid-period check-ins • Strategically planned end of period check-ins • Homework that will be used to provide at least one round of feedback and revision before grading	• Graded quizzes and homework • Unit projects, papers, and exams • Mid-term exams • Marking period exams	• Final exams, projects, and papers
District	• Not applicable	• Common unit exams, mid-terms, and marking period exams • Common quarterly assessments • District placement tests	• Common final exams, projects, and papers • Common assessments for testing out of a course/credit • Common graduation assessments
State	• Not applicable	• State-provided within-year common assessments	• Annual state tests • End of course state tests

CHARACTERISTICS AND USES OF FORMATIVE, INTERIM, AND SUMMATIVE ASSESSMENT



14

Summative Assessment

Characteristics

- Pauses instruction for evaluation
- Controlled by one or more teachers, schools, districts, or states
- Covers a macro unit of instruction (e.g., a semester, course, credit, grade)
- Infrequent (e.g., yearly, finals week)
- Administered after completing a macro unit
- Based on who controls assessment, results may be comparable across students, classrooms, districts, and/or states
- A product

Uses

- Evaluate achievement after a macro unit
- Monitor progress across multiple macro-units
- Corroborate interim assessment
- Evaluate readiness for the next macro unit
- After-the-fact evaluation/ adjustment of broad instructional practices by individual teachers
- After-the-fact evaluation/adjustment of curriculum/programming policies by administrators
- Predict later student outcomes
- Grading and accountability

15

Interim Assessment

Characteristics

- Pauses instruction for evaluation
- Controlled by one or more teachers, schools, districts, or states
- Covers a mid-sized unit of instruction (e.g., a semester, course, credit, grade)
- Somewhat frequent (e.g., weekly to quarterly)
- Administered before and/or after a mid-sized unit
- Based on who controls assessment, results may be comparable across students, classrooms, districts, and/or states
- A product

Uses

- Evaluate achievement after a mid-sized unit
- Monitor progress within a macro-unit (e.g., semester, course, credit, grade)
- Corroborate formative assessment
- Pre-test to tailor unit instructional plans for the group and individual students
- Identify post-unit remedial needs
- Mid-course self-evaluation and adjustment of teacher classroom practices
- Mid-course evaluation and adjustment of school and district policies and programs
- Predict performance on summative assessment
- Grading (and possibly accountability)

16

Formative Assessment

Characteristics

- Facilitate effective instruction (does not pause instruction)
- Learning goals and criteria are clear to students
- Students self-/peer-monitor progress toward learning goals
- Students and teachers receive frequent feedback
- Jointly controlled by each teacher and students
- Covers a micro unit of instruction on a frequent basis (e.g., at least once per class period)
- Tailored to a set of students and an instructional plan
- Might be comparable for a classroom, but not beyond
- Not a product (e.g., quiz, test, bank of questions/tests), a process

Uses

- Engage students in learning/metacognition through frequent feedback and self-/peer-evaluation
- Monitor moment-to-moment student learning
- Diagnose immediate individual and group instructional needs
- Adjust/differentiate instruction in the moment
- Self-evaluate micro-unit instructional effectiveness
- Student results from formative assessment are not appropriate for use in grading or accountability; however, ratings of the quality of formative assessment practice may be appropriate for use in accountability

17

Locus of Effects of Information Derived from Each Tier

- Tier 1: Long-cycle (State or District tests; Summative)
  - Student monitoring
  - Curriculum alignment
- Tier 2: Medium-cycle (Interim; Benchmark)
  - Improved student monitoring of the state of their learning and connections among content
  - Improved teacher cognition about learning
- Tier 3: Short-cycle (Classroom; Formative)
  - Improved classroom practice
  - Improved student engagement
  - Student metacognitive monitoring of the state of their knowledge

What are some key "take away" points?

- Assessment is not a simple matter and not just one thing -- it takes multiple forms for multiple purposes
- Designing good assessment is very challenging -- need solid conceptual foundation about what students should know and how they should know it.
- Assessment needs to be part of an integrated system of curriculum, instruction & assessment
- A major challenge is helping teachers use assessment productively in their ongoing practice -- especially formative assessment

Discussion Topics

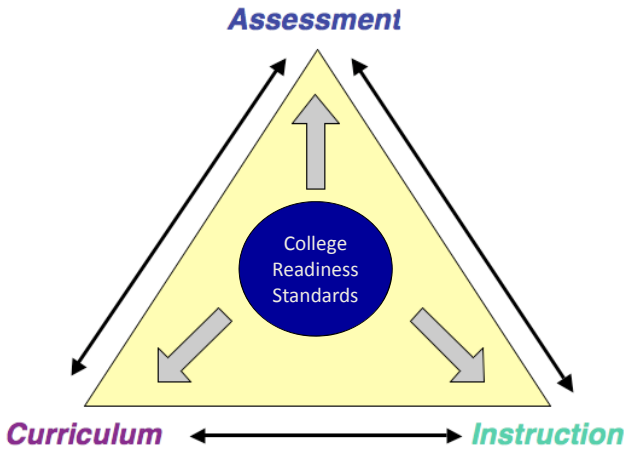
- The Nature of Educational Assessment
- Federal Law, College Readiness Standards, and High Quality Assessments
- Considering a Balanced/Comprehensive Assessment System as Texas and Its Districts Move Forward

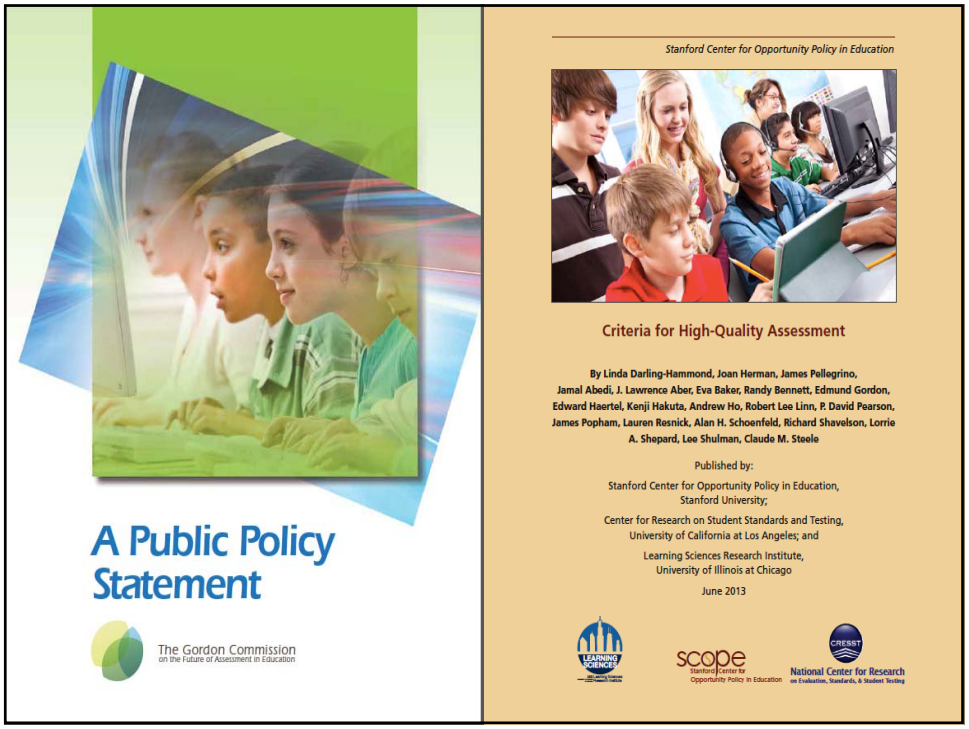
20

Federal Law: NCLB & ESSA Key Requirements

- Annual assessments of all students in Math and Reading for Grades 3-8, and once in grades 9-12
  - Math and Reading annual assessments must be aligned with state academic content and achievement standards
- Annual assessment of students in science no less than once in each of grades 3-5, 6-9 and 10-12
  - Science assessments must be aligned with state academic content and achievement standards
- Reporting in multiple categories for multiple demographic groups
- With ESSA there is more state autonomy than under NCLB, including options for accountability

Using Standards to Align Curriculum, Instruction & Assessment





"To be helpful in achieving the learning goals……, assessments must fully represent the competencies that the increasingly complex and changing world demands…. To do so, the tasks and activities in the assessments must be models worthy of the attention and energy of teachers and students."

-- The Gordon Commission

What does it mean to be "High Quality"?

Criteria for High-Quality Assessment

25

Five Criteria for High-Quality Assessment

1. Assessment of Higher-Order Cognitive Skills
2. High-Fidelity Assessment of Critical Abilities
3. Standards that Are Internationally Benchmarked
4. Items that Are Instructionally Sensitive and Educationally Valuable
5. Evidence of Validity, Reliability, and Fairness

Assessment of Higher Order Cognitive Skills

- A large majority of items and tasks (at least two-thirds) evaluate the conceptual knowledge and applied abilities that support transfer
- At least one-third of the assessment content in mathematics, and at least one-half in English language arts, should evaluate higher-order skills that allow students to become independent thinkers and learners

27

High Fidelity Assessment of Critical Skills

- High fidelity assessment needs to include
  - Research, including synthesis and analysis of information
  - Experimentation and evaluation
  - Oral and written communications
  - Use of technology to access, analyze, and communicate information
  - Collaboration
  - Modeling, design, and problem solving using quantitative skills

28

International Benchmarking

- Calibrated to international assessments such as PISA, International Baccalaureate

29

Instructionally Sensitive and Educationally Valuable

- Research confirms instructional sensitivity
- Rich feedback on learning and performance
- Tasks that reflect and can guide valuable instructional activities

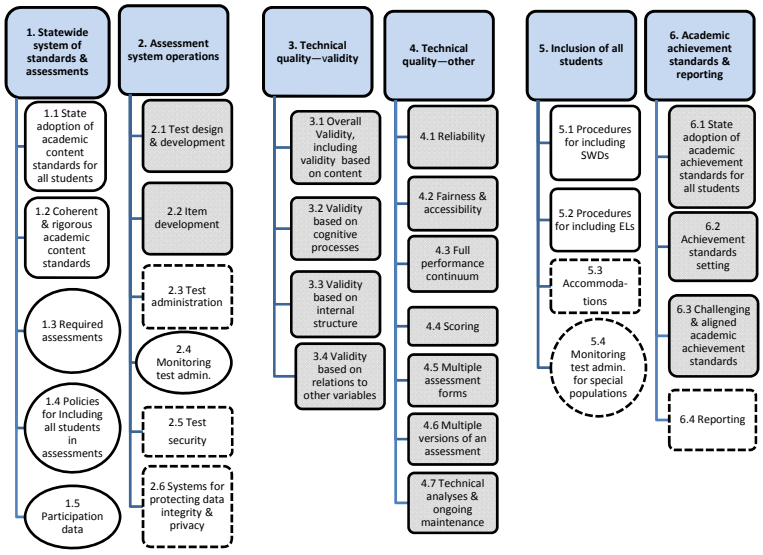
30

Valid, Reliable, and Fair Assessments

- Knowledge and skills are well measured
- Scores are related to abilities they are meant to predict
- Evidence that scores are valid for intended uses
- Evidence that scores are unbiased regarding demographic background and disabilities
- Evidence that scores measure learning accurately along a broad continuum of achievement.
- Evidence that items/tasks tap intended cognitive processes

31

Peer Review: Critical Elements



32

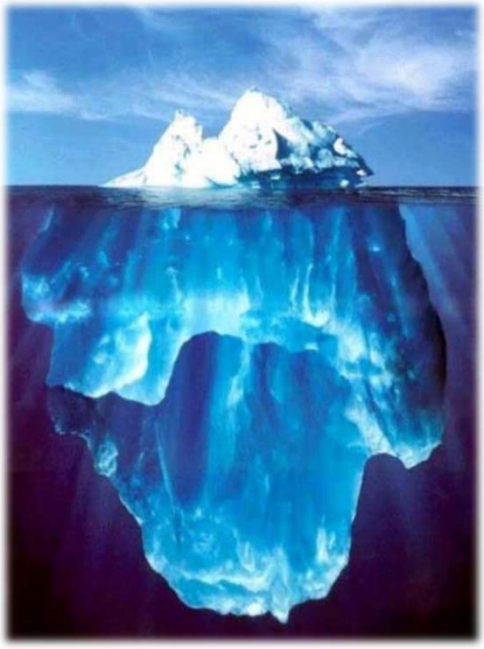
CCSSO Criteria for High Quality Assessments

- Support states as they "develop procurements and evaluate options for high-quality state summative assessments aligned to their college- and career readiness standards."
- Grouped into five broad categories:
  - A. Meet Overall Assessment Goals and Ensure Technical Quality
  - B. Align to Standards – English Language Arts/Literacy
  - C. Align to Standards – Mathematics
  - D. Yield Valuable Reports on Student Progress and Performance
  - E. Adhere to Best Practices in Test Administration

A. Meet Overall Assessment goals and Ensure Technical Quality

- Indicating progress toward college and career readiness
- Ensuring that assessments are valid and required for intended purposes
- Ensuring that assessments are reliable
- Ensuring that assessments are designed and implemented to yield valid and consistent test score interpretations within and across years
- Providing accessibility to all students, including English learners and students with disabilities
- Ensuring transparency of test design and expectations
- Meeting all requirements for data privacy and ownership

QUESTIONS RELATED TO THE TRANSITION TO ASSESSMENT ALIGNED TO COLLEGE AND CAREER READY STANDARDS



35

Why did proficiency rates drop this year in many states?

- Given differences between the most recent tests used by many states and their previous state tests, this should actually be expected.
- The visible "drop" in proficiency is not actually a drop.
- What we see arises from increasing expectations for student achievement and relatively little change in student performance, proficiency, or school effectiveness.
  - States have adopted more challenging academic standards and raised expectations for what students should know and be able to do when they graduate from high school.
  - If states had maintained their former achievement expectations, students would have performed at least as well as students in previous years.
- The new standards and expectations for student achievement better reflect the demands of college and careers.

36

What do proficiency rates tell us? What do the new scores mean?

- Proficiency rates provide insight into the rigor of the test, student performance, and the status of implementation of new content standards.
  - Assessment results provide information about how students perform on the new content standards and expectations for achievement
- The new scores indicate whether and to what extent, students are on track to be successful in college and careers

37

Can results from a test aligned to new standards be compared to results from previous years?

- It is not possible to make a direct or simple comparison between state results on a new assessment and results on your past assessment.
  - The change in assessments, scales, and achievement standards represents a clean break from the past assessment
- Even when statistical linking occurs, interpreting student performance on the new test in terms of old achievement levels and scales is not appropriate because the assessment aligns to new expectations.

w

Discussion Topics

- The Nature of Educational Assessment
- Federal Law, College Readiness Standards, and High Quality Assessments
- Considering a Balanced/Comprehensive Assessment System as Texas and Its Districts Move Forward

39

Need for a "Theory of Action"

A common problem at state and/or district levels is that the assessment components are not conceptually coherent.

They often conflict and as a consequence their use doesn't lead to the desired outcomes of educational improvement.

It is essential to make **EXPLICIT** one's assumptions and "theory of action" for the system of assessments.

A Theory of Action

What is it?

- An empirically and logically stated argument
- A set of underlying assumptions
- A testable hypothesis

…that outlines how and why a given assessment, system, or program, as designed, will support the achievement of specified goals.

- It requires specification of each component of the assessment/evaluation system, the connection between components, and the manner in which they jointly fulfill the requirements of the system.

Pieces of the TOA Puzzle for a CAS

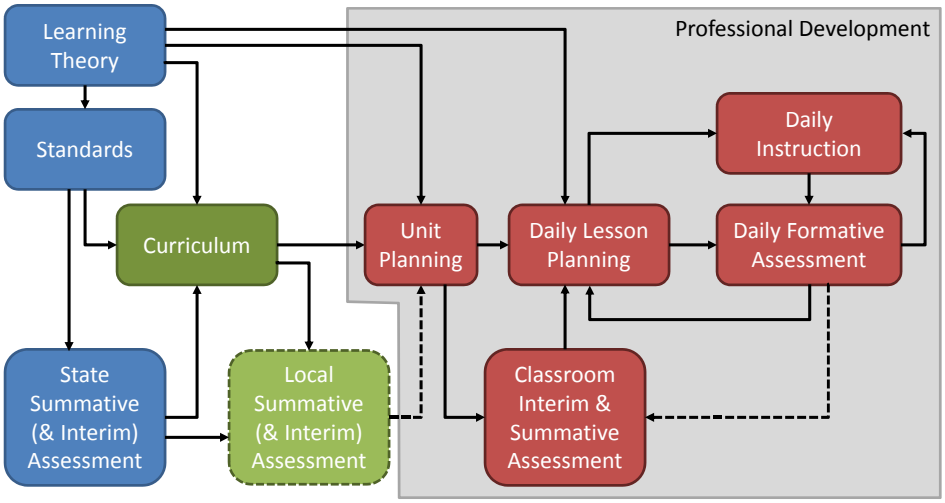
- Purpose
- Theory of Learning
- Priortized Goals of the System
- Intended Uses of Results

→ Must be well articulated prior to assessment system design.

- Overarching Theory as to manner in which the assessment system will bring about desired change (Key Design Principles)
- Design of the system and it's component parts
  - Assessments, Tasks
  - Alignment of each component to goals/intended uses/Key Design Principles
- Mechanism by which component are intended to provide for specified goals.
- Expected relationship among components
- Inferences/assumptions underlying the system working as intended.

→ Articulated as part of assessment system design.

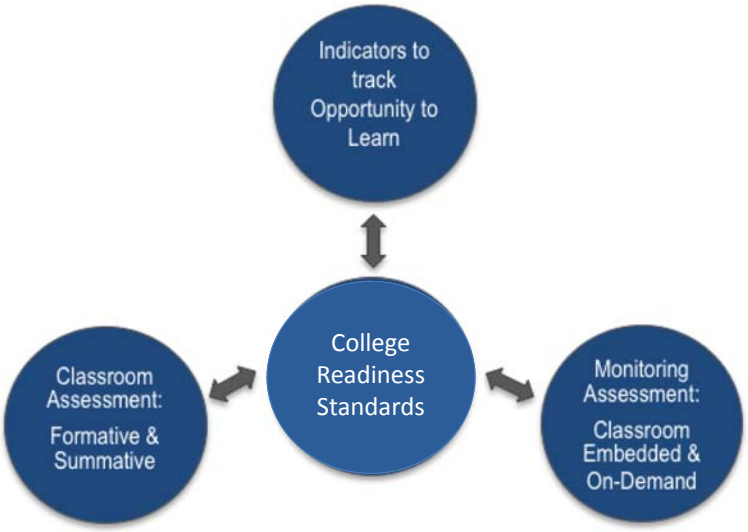
Relating Teaching, Learning, and Assessment: A "Simple" Theory of Action for State & District Officials



Need a Coherent System of Assessments – 3 Major Components

- A system of assessments should include classroom assessment, monitoring (large-scale) assessments, and indicators of opportunity to learn.
  - **Classroom assessment should be an integral part of instruction and should reinforce the type of learning envisioned in standards.**
  - **Monitoring (large-scale) assessments will need to include an on-demand component and a component based in the classroom (classroom-embedded) in order to fully cover the breadth and depth of the standards.**
  - **Indicators of opportunity to learn should document that students have the opportunity to learn in the way called for in standards and that schools have appropriate resources.**

Assessment System Components



1st Major Challenge in Design of the Monitoring Component: Intended uses of the Information

The Complex Space of Monitoring Functions

TABLE 5-1 Questions Answered by Monitoring Assessments

	Levels of the Education System			
Types of inferences	Individual Students	Schools or District	Policy Monitoring	Program Evaluation
Criterion-referenced	Have individual students demonstrated adequate performance in science?	Have schools demonstrated adequate performance in science this year?	How many students in state X have demonstrated proficiency in science?	Has program X increased the proportion of students who are proficient?
Longitudinal and comparative across time	Have individual students demonstrated growth across years in science?	Has the mean performance for the district grown across years? How does this year's performance compare to last year's?	How does this year's performance compare to last year's?	Have students in program X increased in proficiency across several years?
Comparative across groups	How does this student compare to others in the school/state?	How does school/ district X compare to school/district Y?	How many students in different states have demonstrated proficiency in science?	Is program X more effective in certain subgroups?

2nd Major Challenge in Design of the Monitoring Component: Possible Sources of Evidence

State Assessments for Monitoring

Combine two types of external assessment strategies, in conjunction with OTL indicators:

On-Demand Assessments

- Developed by the state
- Administered at a time mandated by the state

Classroom-Embedded Assessments

- Developed by the state or district,
- Administered at a time determined by the district/school that fits the instructional sequence in the classroom

Possible Options for the On-Demand Assessment Components

- Mixed item formats, including extended constructed response
  - Such as AP exams
- Mixed item formats with performance tasks
  - might involve both group and independent activities
  - might involve some hands-on tasks
- Use matrix sampling, depending on the intended use and the need to report scores for individuals versus for groups.

Possible Options for the Classroom-Embedded Assessment Components

- **Replacement units** (curriculum materials + assessments) developed outside of the classroom (by state or district)
- **Item banks of tasks**, developed outside of the classroom
- **Portfolio collections of work samples**, with tasks specified by state or district

How might states and districts organize the different assessments that they seem to need?

- Desired end product is a multilevel system
  - Each level fulfills 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, interim, summative
    - 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 are 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

What else is needed for change to occur?

Much of the change in the productive use of assessment requires training in the use of new tools and systems

- A substantial professional development effort is needed across levels of the system
  - Teachers, principals, and district leadership
- Processes for the effective collection and use and interpretation of assessment information need to be implemented
  - Focus of many assessment literacy efforts
- New technologies and data systems may need to be created and accommodated in the system's business practices

One Vision for how integration of assessment and instruction happens in districts and schools

District Level

- District has a vision for high quality teaching and learning.
- High quality tasks are embedded into the K-12 curriculum.
- Standards Based Grading is aligned.
- District supports ongoing professional learning for staff.

School Level

- Common collaboration for grade level teams is in place.
- Administrators support this work.

Classroom Level

- Students engage in ongoing problem solving and challenging assessments.
- Teachers engage in formative assessment processes.

What else is needed to make assessment useful in promoting student achievement?

Assessment Should not be the
"Tail that Wags the Educational Dog"