

Challenges to Developing and Implementing Instructionally Supportive Local Assessment Systems

Marge Petit & Scott Marion
Center for Assessment

June 24, 2003

CCSSO Conference on Large-Scale Assessment

www.nciea.org



The business of schools is to invent tasks, activities, and assignments that the students find engaging and that bring them into **profound interactions with content and processes** they will need to master to be judged well educated [emphasis added].

Schlechty, P.C. (2001) *Shaking up the schoolhouse*. San Francisco: Jossey-Bass



Session Outline

- Scott-Introduction
 - Marge-Aggregation
 - Brian-Commonality
 - Kim-Implementation Fidelity
-
- We are drawing on experiences across several states, not focusing on any one state.



Building a Conceptual Framework

- The National Research Council through the publication of “Knowing What Students Know” and workshops/study groups about “Bridging the Gap” offers the following characteristics of a balanced and coherent assessment system. These are a helpful for framing our work:



Comprehensiveness

- A range of measurement approaches are used to support educational decisions.
- Systems can (should) include both formative and summative assessments.
- The system can also include measures that address the quality of instruction.



Coherence

- **The conceptual base or models of learning underlying the assessments at all levels (large-scale to classroom) are compatible.**
- **Alignment is needed among standards, curriculum, instruction, assessment, and professional development so that all of the parts are working toward a common set of learning goals.**
 - ◆ This is especially challenging due a lack of understanding or, at least, a lack of explicitness about learning theory.



Continuous

- **In a coordinated system, assessments measure student progress over time, for example, over a school year, over several grades, or over a student's entire school experience.**



Integration

- **A coherent assessment system is integrated into a larger, coherent educational system that provides resources and professional development to ensure that teachers have the capacity to do what is expected based upon standards and assessments .**



Quality Assessments

- **In a coherent system, the large scale and classroom assessments that are included in the system exhibit are of high quality.**
 - ◆ This is often the weakest link in the chain.



What the Framework Omits

- In today's accountability world, local assessment systems are being asked to do more than “only” inform instruction—especially if they are to garner the resources necessary for sustaining these systems.
- We are being expected to define “proficient” or good enough as a result of these assessment systems. Some issues that surface when trying to use local assessment systems to define proficiency (or some other performance level) include:
 - ◆ Representation, Aggregation, and Feasibility
 - ◆ Commonality and replaceability of task
 - ◆ Implementation Fidelity



Representation and Aggregation



As the Center has worked with some states to build Local Assessment Systems, a *key question* has been...

If you have a wide array of assessments (multi-week to 10 minute quizzes), how can they

- be mapped to ensure adequate coverage (representation),
- be combined (aggregation) to provide reliable and valid decisions,
- support a coherent instructional program,
- and be feasible to implement?



Some Local Assessment Systems Purposes

- To inform instruction;
- To hold students accountable for learning at local or state level (e.g., graduation);
- To provide a rich source of data for school improvement.



Some Aggregation and Representation Considerations

- Structure of standards – where one aggregates to, and where one aggregates from;
- Link between assessment and purposes;
- Assessments – depth vs. breadth and how to capture (weight) the depth of the assessment;
- Common metric for aggregation;
- Standard Setting.



Standards Structure

Content Area

Content Area

Content Area

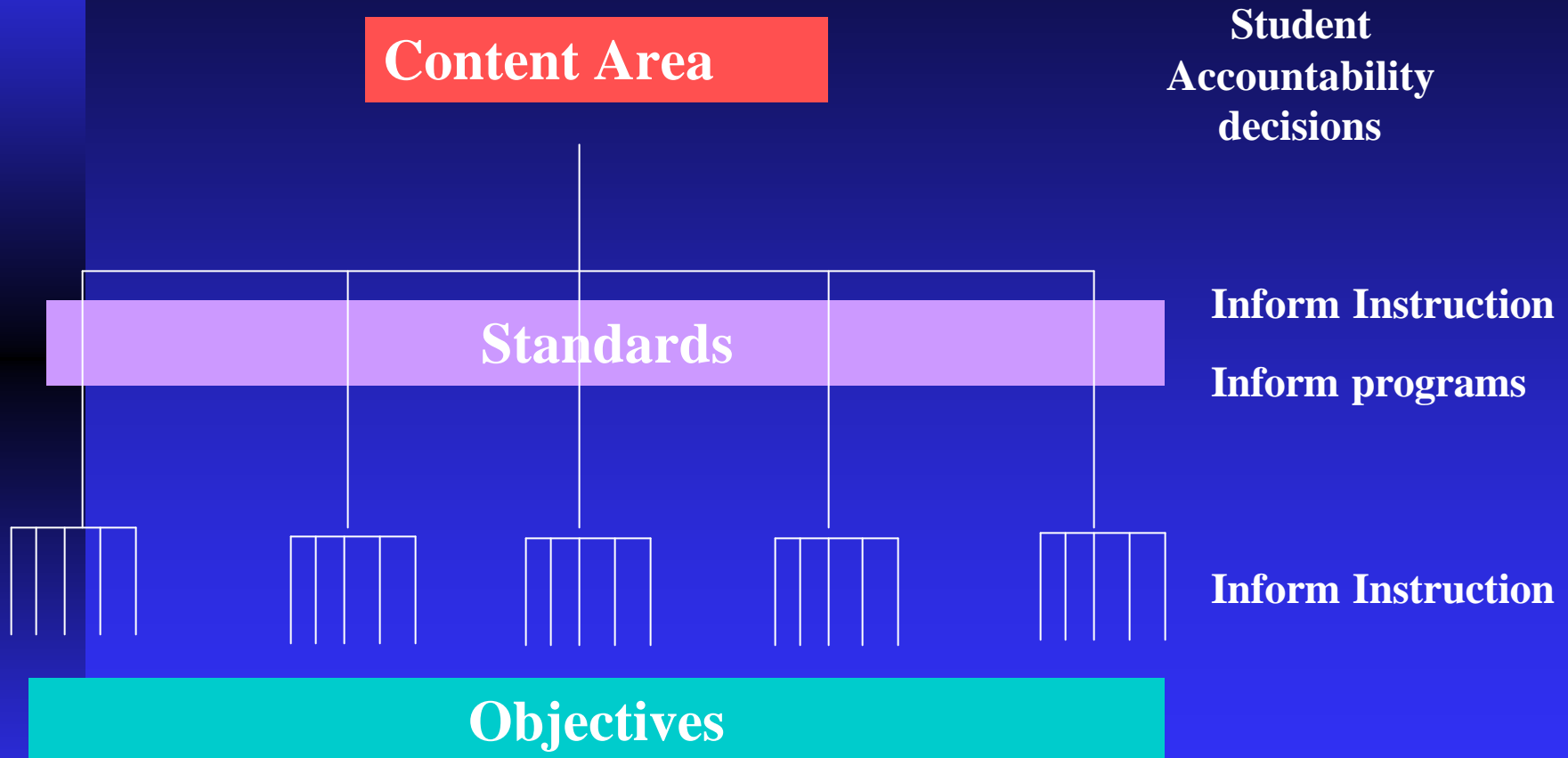
Content Area



Aggregation at
what level – for
what purposes?



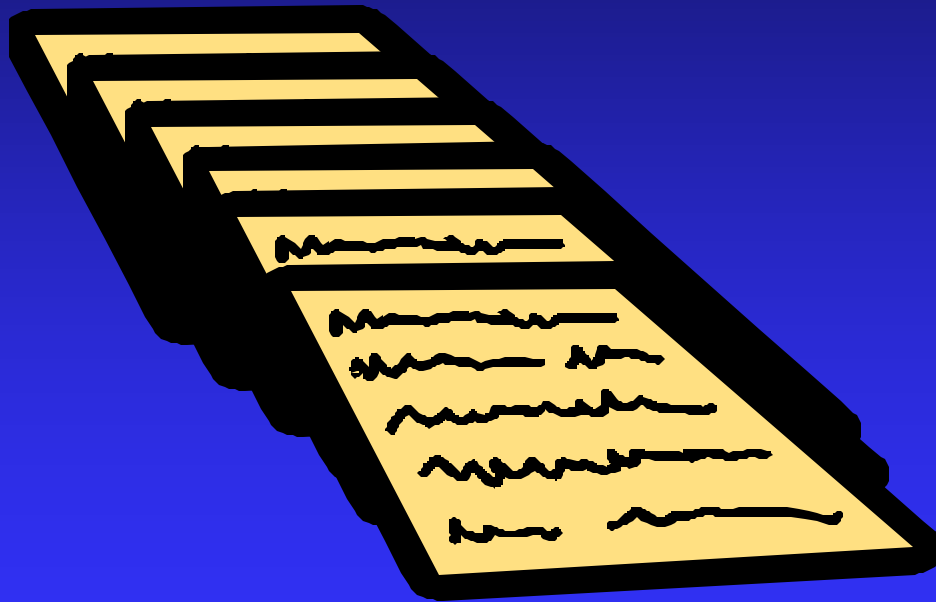
Relationship between Representation and Aggregation Linked to Purposes



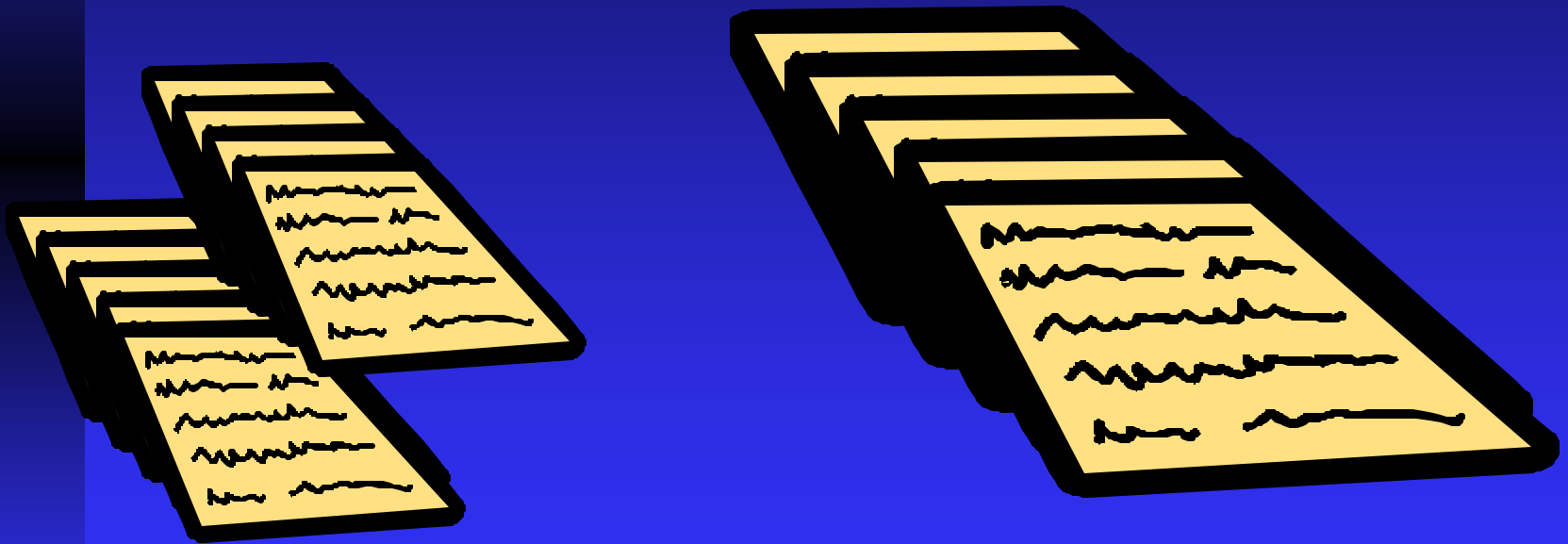
Representation: Lots of small pieces of information – heavy representation at the objective level



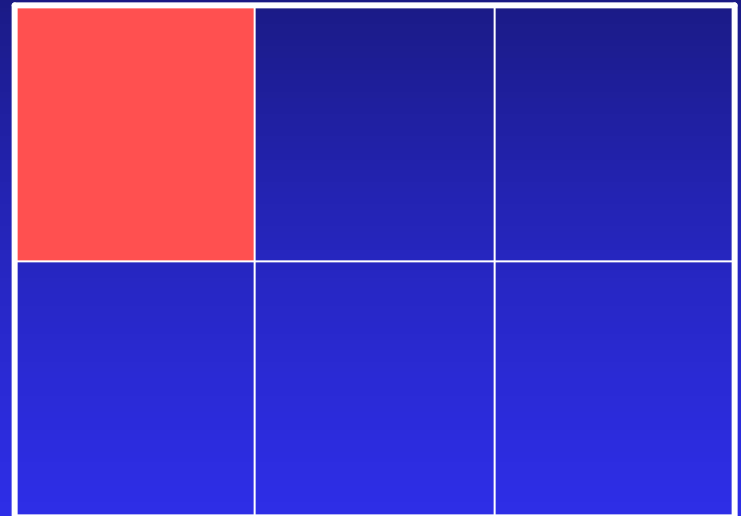
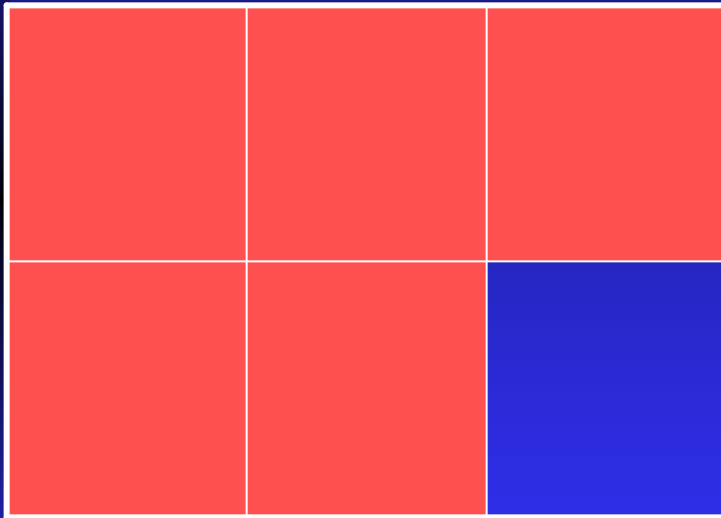
Representation: Fewer, but larger assessments



Representation: Mix of Assessments that address depth and breadth of expectations in standards

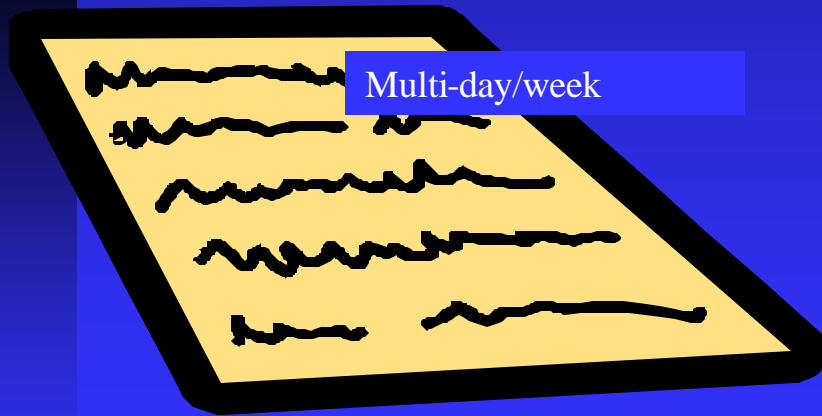


Two Examples: Proportion of All Assessments in LAS Applied to Student Accountability Decisions

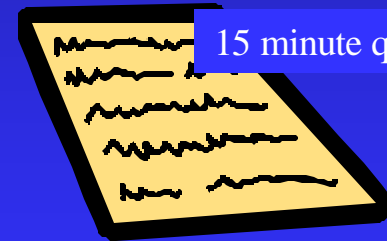


Representation: Appropriately Capturing the Depth and Breadth of Assessments

Caution: It is possible to assign a greater weight to assessments that assess breadth, rather than depth depending upon where you aggregate from.



Multi-day/week



15 minute quiz

Place Assessments on a Common Metric for Aggregation

Example - If a four point scale is used – the meaning of each point should be the same across assessments to be aggregated.

4 – Success in relationship to the demands on the assessment related to the expectations in the standards

3 – XXXX

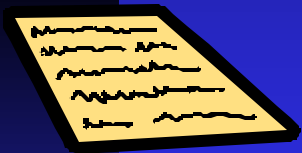
2 – AAAA

1 - CCCC

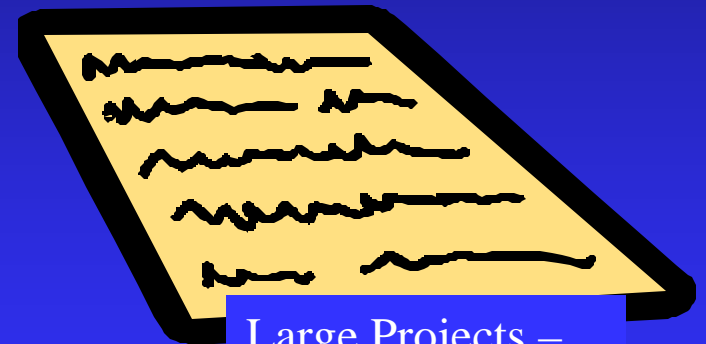
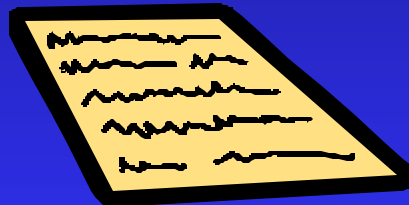


Place Assessments on a Common Metric for aggregation

The relative depth and breadth of the assessment should be reflected in the total points earned, but maintain the common metric.



Quizzes – Single 4 point scale



Large Projects – Multiple 4 points scales

A Case Study – Language Arts High School Graduation

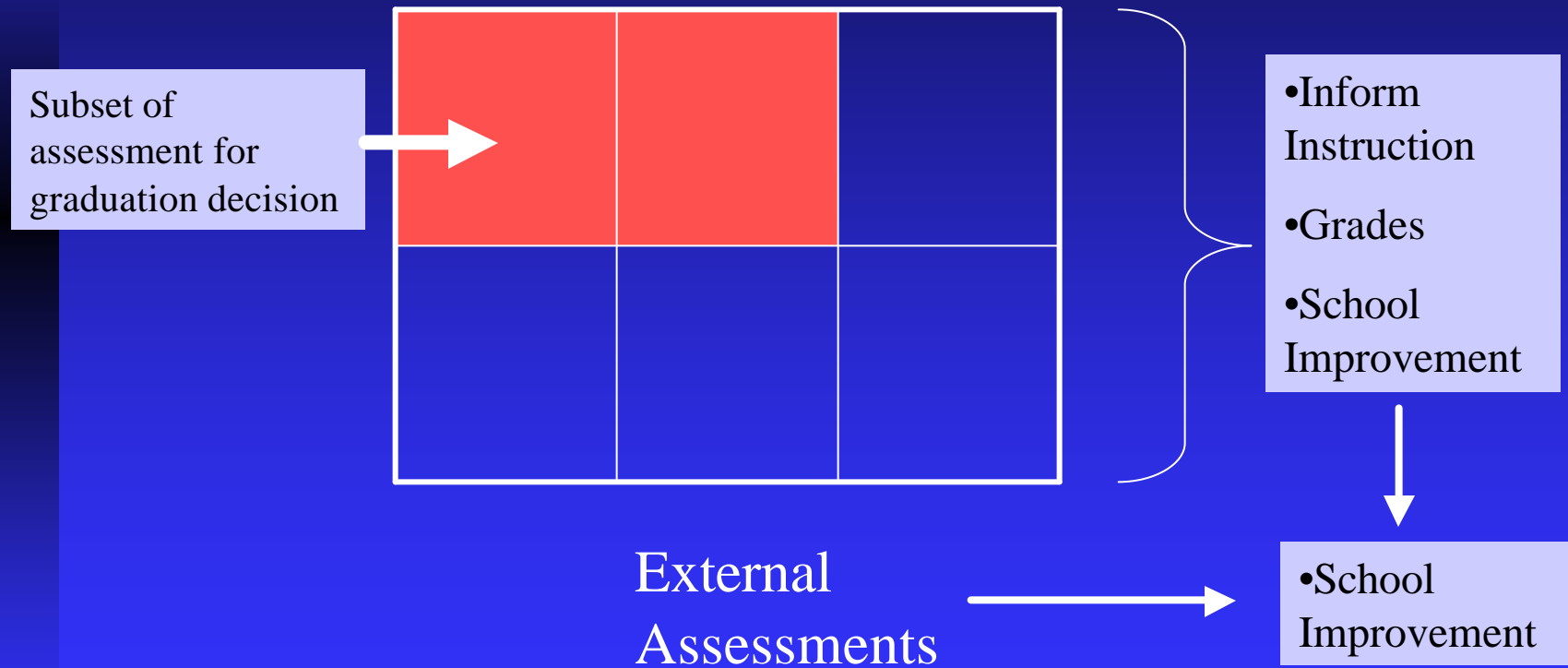


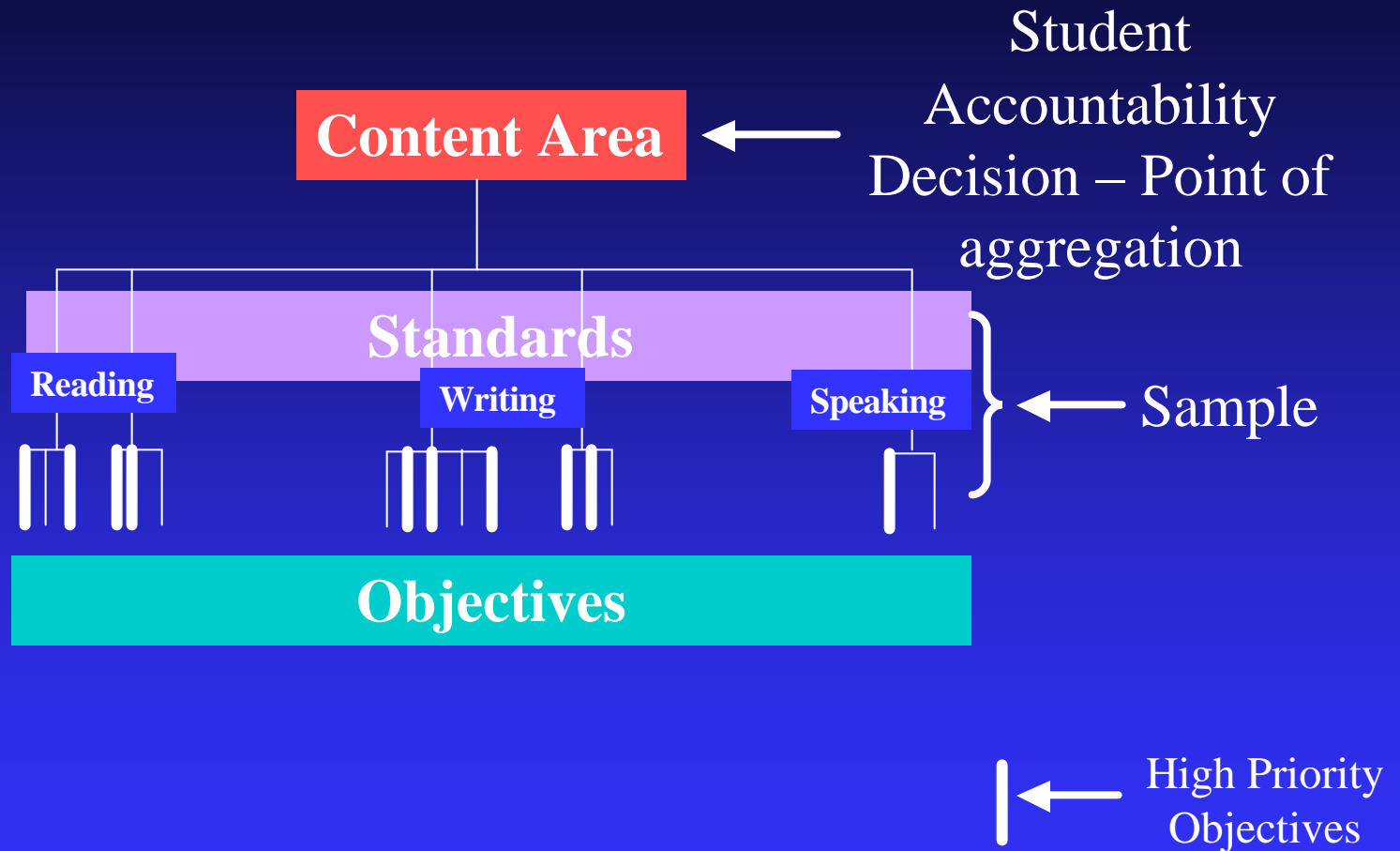
Decisions

- Align curriculum and course sequence to standards to assure that students have a full and fair opportunity to learn;
- Decide the proportion of assessments that serve the different purposes;
- Embed medium to large assessments into curriculum based on course sequence in school and opportunity to learn;
- Sample at the standards level based on a prioritization of the objectives within each standard to assure “sufficient” representation of the important concepts and skills within a standard;



Proportion of All Assessments in LAS Applied to Different Purposes





Embed Assessments into Curriculum

	Standards	English I	English II – World Literature	English III- American Literature	English IV - Electives	Notes
# 1	Reading Speaking and			0		
# 2	Writing Reading Writing		0	0		
# 3	Writing		0			
# 4	Reading Writing	0				
# 5	Reading and Writing	0	0			
# 6	Reading Speaking				0	
# 7	Writing	0				
# 8	Reading and Writing	0				
# 9	Writing			0		
# 10	Reading and Writing	0				
# 11	Reading and Writing				0	
# 12	Reading - Writing		0			



More Decisions

- Build rubrics consistent with demand in standards that have the same meaning at each point;
- Weight by the number of standards assessed. Larger assessment are designed to assess more standards.
- Aggregate at the highest level within each content area;
- Establish a “cut point” based on a pre-established definition of “proficient” and the meaning at each point on rubrics. Validate the “cut point.”



Representation and Aggregation

	Language Arts Standards				
	Reading		Writing		Speaking
	Reading I	Reading II	Writing I	Writing II	
# 1	4		4	4	4
# 2		4	4	4	
# 2	4		4	4	
# 3			4	4	
# 4	4		4	4	
# 5	4	4	4	4	
# 5	4	4	4	4	
# 6	4				8
# 7			4	4	
# 8	4	4	4	4	
# 9			4	4	
# 10	4		4	4	
# 11		4	4	4	
# 12	4		4	4	
	120 points possible (Cut point at 78 points)				



Validate the Cut Point with a Body of Student Work



Learnings

- Building on a coherent set of standards is essential;
- Prioritize standards and objectives to be assessed – you can't do it all;
- Include a mix of assessments that capture the depth and breadth of the standards;
- Place assessments on a common metric;
- Make decisions about the “weighting” of assessments to appropriately capture the depth and breadth of the assessments to be aggregated which is consistent with the common metric;



More Learnings

- Make clear decisions about the level of aggregation and for what purposes;
- Aggregate at the highest level for student accountability decisions to obtain the most reliable decision;
- Decide on systems of aggregation – compensatory (E.g. Mean, patterns of performance (E.g., Mode))
- Keep it feasible.



For more information

The Center for Assessment

www.nciea.org

mpetit@nciea.org

smarion@nciea.org



Check website for information on 2003 RILS conference sponsored by the Center and WestEd

- ◆ What states are doing with *No Child Left Behind*
- ◆ Reliability and *NCLB*
- ◆ Alignment and *NCLB*

