



University of Illinois at Chicago
Learning Sciences Research Institute



Next Generation Assessment: Going From CCSS & NGSS to SBAC & Beyond

Jim Pellegrino
University of Illinois
at Chicago



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.



Segment Topics

- Framing some of the National Discussion
- The SBAC Assessment System
 - Assessment Design within SBAC
 - SBAC College Readiness Policy & Achievement Level Descriptions
- Assessment aligned with Deeper Learning
- Coordinated & Coherent Assessment Systems
- Some Final Thoughts & Exhortations



A View from the White House

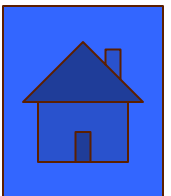
“I am calling on our nation’s Governors and state education chiefs to develop standards and assessments that don’t simply measure whether students can fill in a bubble on a test, but whether they possess 21st century skills like problem-solving and critical thinking, entrepreneurship and creativity.”

— President Barack Obama, March 2009



Some Elements of the National Conversation

- 2010: Common Core State Standards in English Language Arts and Mathematics
- 2011-13: NRC Science Framework & Next Generation Science Standards
- 2012-13: Calls for Assessment tied to Deeper Learning & 21st Century Skills
 - *Education for Life & Work: Developing Transferable Knowledge and Skills in the 21st Century*
 - *Gordon Commission on the Future of Educational Assessment*
 - *Policy Report on Criteria for High Quality Assessment*
- 2011-14: State Assessment Consortia – PARCC, SBAC, NCSC & DLM



COMMON CORE STATE STANDARDS FOR

English Language Arts
&
Literacy in History/Social Studies,
Science, and Technical Subjects





Big Shifts Signaled by CCSS

- Challenging text
- Close reading
- Informational text
- Multiple texts
- Disciplinary literacy (Grades 6-12)
- Argumentation
- 21st century research and communication tools
- Writing about sources

COMMON CORE STATE STANDARDS FOR

Mathematics



A FRAMEWORK FOR K-12 SCIENCE EDUCATION

Practices, Crosscutting Concepts, and Core Ideas

NATIONAL RESEARCH COUNCIL
OF THE NATIONAL ACADEMIES



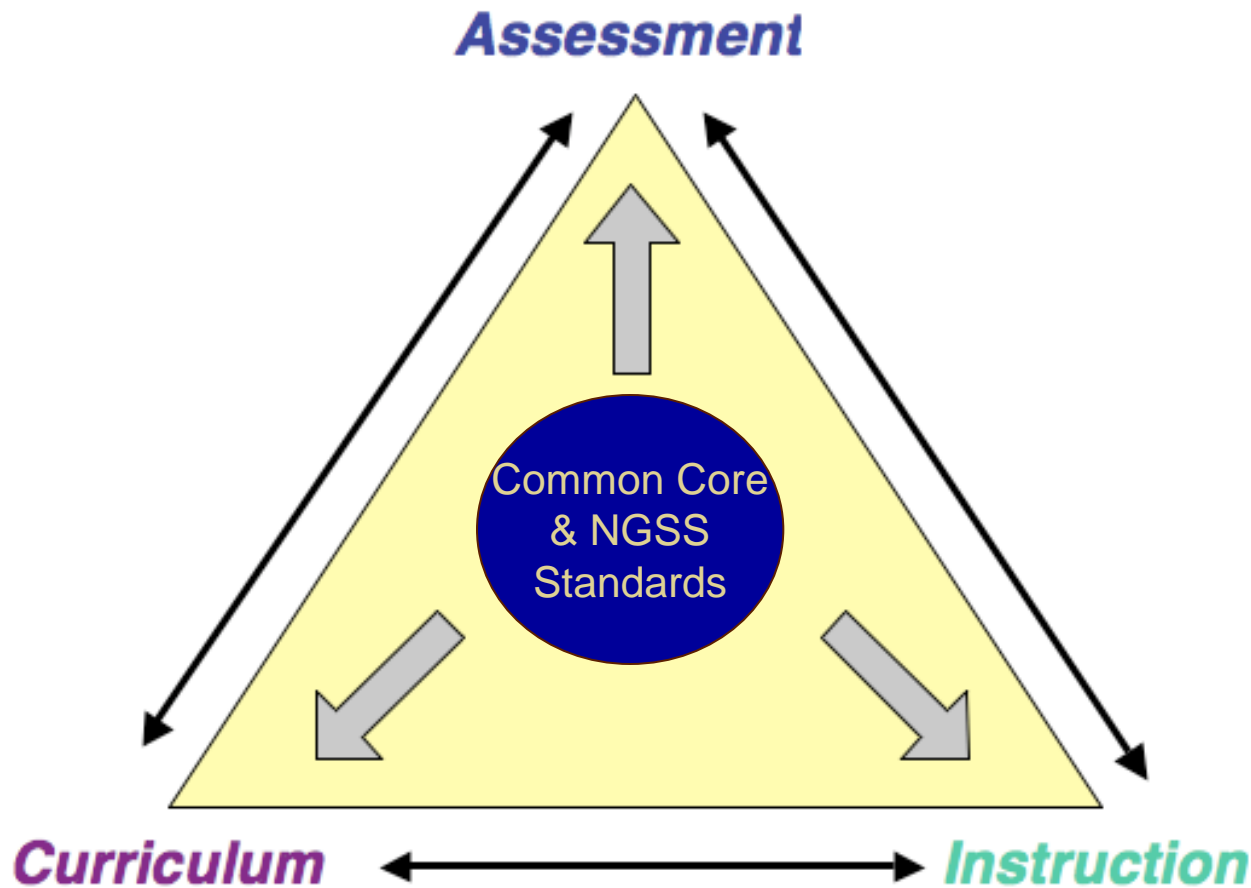


New Definitions of Competence

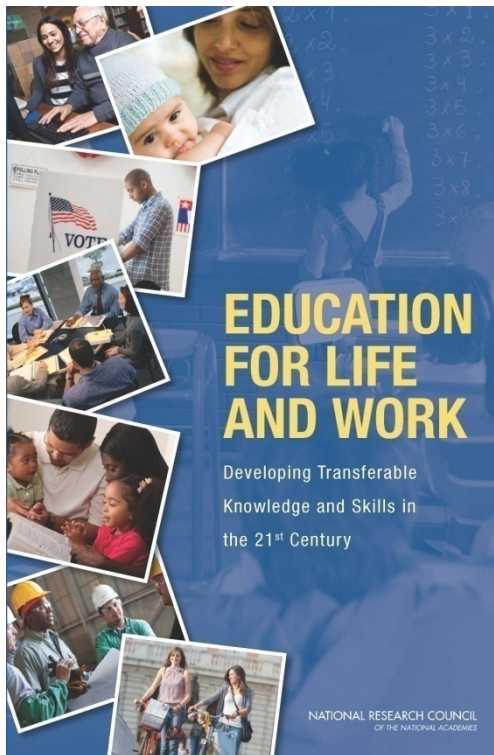
- Both the CCSS for Mathematics and the NRC Science Framework have proposed descriptions of student competence as being the intersection of knowledge involving:
 - **important disciplinary practices** and
 - **core disciplinary ideas**, with
 - **performance expectations** representing the intersection of core content and practices.
- Both view competence as something that develops over time & increases in sophistication and power as the product of coherent curriculum & instruction



Using Standards to Align Curriculum, Instruction & Assessment



Education for Life and Work: Developing Transferable Knowledge and Skills in the 21st Century



**Committee on Defining Deeper Learning and
21st Century Skills**

Division of Behavioral and Social Sciences and Education
National Research Council

Clarifying Terms

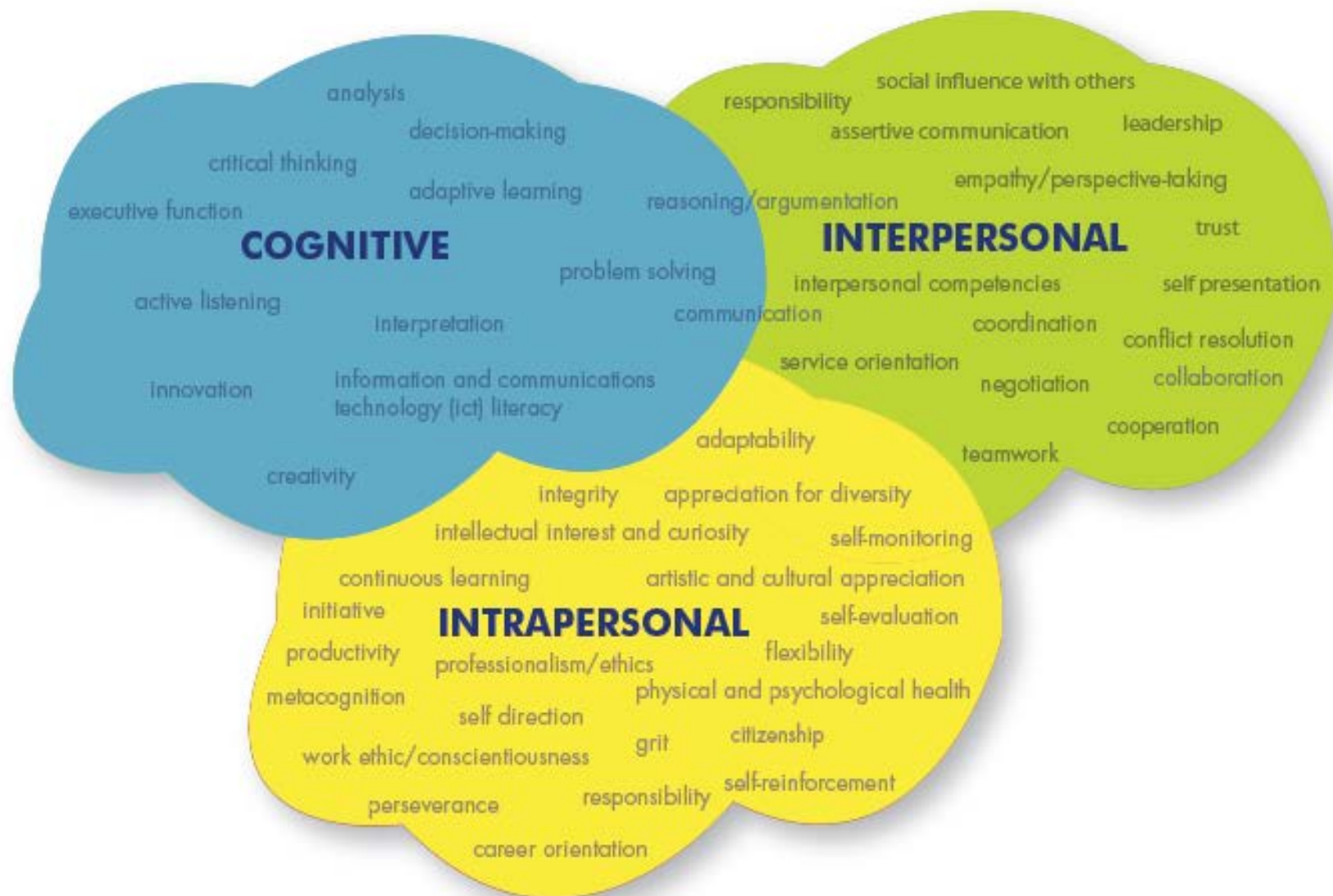
- Deeper learning is the process of learning for transfer. It enables an individual to take what was learned in one situation and apply it to new situations.
- The product of deeper learning is transferable knowledge, including content knowledge in a subject area and procedural knowledge of how, why, and when to apply this knowledge to answer questions and solve problems in the subject area.
- We refer to this transferable knowledge as “21st century competencies” to reflect that both skills and knowledge are included.



Three Domains of Competence

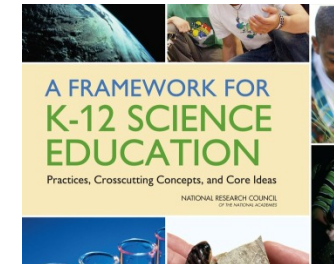
- **Cognitive:** reasoning and memory
- **Intrapersonal:** self-management
- **Interpersonal:** expressing ideas and interpreting and responding to others' messages
- The 3 domains are intertwined





Deeper Learning in the Disciplines

- The math and English CCSS and the NRC Science Framework each call for deeper learning
- A cluster of cognitive competencies – including critical thinking and constructing and evaluating evidence-based arguments – is strongly supported across all three disciplines.
- Coverage of competencies in the intrapersonal and interpersonal domains is present but uneven.
- Where standards documents do not explicitly overlap with 21st century competencies, there is little evidence of conflict between them.



Mathematics

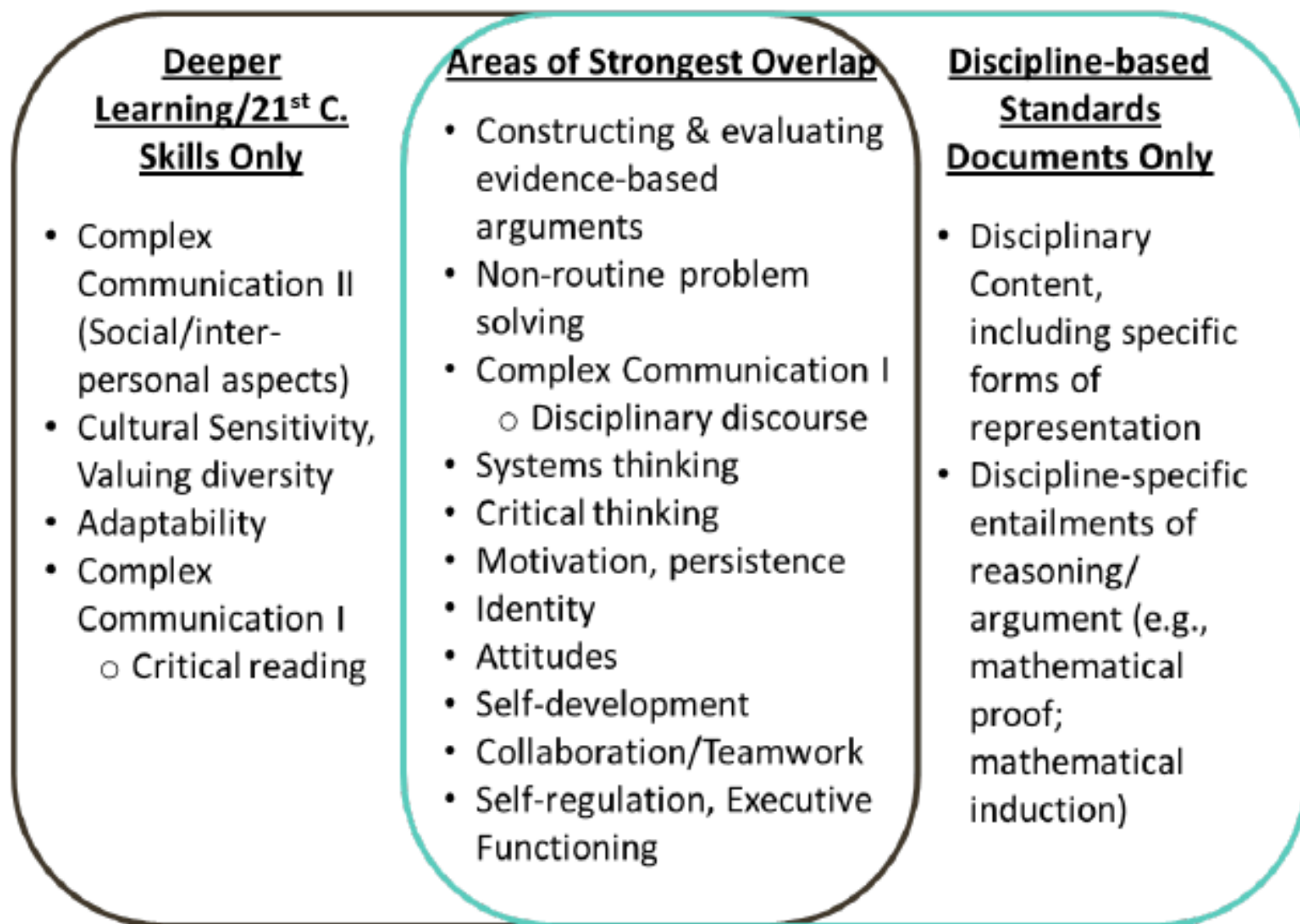
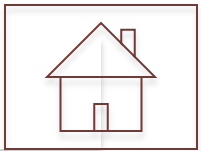


FIGURE 5–2 Overlap between CCSS math standards and 21st century skills.

Assessment Challenges

- Current educational policies and accountability systems rely on standardized assessments that focus primarily on recall of facts and procedures.
- These assessments are easily scored and quantified for accountability purposes. Although inexpensive, they are not optimal for assessing 21st century competencies.
- The extent to which the 21st century competencies articulated in standards documents will be emphasized depends on their inclusion in assessments.





A Public Policy Statement



The Gordon Commission
on the Future of Assessment in Education

Stanford Center for Opportunity Policy in Education



Criteria for High-Quality Assessment

By Linda Darling-Hammond, Joan Herman, James Pellegrino, Jamal Abedi, J. Lawrence Aber, Eva Baker, Randy Bennett, Edmund Gordon, Edward Haertel, Kenji Hakuta, Andrew Ho, Robert Lee Linn, P. David Pearson, James Popham, Lauren Resnick, Alan H. Schoenfeld, Richard Shavelson, Lorrie A. Shepard, Lee Shulman, Claude M. Steele

Published by:

Stanford Center for Opportunity Policy in Education,
Stanford University;

Center for Research on Student Standards and Testing,
University of California at Los Angeles; and

Learning Sciences Research Institute,
University of Illinois at Chicago

June 2013



scope
Stanford Center for
Opportunity Policy in Education



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



Gordon Commission View

“To be helpful in achieving the learning goals laid out in the Common Core, 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



Criteria for High-Quality Assessment



Five Major Criteria

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. Assessments that are Valid, Reliable & Fair

1: 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 (e.g., Depth of Knowledge Levels Two, Three, or Four in Webb's Taxonomy or the equivalent)
- ✓ 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 (DOK Levels Three or Four)

2: High-Fidelity Assessment of Critical Abilities

Critical abilities outlined in the Standards are evaluated using high-fidelity tasks that use the skills in authentic applications:

- ✓ Research, including analysis and synthesis of information
- ✓ Experimentation and evaluation
- ✓ Oral communications: speaking and listening
- ✓ Written communications: reading and writing
- ✓ Use of technology for accessing, analyzing, and communicating information
- ✓ Collaboration
- ✓ Modeling, design, and problem solving using quantitative tools

3: Standards that Are Internationally Benchmarked

- ✓ Calibration to PISA, International Baccalaureate, or other internationally comparable assessments (based on evaluation of content comparability, performance standards, and analysis of student performance on embedded items)

4: Items that Are Instructionally Sensitive and Educationally Valuable

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

5: Assessments that Are Valid, Reliable, and Fair

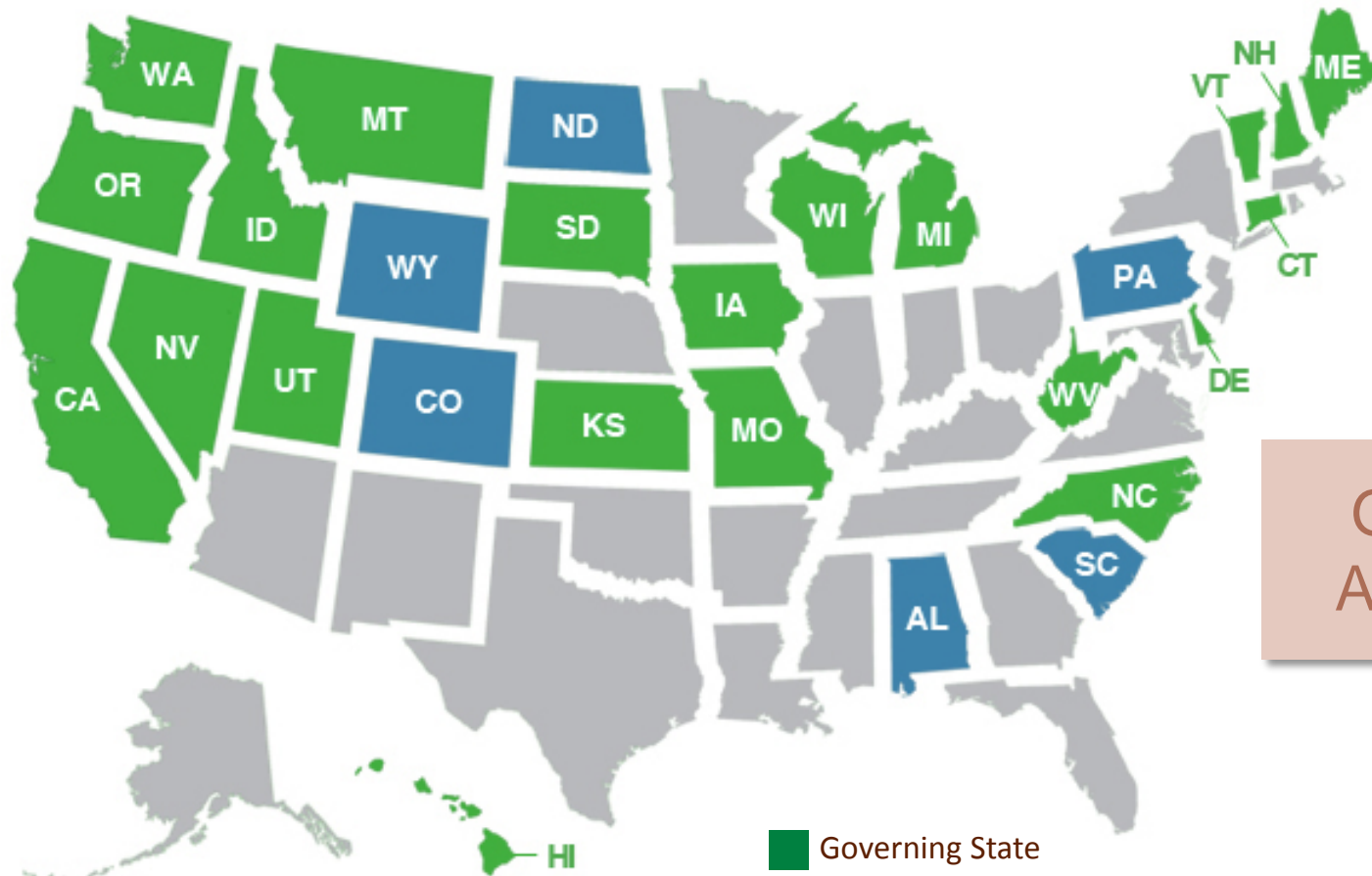
- ✓ Evidence that the intended knowledge and skills are well measured
- ✓ Evidence that scores are related to the abilities they are meant to predict
- ✓ Evidence that the assessments are well-designed and valid for each intended use — and that uses are appropriate to the test purposes and validity evidence
- ✓ Evidence that the assessments are unbiased and fairly measure the knowledge and skills of students from different language, cultural, and income backgrounds, as well as students with learning differences
- ✓ Evidence that the assessments measure students learning accurately along a continuum of achievement, consistent with the purposes the assessments are intended to serve





Learning Sciences Research Institute
University of Illinois at Chicago

Smarter Balanced Assessment Consortium



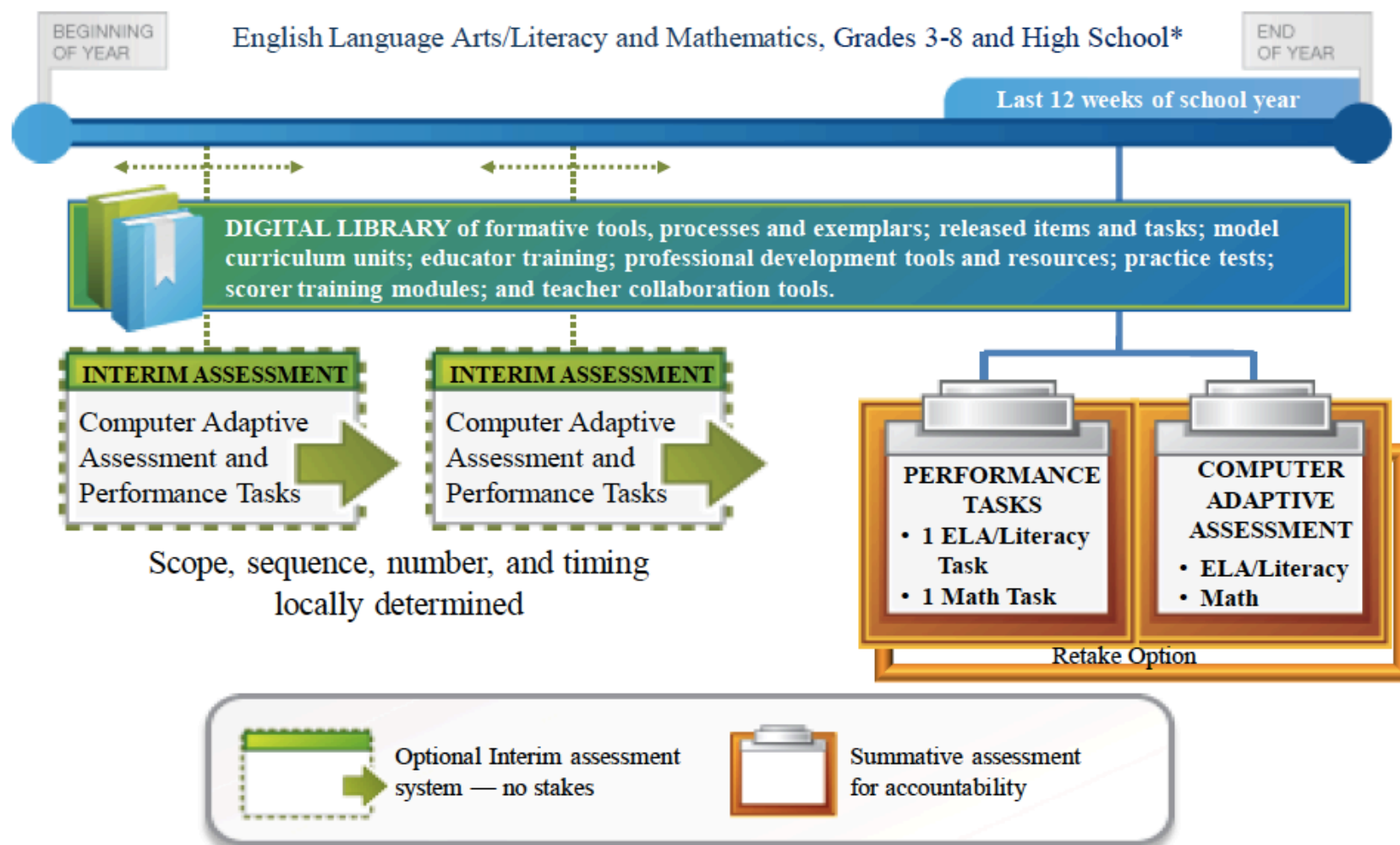
Next
Generation
Assessment
System

Governing State

Advisory State

Membership status as of March 6, 2012

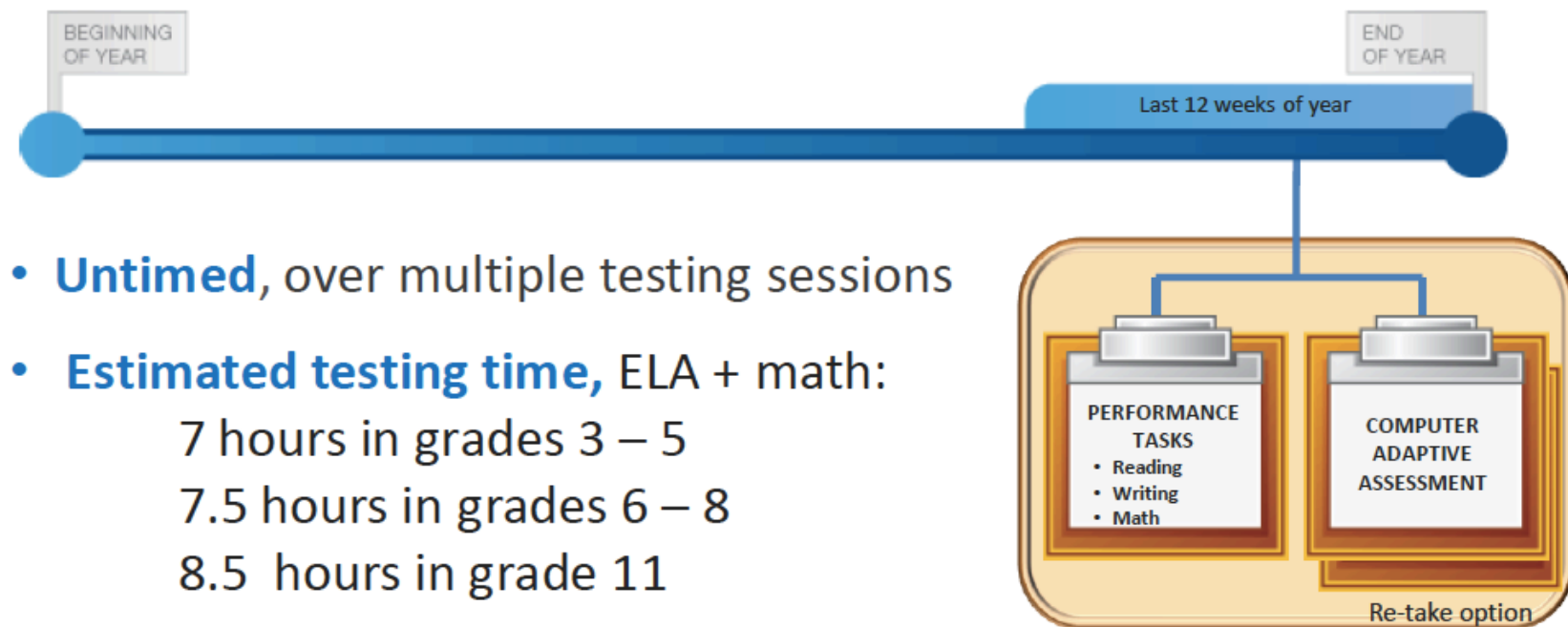
Smarter Balanced Assessment System



* Summative and interim assessments for grades 3 – 8 and 11, with additional supporting assessments for grades 9 and 10.

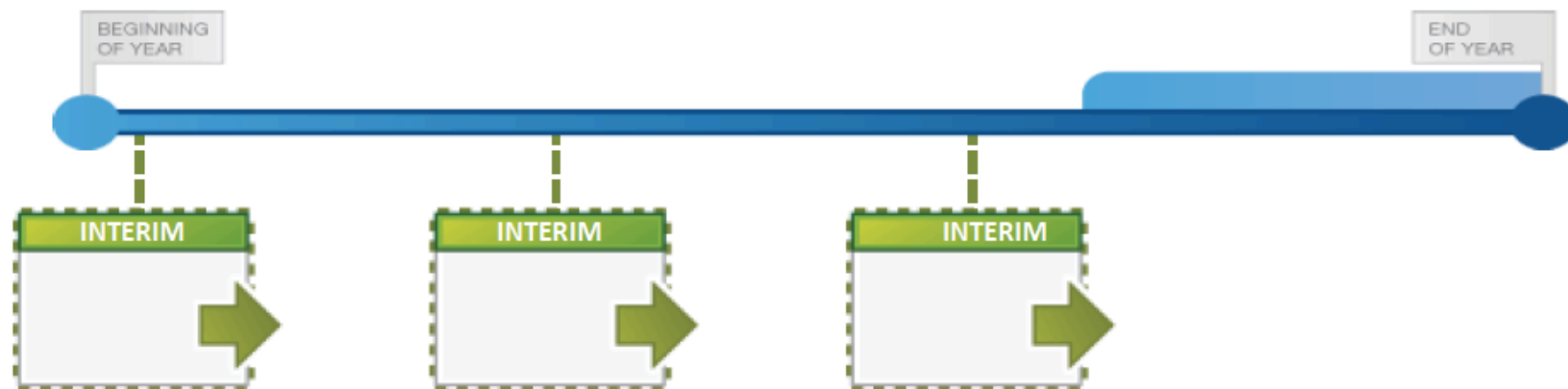
** Time windows may be adjusted based on results from the research agenda and final implementation decisions.

Smarter Balanced Summative Components



- **Untimed**, over multiple testing sessions
- **Estimated testing time**, ELA + math:
 - 7 hours in grades 3 – 5
 - 7.5 hours in grades 6 – 8
 - 8.5 hours in grade 11
- **PT and EOY scores combined** for accountability
- **1 Retake**, if locally approved, for testing irregularities
- **Paper and pencil version for 3 years**; thereafter as accommodation

Smarter Balanced Supports: Optional Interim Assessment System



- **Computer adaptive**, multiple item types from EOY summative
- **Customizable**: number, scope, timing
- **Non-secure** and fully accessible -- teachers can see items and student responses
- **Produces links** to student resources and PD resources



Smarter Balanced Supports: The Digital Clearinghouse



- **Reporting suite** with differentiated tools for users
- **Online Practice Test utility**
- **Assessment literacy modules**
- **Formative tools, exemplar instructional modules, vetted and rated units**
- **Released performance tasks and rubrics**
- Issue-focused **chat rooms** for teachers

Smarter Balanced: Supports and Timeline

Plans as of spring 2013, subject to change

2012 - 2013

- **Prototype items & tasks** released (www.smarterbalanced.org)
- **Practice Test** utility available to teachers, students, public

Summer/fall 2013

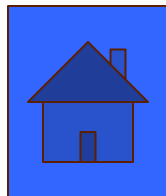
- **Teacher cadres** (avg. 90 per state) trained in use of formative tools and PD modules; ready to lead in-state trainings (Summer)
- First set of **Exemplary Instructional Modules** released (Fall)

Winter 2014

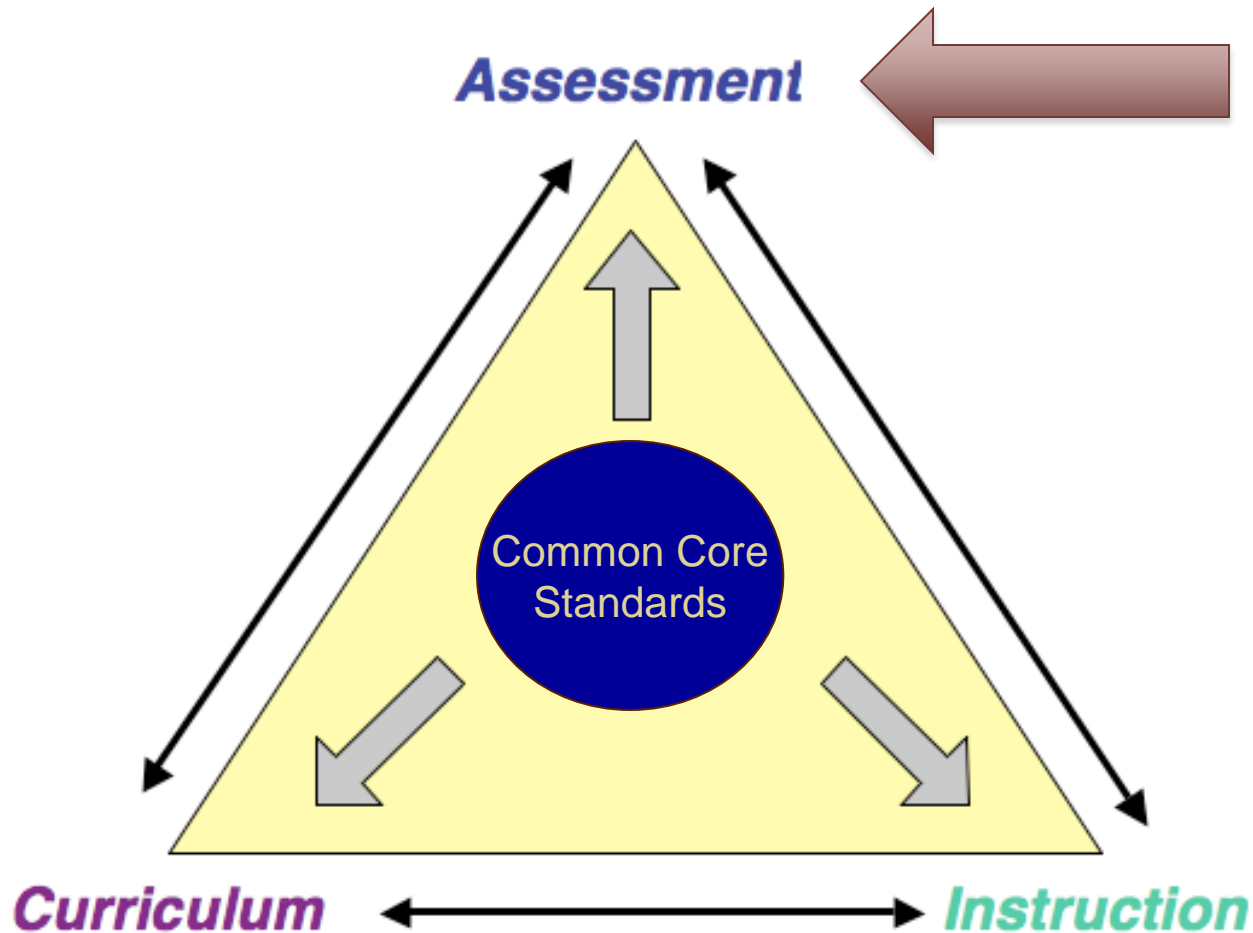
- Second set of **Exemplary Instructional Modules** released

Fall 2014

- Comprehensive Electronic Platform, including **Digital Library** launched
- **Interim assessments** available



Aligning Large-Scale Assessment with the Common Core Standards





Learning Sciences Research Institute
University of Illinois at Chicago

Common Core State Standards

COMMON CORE STATE STANDARDS FOR

English Language Arts
&
Literacy in History/Social Studies,
Science, and Technical Subjects

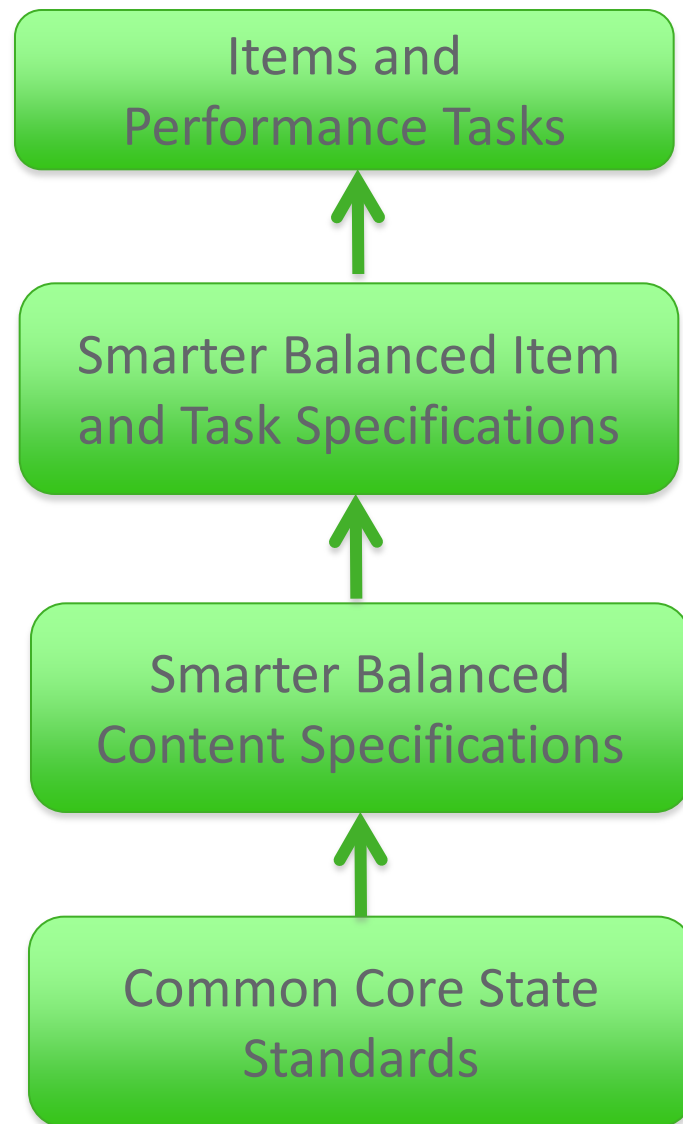


COMMON CORE STATE STANDARDS FOR

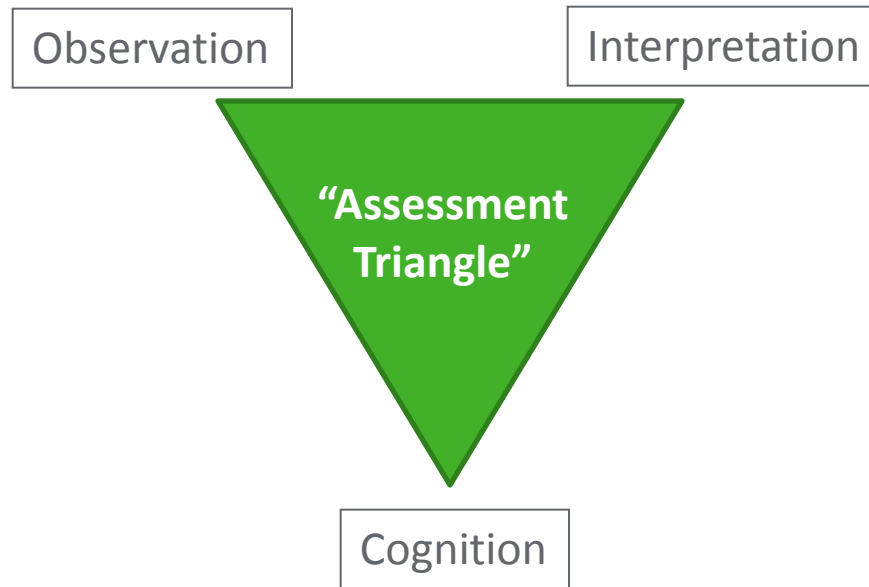
Mathematics



Smarter Balanced and Evidence-Centered Design

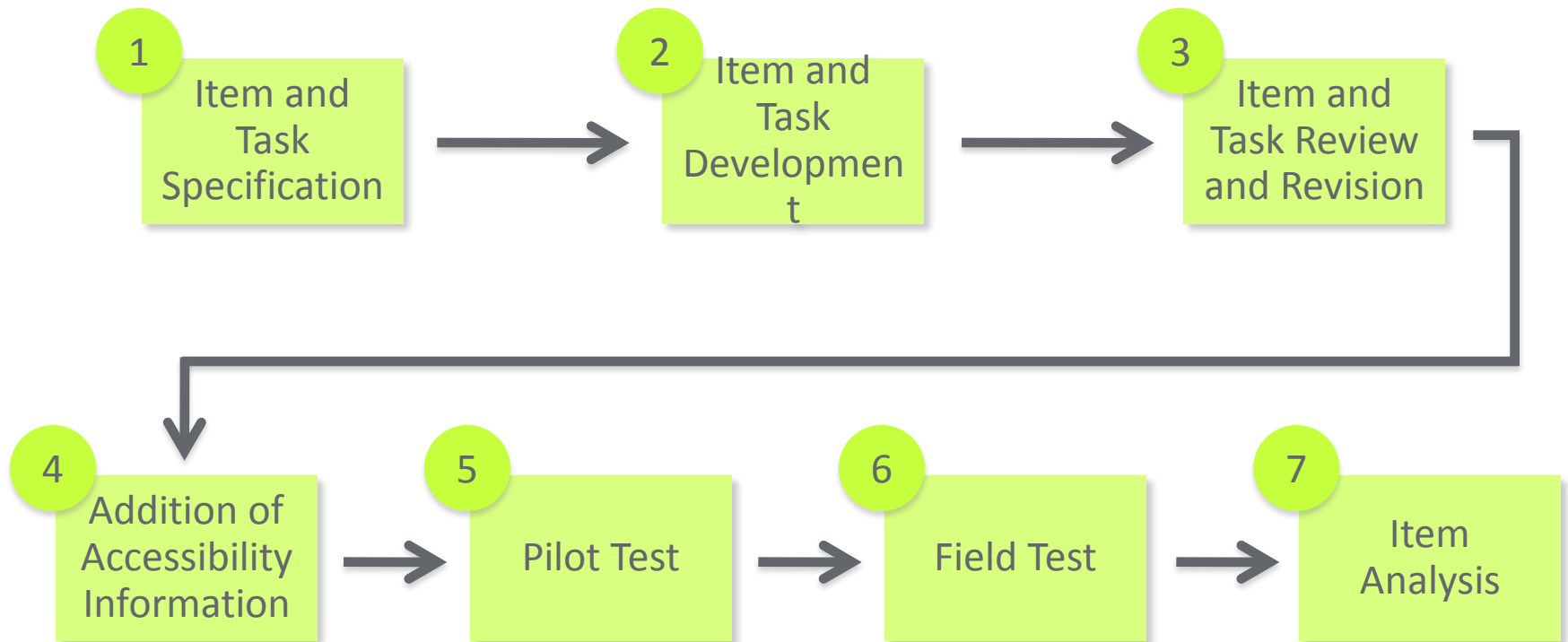


Evidence-Centered Design



- Item development approach that defines claims about students and their learning
- Evidence needed to support claims
- Types of items and tasks needed to collect evidence

Smarter Balanced Item Development Process





Six Item Types

- Selected Response
- Constructed Response
- Extended Response
- Performance Tasks
- Technology-Enabled
- Technology-Enhanced



Selected Response

Multiple Correct Options

Which of the following statements is a property of a rectangle? Select all that apply.

- ☐ Contains three sides
- ☐ Contains four sides
- ☐ Contains eight sides
- ☐ Contains two sets of parallel lines
- ☐ Contains at least one interior angle that is acute
- ☐ Contains at least one interior angle that is obtuse
- ☐ All interior angles are right angles
- ☐ All sides have the same length
- ☐ All sides are of different length



Constructed Response

Extended Response

Ms. McCrary wants to make a rabbit pen in a section of her lawn.

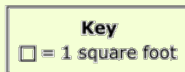
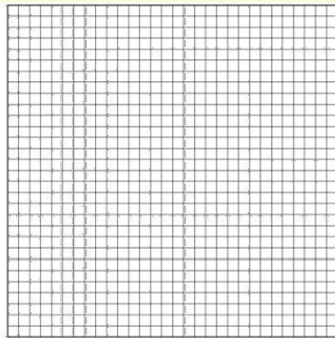
Her plan for the rabbit pen includes the following:

- It will be in the shape of a rectangle.
- It will take 24 feet of fence material to make.
- Each side will be longer than 1 foot.
- The length and width will measure whole feet.

Part A

Draw 3 **different** rectangles that can each represent Ms. McCrary's rabbit pen. Be sure to use all 24 feet of fence material for each pen.

Use the grid below. Click the places where you want the corners of your rectangle to be. Draw one rectangle at a time. If you make a mistake, click on your rectangle to delete it. Continue as many times as necessary.



Use your keyboard to type the length and width of each rabbit pen you draw. Then type the area of each rabbit pen. Be sure to select the correct unit for each answer.

[Students will input length, width, and area for each rabbit pen. Students will choose unit from drop down menu.]

Pen 1:

Length: (feet, square feet)

Width: (feet, square feet)

Area: (feet, square feet)

Pen 2:

Length: (feet, square feet)

Width: (feet, square feet)

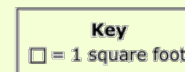
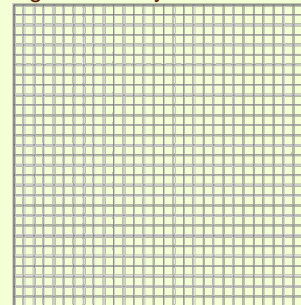
Area: (feet, square feet)

Part B

Ms. McCrary wants her rabbit to have more than 60 square feet of ground area inside the pen. She finds that if she uses the side of her house as one of the sides of the rabbit pen, she can make the rabbit pen larger.

- Draw another rectangular rabbit pen.
- Use all 24 feet of fencing for 3 sides of the pen.
- Use one side of the house for the other side of the pen.
- Make sure the ground area inside the pen is greater than 60 square feet.

Use the grid below. Click the places where you want the corners of your rectangle to be. If you make a mistake, click on your rectangle to delete it.



Use your keyboard to type the length and width of each rabbit pen you draw. Then type the area of each rabbit pen. Be sure to select the correct unit for each answer.

Length: (feet, square feet)

Width: (feet, square feet)

Area: (feet, square feet)



Performance Task

Student Directions:

Part 1 (35 minutes)

Your assignment:

You will read a short story and article, watch a video, review research statistics, and then write an argumentative essay about your opinion on virtual schools.

Steps you will be following:

In order to plan and compose your essay, you will do all of the following:

1. Read a short story and article, watch a video, and review research statistics.
2. Answer three questions about the sources.
3. Plan and write your essay.

Directions for beginning:

You will now read the sources and watch a video. Take notes, because you may want to refer back to your notes while writing your essay. You can refer back to any of the sources as often as you like.

- (short story)
- (article 1)
- (video)
- (research statistics)

Questions

Use your remaining time to answer the questions below. Your answers to these questions will be scored. Also, they will help you think about the sources you've read and viewed, which should help you write your essay. You may click on the appropriate buttons to refer back to the sources when you think it would be helpful. You may also refer to your notes. Answer the questions in the spaces provided below them.

1. Analyze the different opinions expressed in "The Fun They Had" and the "Virtual High School Interview" video. Use details from the story and the video to support your answer.
2. What do the statistics from "Keeping Pace with K–12 Online Learning" suggest about the current trends of virtual schools in the U.S.? Use details from the charts to support your answer.

3. Explain how the information presented in the "Virtual High School Interview" video and the article "Virtual Schools Not for Everyone" differs from the information in the research statistics? Support your answers with details from the video and the articles.

Part 2 (85 minutes)

You will now have 85 minutes to review your notes and sources, and to plan, draft, and revise your essay. You may also refer to the answers you wrote to the questions in part 1, but you cannot change those answers. Now read your assignment and the information about how your essay will be scored, then begin your work.

Your Assignment

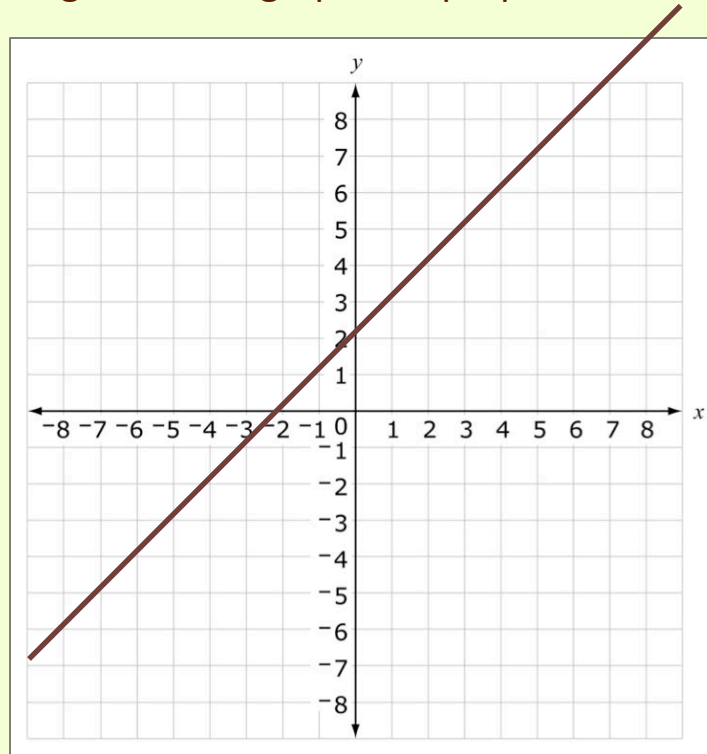
Your parents are considering having you attend a virtual high school. Write an argumentative essay explaining why you agree or disagree with this idea. Support your claim with evidence from what you have read and viewed.



Technology-Enhanced

Collects Evidence through a Non-Traditional Response

The value of y is proportional the the value of x . The constant of proportionality for this relationship is 1. On the grid below, graph this proportional relationship.





SBAC College Readiness

College Content-Readiness Definition

English Language Arts/Literacy ³	Students who perform at the College Content-Ready level in English language arts/literacy demonstrate reading, writing, listening, and research skills necessary for introductory courses in a variety of disciplines. They also demonstrate subject-area knowledge and skills associated with readiness for entry-level, transferable, credit-bearing English and composition courses.
Mathematics	Students who perform at the College Content-Ready level in mathematics demonstrate foundational mathematical knowledge and quantitative reasoning skills necessary for introductory courses in a variety of disciplines. They also demonstrate subject-area knowledge and skills associated with readiness for entry-level, transferable, credit-bearing mathematics and statistics courses. .



SBAC's Achievement Levels

Policy ALDs. The overall claim was delineated into the following four levels (with the defining phrases⁴ bolded):

The Level 4 student demonstrates **thorough understanding of and ability to apply** the English language arts and literacy (mathematics) knowledge and skills needed for success in college and careers, as specified in the Common Core State Standards.

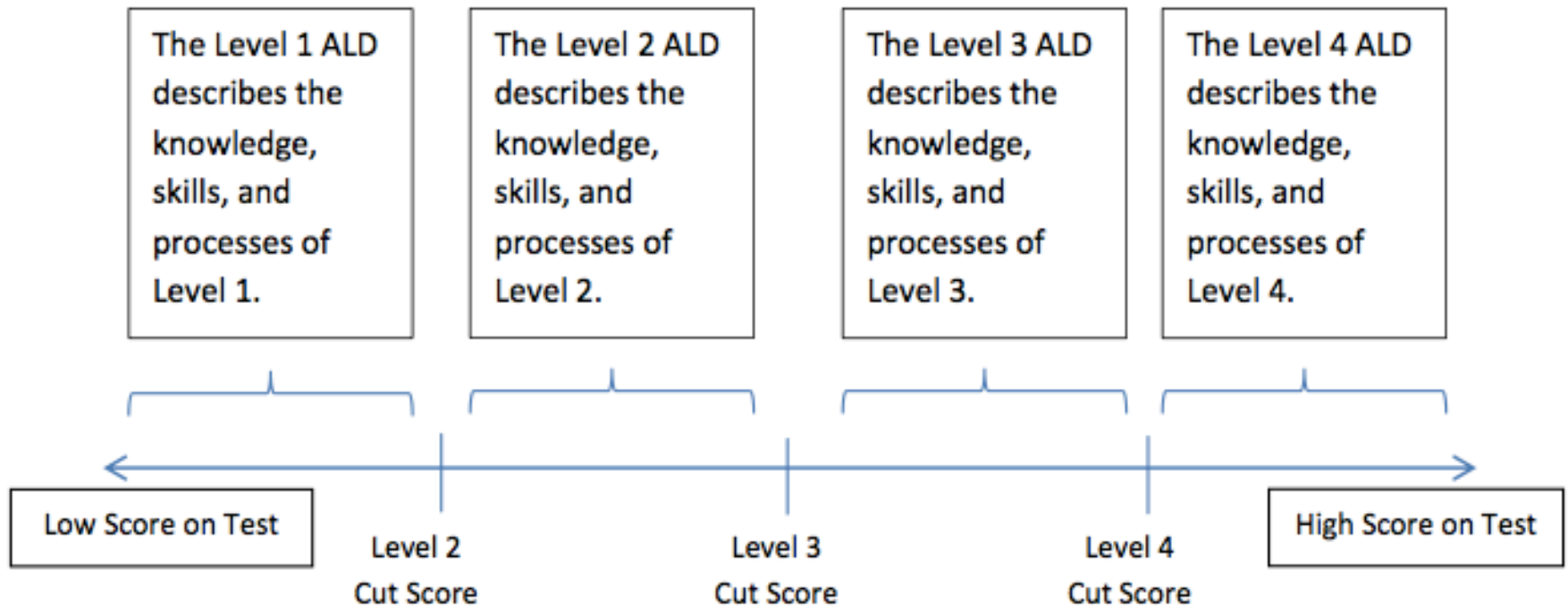
The Level 3 student demonstrates **adequate understanding of and ability to apply** the English language arts and literacy (mathematics) knowledge and skills needed for success in college and careers, as specified in the Common Core State Standards.

The Level 2 student demonstrates **partial understanding of and ability to apply** the English language arts and literacy (mathematics) knowledge and skills needed for success in college and careers, as specified in the Common Core State Standards.

The Level 1 student demonstrates **minimal understanding of and ability to apply** the English language arts and literacy (mathematics) knowledge and skills needed for success in college and careers, as specified in the Common Core State Standards.



Levels & Test Cutscores





Level 3 Policy Claims

3

Student demonstrates adequate understanding of and ability to apply the knowledge and skills associated with college content-readiness.

Student is conditionally exempt from developmental course work, *contingent on evidence of sufficient continued learning in Grade 12.*

Within each state, higher education and K-12 officials may jointly determine appropriate evidence of sufficient continued learning (such as courses completed, test scores, grades or portfolios).

Students are encouraged to take additional 4th year courses as well as appropriate advanced credit courses leading to college credit while in high school.

For students who demonstrate evidence of sufficient continued learning in Grade 12, colleges may evaluate additional data (courses completed, grades, portfolios, placement test scores, etc.) to determine appropriate course placement at or above the initial credit-bearing level.

For students who fail to demonstrate evidence of sufficient continued learning in Grade 12, colleges also may evaluate the same types of additional data to determine placement in developmental or credit-bearing courses.



CRESST REPORT 823

On the Road to Assessing Deeper Learning:
The Status of Smarter Balanced and PARCC
Assessment Consortia

January, 2013

Joan Herman & Robert Linn
CRESST/University of California, Los Angeles



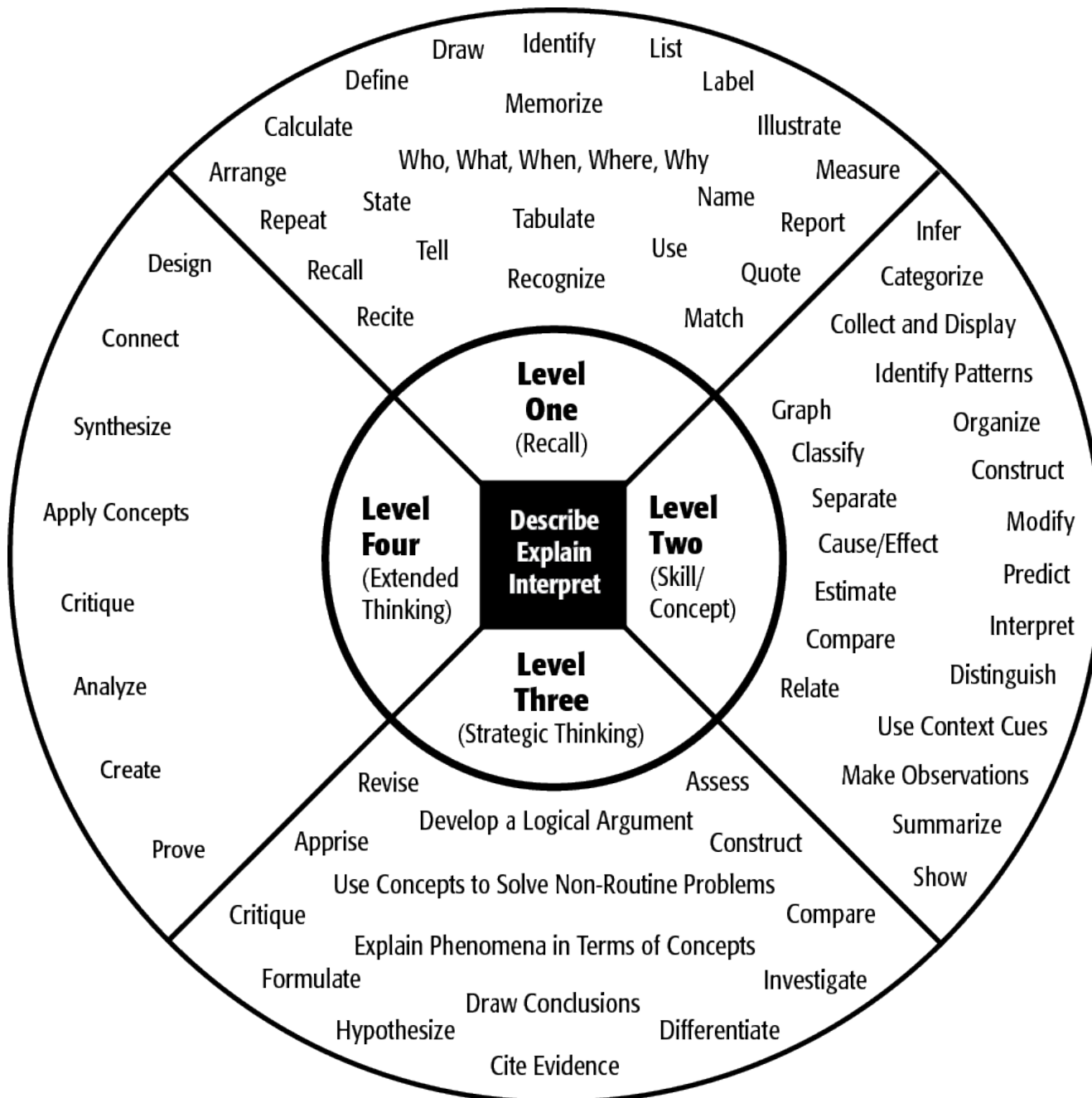
National Center for Research
on Evaluation, Standards, & Student Testing

UCLA • Graduate School of Education & Information Studies



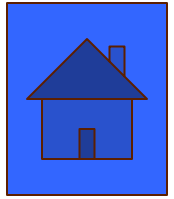
Herman & Linn Questions

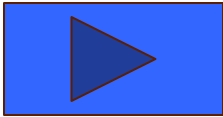
- To what extent are the SBAC and PARCC assessments tapping aspects of Deeper Learning?
- Frame the analysis in terms of Webb's *Depth of Knowledge* Framework
- How do the results compare with similar analyses of typical achievement test items?
- What are the implications regarding the need (and support) for classroom instruction targeted at deeper learning?

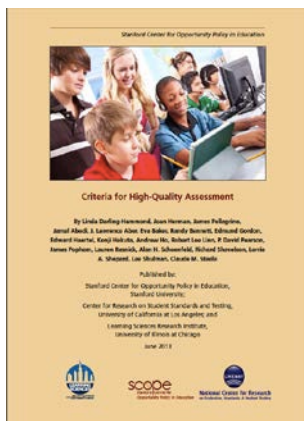




Herman & Linn Results



- Both the SBAC and PARCC assessment blueprints and task designs include multiple examples at DOK Levels 3 and 4.
- In SBAC there is a task distribution across all 4 levels --- depends on the claim. 
- Typical state achievement tests have the preponderance of items at lower DOK Levels
- Instruction and classroom assessment will need to shift to emphasize Deeper Learning, e.g., the close reading of complex texts



Criterion Report: Basic Premises

- No single assessment can evaluate all of the kinds of learning we value for students or meet all of the goals held by parents, practitioners, and policymakers. In a coordinated system of assessments, different tools should be used for different purposes: formative and summative, diagnostic, and large-scale reporting. However, all assessments should faithfully represent the Standards and model good teaching and learning practice. We urge that systems be evaluated by five explicit criteria.



States & Districts Require Different Assessments Aligned to Purposes

- ❑ 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, 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.



Learning Sciences Research Institute
University of Illinois at Chicago

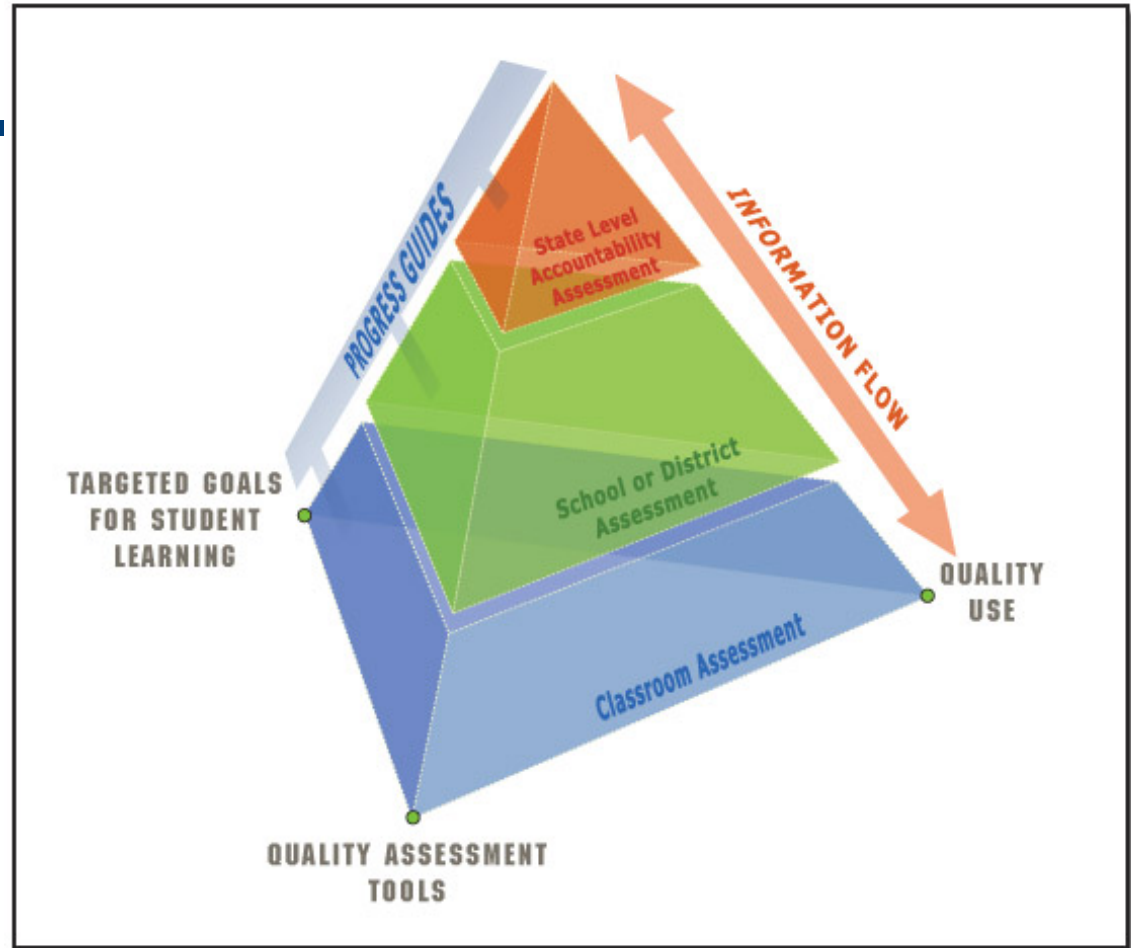
An Integrated System

Coordinated across levels

Unified by common learning goals

Synchronized by unifying progress variables

CAESL ASSESSMENT SYSTEM



Multilevel Assessment System



The Key Design Elements of a Comprehensive System

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

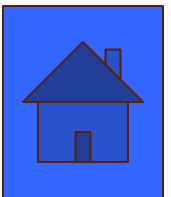
- ❑ The system uses tasks, tools, and technologies appropriate to the desired inferences about student achievement
 - Doesn't force everything into a fixed testing/task model
 - Uses a range of tasks: performances, portfolios, projects, fixed- and open-response tasks as needed
 - **Example:** Use of the performance tasks as the basis for classroom formative assessment – focus on deep engagement with multiple forms of challenging text



Learning Sciences Research Institute
University of Illinois at Chicago

State & District Considerations Going Forward

- What assessment system elements are needed?
- Federal, State and Local mandates
 - What's fixed and what's variable?
- What assessment system elements are currently in place?
- Why are we doing what we are doing?
 - What are the policy and practice drivers?
- How do we move forward and plan for 2014-15 and beyond?





What's Left to Do? – A LOT!!!

Learning Sciences Research Institute
University of Illinois at Chicago

- We need to translate the standards into effective models, methods and materials for curriculum, instruction, and assessment.
 - Need to unpack & clarify performance expectations
 - Need precise claims & evidence statements
 - Need task models & templates
- We need to use what we know already to **evaluate** and **improve** the assessments that are part of current practice, e.g., classroom assessments, large-scale exams, etc.



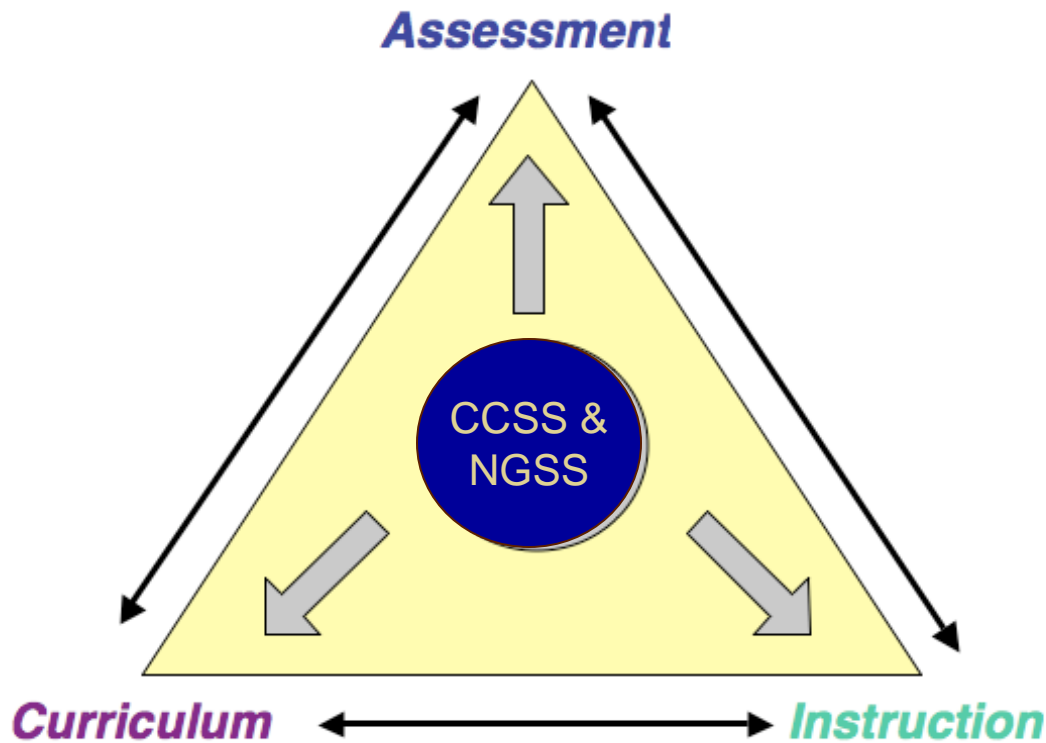
Will We Have Assessments Worth Teaching With & To?

- Desires and timelines of the policy community may conflict with the capacities of the R&D & Practice communities
 - e.g., Worst thing we could do is leap to designing a new “NGSS Aligned” High Stakes Test
- ***Standards are the beginning not the end*** – not a substitute for the thinking and research needed to define ***progressions of learning*** that can serve as a basis for the integration of ***curriculum, instruction and assessment.***



Learning Sciences Research Institute
University of Illinois at Chicago

Assessment Should not be the “*Tail Wagging the Education Dog*”





Voltaire

“The perfect
is the enemy
of the good.”

