

Next Generation Assessment Systems: Toward Coherence and Utility

Joan L. Herman

*National Academies Workshop II:
Best Practices for State Assessment Systems*

Washington, DC
April 6-7, 2010

National Center for Research
on Evaluation, Standards, & Student Testing



Overview

- Propitious time for assessment
- Nature of coherent systems
- Fundamental system criteria
- Implications

Propitious Time



- RTT: \$4.3 billion enacted as part of ARRA
- Common Core Learning Standards
- RTT assessment grants
- Belief in power of [formative] assessment

Needed: *Coherent Assessment Systems*

Coherent

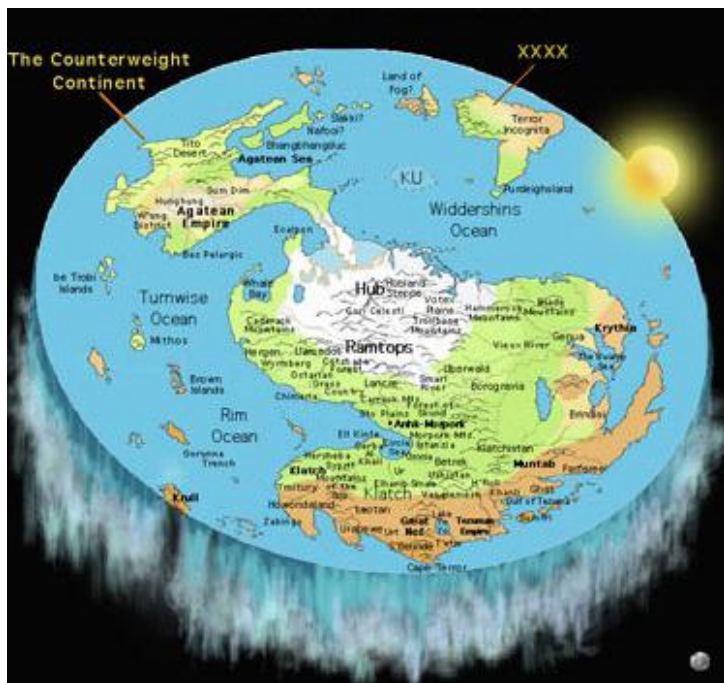
System of assessments

Implicit: System validity
for intended purposes

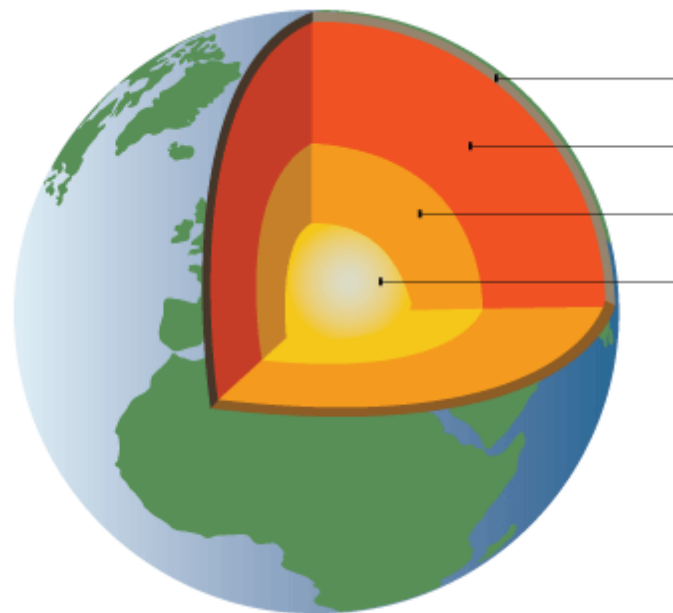
Why a *System* of Assessments?

- Multiple measures to adequately represent valued learning

Reflects Depth and Breadth of Expected Learning



VS.



Today's accountability tests:
limited treatment of target
domains

Multiple measures to
capture the full domain,
balanced and deep
coverage

Reflects Depth and Breadth of Expected Learning



Today's tests

VS.



Systems of assessment to capture rich portrait of proficiency

Performance and Other Interactive Assessments for Accountability

- Address capacities not well addressed by MC and short answer items:
 - ✓ *Depth of thinking and reasoning*
 - ✓ *Ability to apply knowledge, solve problems*
 - ✓ *Communicate knowledge*
 - ✓ *Utilize technology*
 - ✓ *Embedded in content*
- Model good practice

Why a *System* of Assessments?

- Multiple measures to adequately represent valued learning
- Multiple measures to serve intended purposes
- Coordinated use of data

WHY COHERENCE?

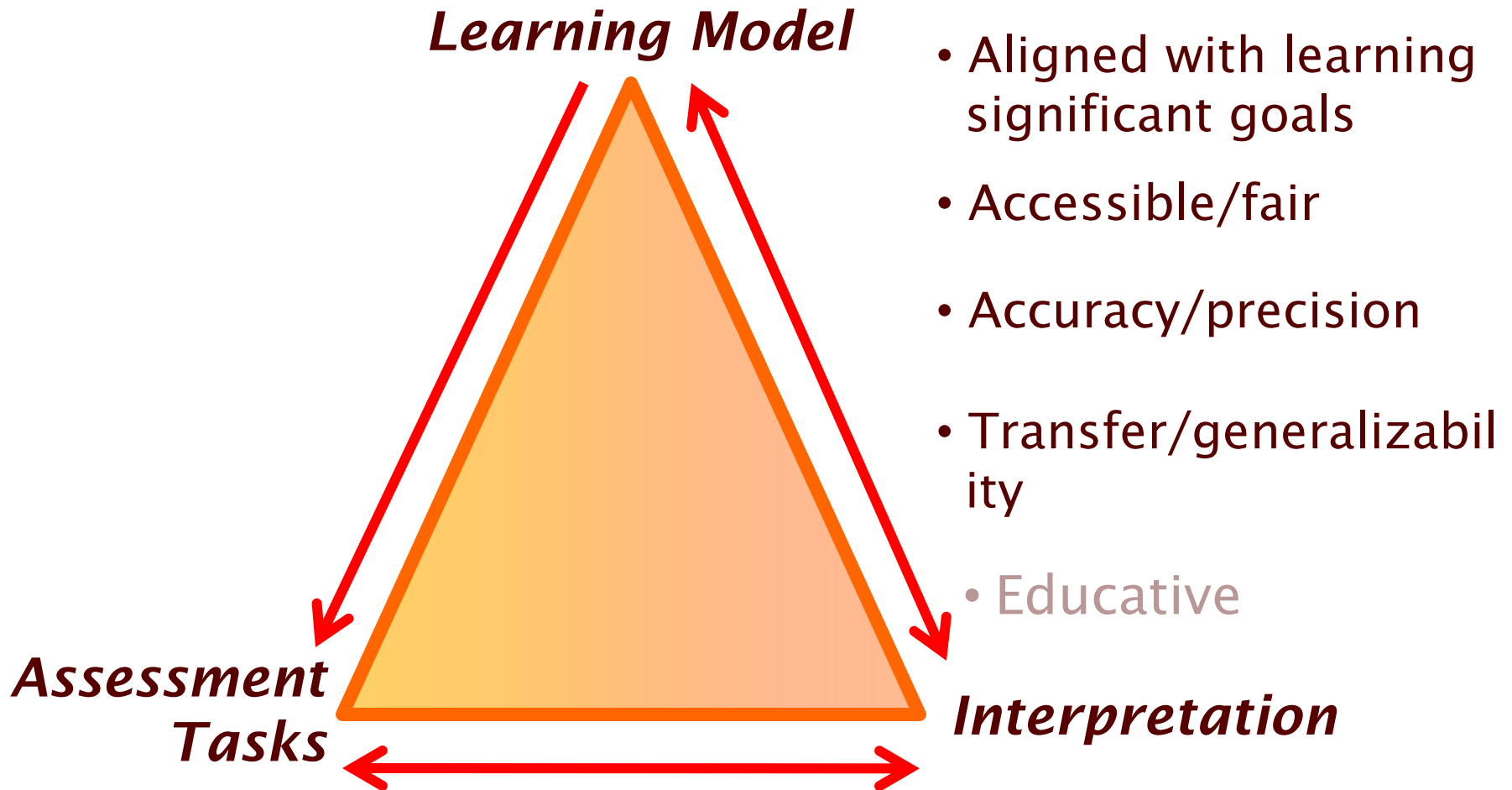
Consider the Alternative



Coherence:

- Fundamental: Measures significant learning
- Horizontal
- Developmental
- Vertical

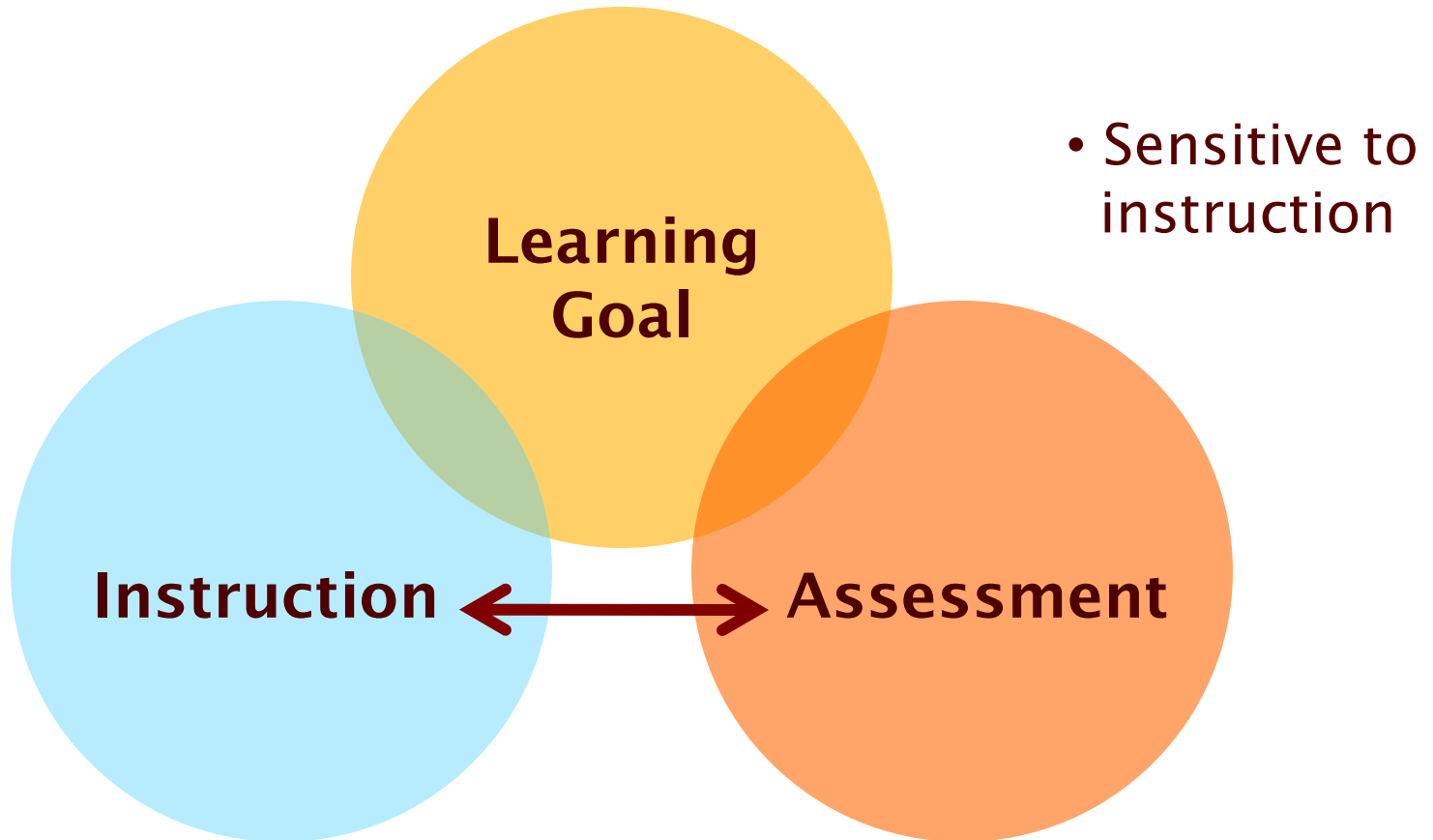
Fundamental: Valid Measure of Learning



SIDE BAR: What “Learning?”

- Prepared for what? RAPID CHANGE
 - ✓ *Knowledge, tools, jobs that don't yet exist*
 - ✓ *Knowledge economies, global and environmental pressures*
- 21st century skills embedded in CCL
 - ✓ *Ways of thinking*
 - ✓ *Ways of working*
 - ✓ *Tools for Working*
 - ✓ *Living in this world*

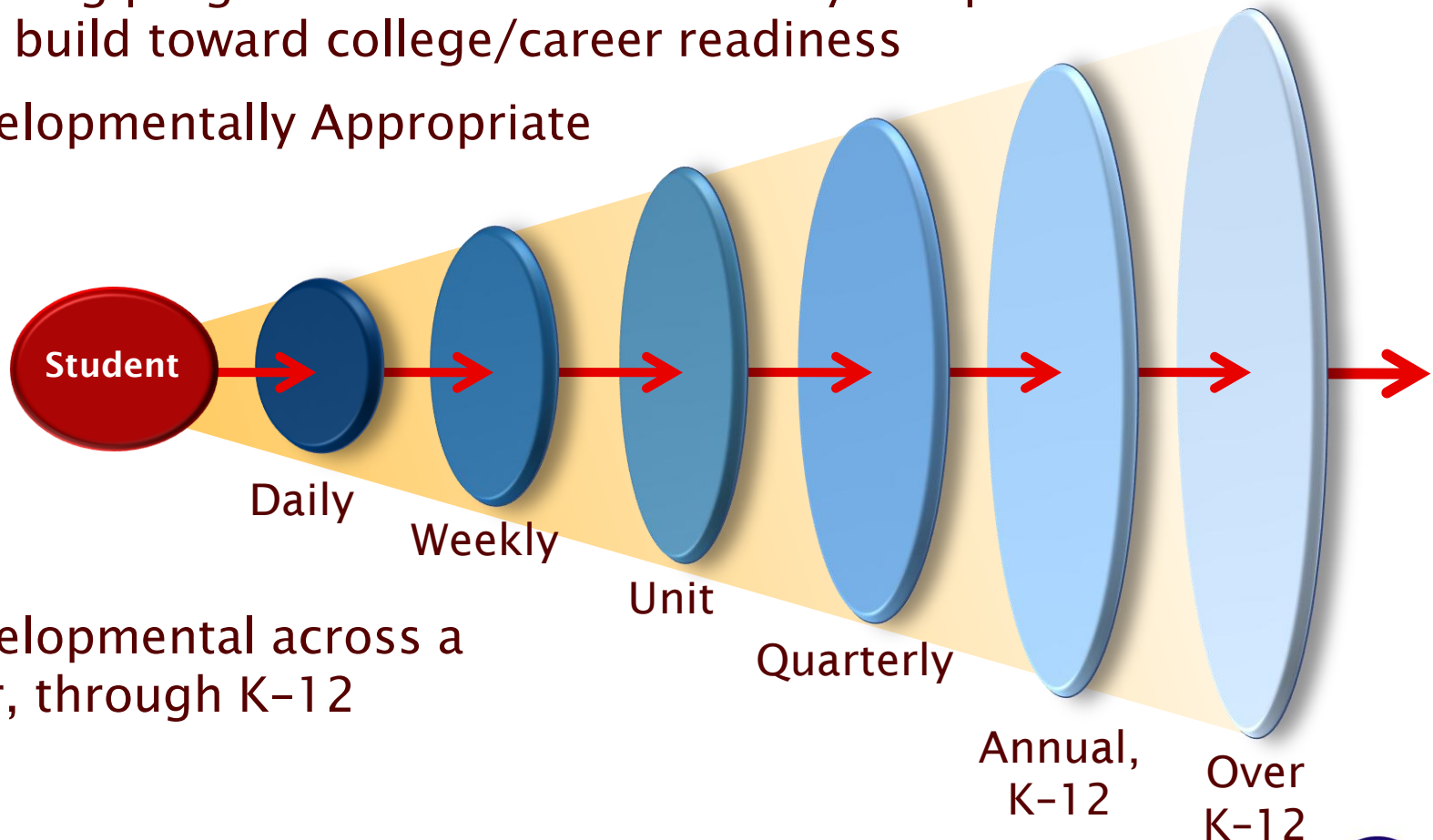
Horizontal Coherence



Developmental Coherence

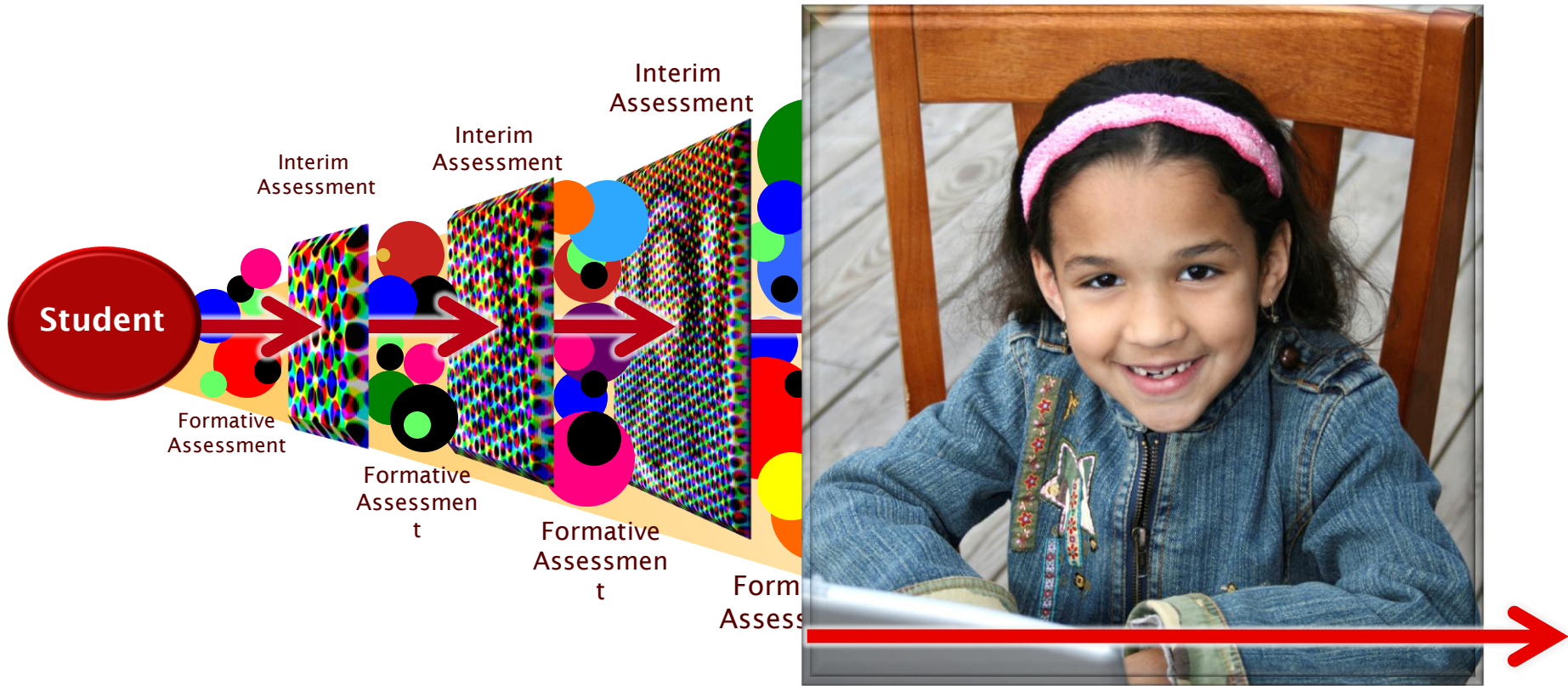
Learning Goals/Progression

- Learning progression-based: Assess key competencies that build toward college/career readiness
- Developmentally Appropriate



- Developmental across a year, through K-12

Coherence Supporting Learning



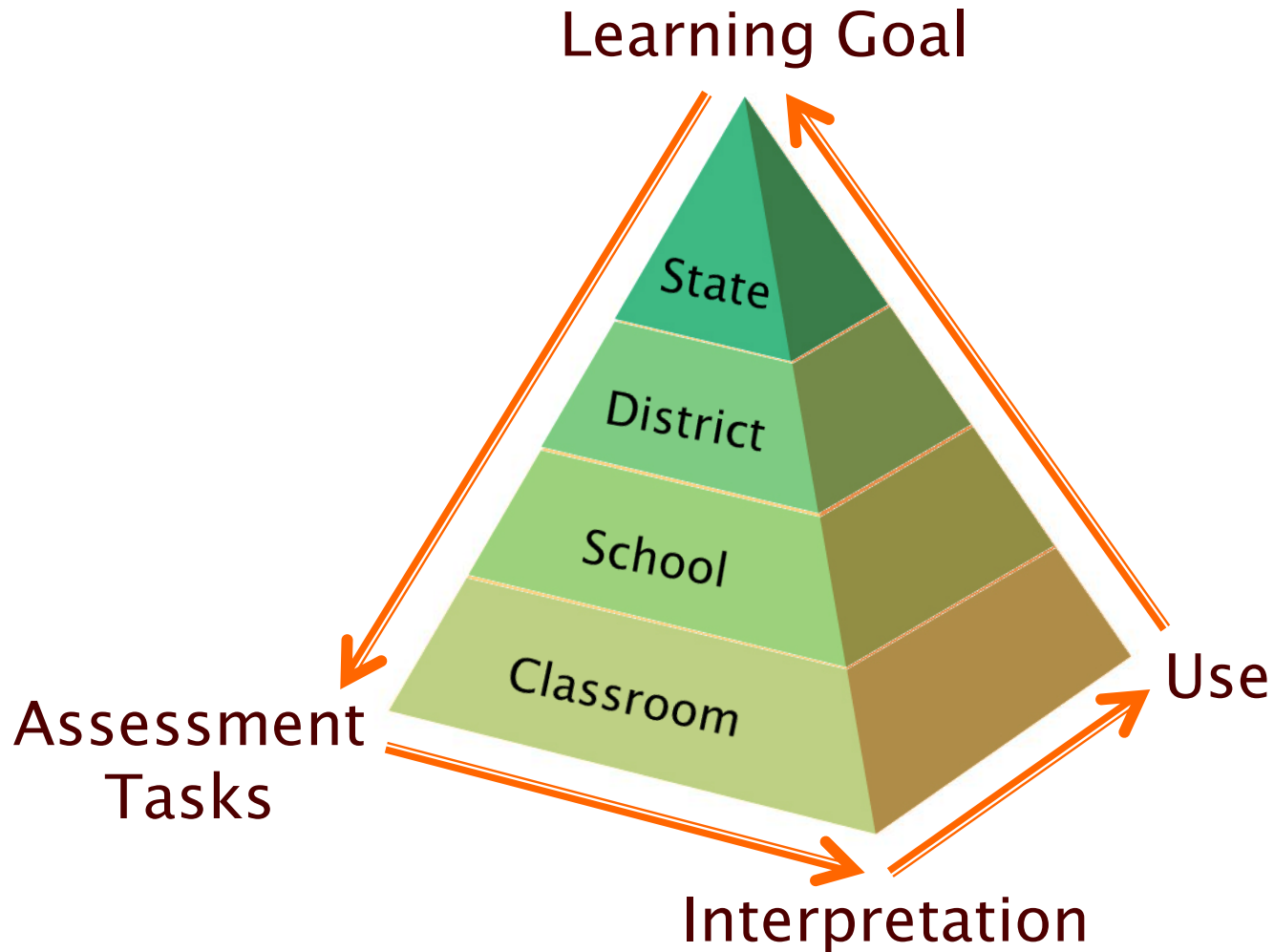
End of the Year Assessment
On track to college readiness

K-12 Coherence



Assessments Builds K-12
to College Readiness

Vertical Coherence



Federal Expectations

Purpose	Assessment Type	Primary Users	Use
Accountability/ Evaluation	On-demand annual	State, District, Schools, Teachers, Parents, Students, Public	<ul style="list-style-type: none"> ▪ School/Teacher/Principal effectiveness and capacity building ▪ Status/Growth re: college career readiness ▪ General Feedback curriculum strengths and weaknesses, student strengths and weaknesses? ▪ Recognize and build on excellence
Accountability/ Evaluation	Through Course, Curriculum Embedded	Same as above	<ul style="list-style-type: none"> ▪ Same as above
Monitoring/Super vision	Interim Assessment, Unit Exams, Curriculum embedded assignments	District /School Administrators, Teachers, Parents, Students	<ul style="list-style-type: none"> ▪ Identify struggling students, teachers schools ▪ Diagnose/adjust? ▪ Identifying promising practices
Formative	Curriculum-embedded, student work, class discourse, discussion	Teachers, Students	<ul style="list-style-type: none"> ▪ Inform immediate teaching and learning

Assessment Type: On-demand Annual

Purpose	Primary Users	Use
Accountability/ Evaluation Communicate Model	State, District, Schools, Teachers, Parents, Students, Public	<ul style="list-style-type: none">▪ School/Teacher/Principal effectiveness and capacity building▪ Status/Growth re: college career readiness▪ General Feedback: curriculum strengths and weaknesses, student strengths and weaknesses?▪ Recognize and build on excellence

Assessment Type: Through Course, Curriculum Embedded

Purpose	Primary Users	Use
<p>Accountability/ Evaluation</p> <p>Communicate</p> <p>Model Practice</p>	<p>State, District, Schools, Teachers, Parents, Students, Public</p>	<ul style="list-style-type: none"> ▪ School/Teacher/Principal effectiveness and capacity building ▪ Status/Growth re: college career readiness ▪ General Feedback: curriculum strengths and weaknesses, student strengths and weaknesses? ▪ Recognize and build on excellence

Assessment Type: Interim Assessment; Unit Exams; Curriculum Embedded Assignments or Projects

Purpose	Primary Users	Use
Monitoring/ Supervision Communicate Model	District/Schools Administrators, Teachers, Parents, Students	<ul style="list-style-type: none">▪ Identify struggling students, teachers schools▪ Diagnose/adjust ?▪ Identifying promising practices

Assessment Type: Curriculum-embedded, Student Work, Class Discourse, Discussion

Purpose	Primary Users	Use
Formative Communicate Model	Teachers, Students	<ul style="list-style-type: none">Inform immediate teaching and learning

System Validity

- To what extent does the system, its individual and collective measures, well serve intended purposes?
- To what extent do individual and collective measures provide accurate information for intended decision-making?
- To what extent do individual and collective measures enable all students to show what they know, demonstrate progress?
- What are system consequences? As intended? For individuals from special subgroups?
- What claims need to be satisfied?
- What evidence will satisfy those claims?
- What are design implications?

Some Validity Criteria For Accountability Assessments

- Learning-based, aligned
- Comprehensiveness
- Fairness, sensitive to growth on the full continuum
- Precision
- Comparability
- Transfer, predicts subsequent success
- Instructional sensitivity
- Educative, models good practice
- Consequences and capacity

Some Validity Criteria for Monitoring/ Supervision Assessments

- Depends on intended use
- Predictive: Performance on accountability
- Reliable diagnosis: At a gross level, at least
- Learning—trajectory-based?
- Instructionally tractable – for whom?
- Timely
- Educative? For whom
- Consequences, system capacity for use

Special Validity Criteria for Formative Assessments

- Learning-based
- Continuous
- Supports learning/instructional value
- Diagnostic
- Instructionally tractable for teachers and students
- Unique, responds to individual and class
- Educative
- Consequences, capacity to use

Going from Rhetoric to Reality

- Concept of “system,” rather than a single, annual test, is important to move forward
- But, there’s no magic band-aid
- Good systems, like good assessments, start with definition of purpose(s)
- Feasible systems start with clear priorities
- Systems that try to serve all, will well serve none

System Support for Change

- The goal: All students achieving to high levels, prepared for success in the 21st Century
- The goal: Dramatically different expectations for teaching and learning
- New assessments – system or not – do not create effective opportunity to learn nor change practice
 - ✓ *Necessary but not sufficient*
 - ✓ *Coordinated systems of support at and for every level:*
 - ❖ *Professional: intellectual, social capital*
 - ❖ *Technological*

Final Thoughts

- COHERENCE with priority purposes needs to be designed in
- Complex design problem with competing tensions
- Coordinated support systems must be part of the design process
- New methodologies and technologies will need to be invented
- Build and substantiate validity argument
- There is no single right answer



National Center for Research on Evaluation, Standards, & Student Testing

UCLA | Graduate School of Education & Information Studies

VISIT US ON THE WEB

cresst.org

herman@cse.ucla.edu