COLORADO PERFORMANCE TASK DEVELOPMENT WORKSHOP

Presented by:

Jessica M. Loughran

Center for Educational Testing & Evaluation

University of Kansas

Training Goals

At the end of the training, you should be able to:

1. Understand performance-based assessment:
   - Elements
   - Development process

2. Understand the basic principles of rubric development

3. Create a performance-based assessment for your content area/grade band using a template

4. Prepare materials for assessment trials

Goals of Assessments

- Make a valid connection between a student's performance in an assessment to what the student knows or can do.
- To infer from students' performance about their knowledge/skills

Performance

Knowledge

Skills

Competencies

Goals of Assessments

- **Validity**:
  - A test is valid if the inferences we make about the student based on their test scores are appropriate
  - Validity is the most important concept in an assessment

- **Reliability**
  - Consistency
  - A test is reliable to the extent that it produces stable, consistent results for each examinee
  - Reliability is important because it can limit the validity of an assessment

Goals During Assessment Development

- Align the assessment to the target standard and grade level
- Assure that the assessment is written at the appropriate cognitive level and difficulty level
- Assure that the claims we make about students are **fair** and **reliable**
- Build evidence of **validity** into the assessment development process
- Allow for **accessibility** through universal design and accommodations

What are Performance-Based Assessments?

- What do you envision?
  - Take a few minutes to think about this
 - Write down your general impressions about performance-based assessments

Performance-Based Assessment

- What do you envision?

- We tend to initially think of the "operational" aspects of an assessment.

- In developing performance-based assessments, we need to think more deeply about what we wish to measure, how to collect evidence of the students' knowledge/skill levels, etc.

Performance-Based Assessment at New York's School of the Future
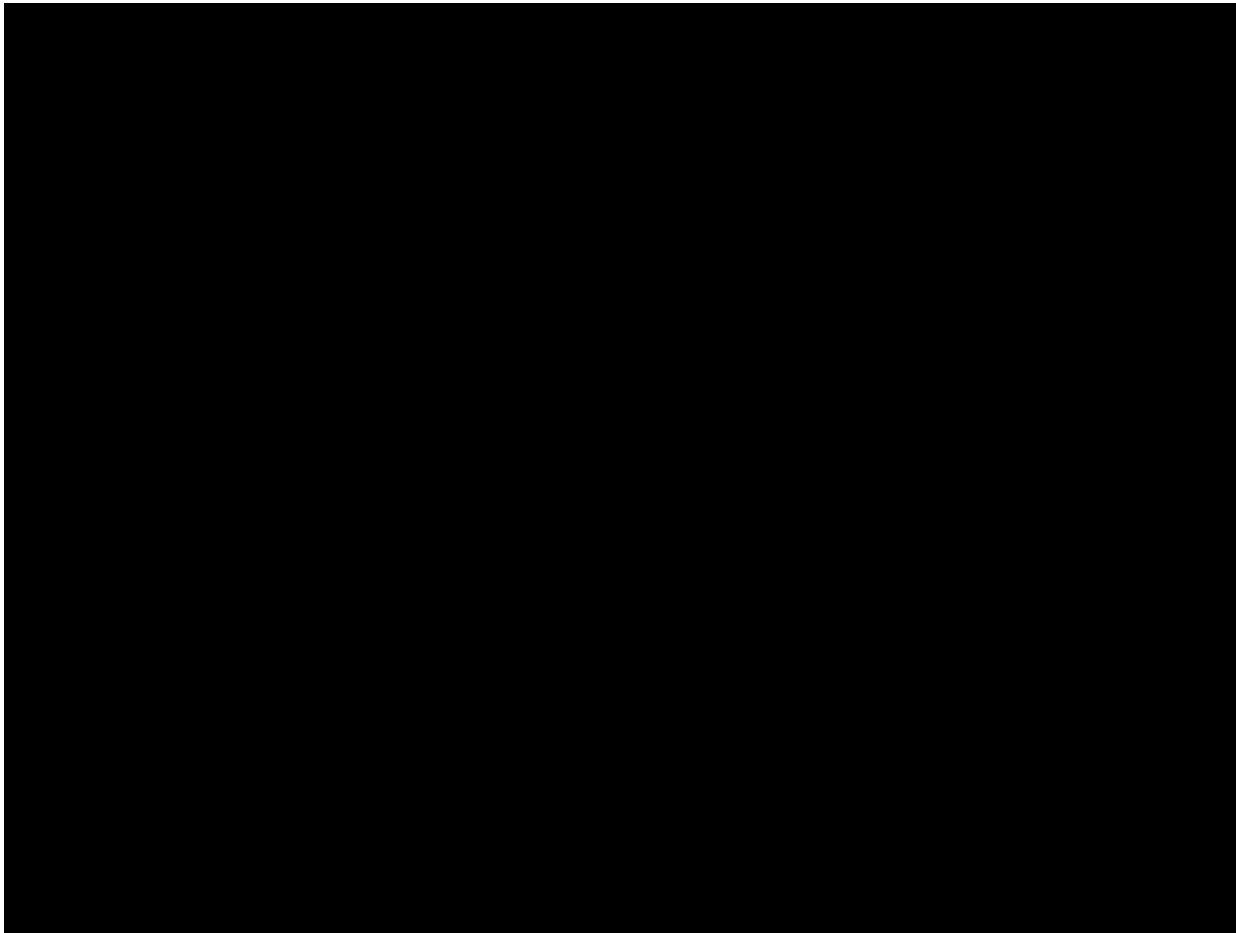
Performance-Based Assessments

- What are performance-based assessments?
  - "Authentic Assessments"
  - The student constructs/supplies an answer, creates some product, or performs an activity (Stecher, 2010)
  - Performance-based assessments are collections of <u>performance tasks</u>.

Performance-Based Assessments

- **What are performance tasks?**
  - What the student does during the assessment
  - A performance-based assessment usually consists of several performance tasks.

- **Performance tasks consist of:**
  - Stimulus material
  - Instructions/request for information
  - Student's responses
  - Methods of recording student's responses
  - Methods of scoring student's responses

Performance-Based Assessments

- **Advantages**:
  - More "<u>authentic</u>" to the teaching/learning environment
  - Allow students to demonstrate knowledge/skills that are, otherwise, very difficult to measure on traditional tests
  - Higher <u>Depth of Knowledge (DOK)</u>
  - Higher <u>face validity</u>, thus more credible
  - Provides the opportunity for <u>detailed feedback</u> about performance
  - Complementary to "traditional" assessments

Performance-Based Assessments

- **Challenges** (consider all of these during task development):
  - Cost
  - Time to develop
  - Feasibility
  - Providing accommodations
  - Reliability of scoring

Major Assessments with Performance-Based Components :

- Advanced Placement
- Career Pathways Assessments (cPass)
- Future Farmers of America (FFA)
  - Career Development Events
- National Assessment of Educational Progress (NAEP)
- National Occupational Competency Testing Institute (NOCTI)
  - "Job Ready" assessments
- National Vocational Qualifications (NVQs)—United Kingdom

Performance-Based Assessments in Colorado

- For what purposes will performance-based assessments be used in the Colorado <u>K-12 schools and districts</u>?
- For what purposes will performance-based assessments be used in Colorado <u>Career & Technical Education</u>?
- How will outcomes of performance-based assessments affect students and educators?

Development Process: Traditional Approach

- Align the content of the item to the content standard
- Example from the first grade Mathematics standards
  - Standard: 1. Number sense, Properties, and Operations
  - Use the equal sign to demonstrate equality in number relationships.

Frank wrote the following equation:

7 = 8-1

Explain why this is true.

Example adapted from: Smarter Balanced (2012)

Evidence-Centered Design (ECD):
A structured development process

- ECD describes:
  - the elements of <u>assessment design</u>
  - the elements of the <u>operational assessment</u>
  - <u>test items/performance tasks</u> that are carefully designed so we can collect evidence of the knowledge/skills
  - <u>scoring procedures</u> designed so that we are recording observable variables that relate to the knowledge/skills
  - how these elements influence one another

Assessment Development Process: The "Layers of a House" Metaphor

Evidence-Centered Design (ECD): Vocabulary

- **Claims**
  - Inferences that we make about a student's knowledge, skills, abilities, or other characteristics (KSAOs)
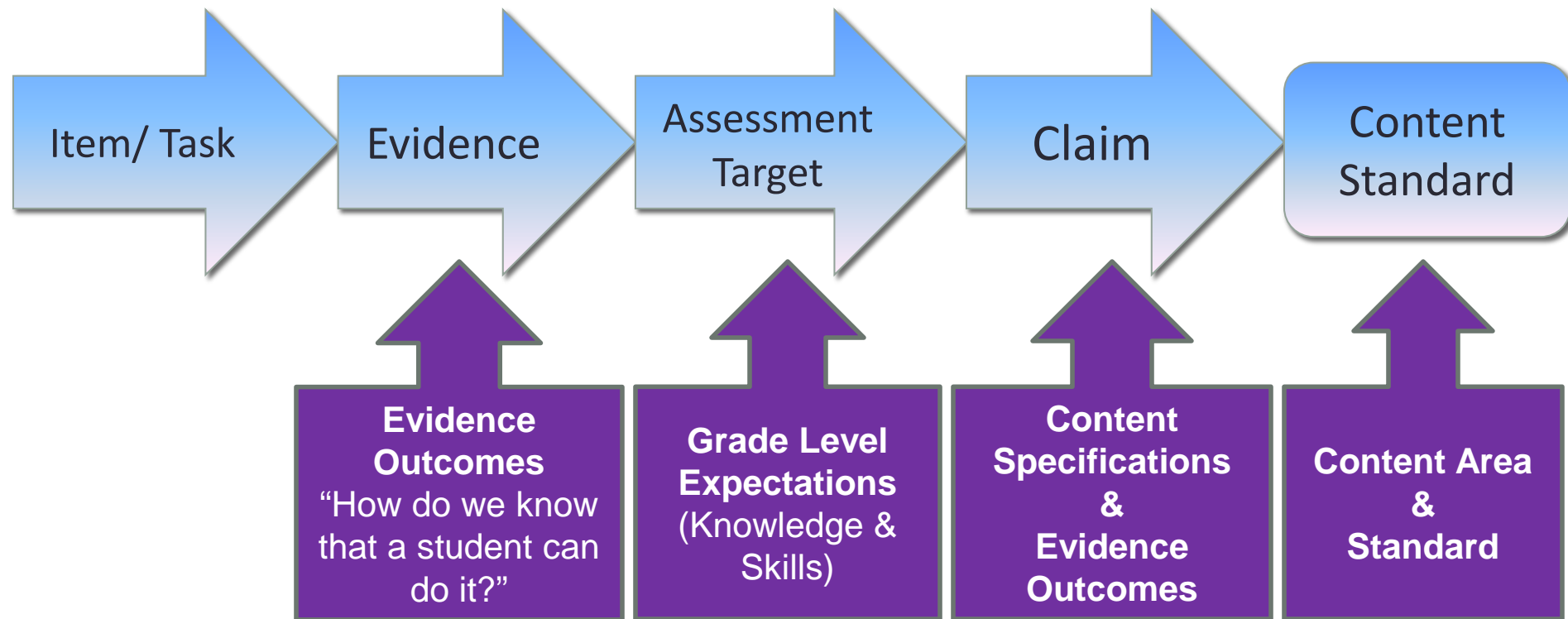
- **Evidence**
  - "Information/observations that allow claims to be made about an examinee's proficiency from evaluations of observable behaviors in given performance situations" (Almond, Steinberg, & Mislevy, 2002, p. 59)
  - Note: Proficiency is not observable! Evidence of the proficiency can be observable.

Evidence-Centered Design

Item/ Task → Evidence → Assessment Target → Claim → Content Standard

Image Source: Smarter Balance Assessment Consortium (2012)

Evidence-Centered Design & CDE Academic Standards

Item/ Task	→	Evidence	→	Assessment Target	→	Claim	→	Content Standard
		Evidence Outcomes "How do we know that a student can do it?"		**Grade Level Expectations** (Knowledge & Skills)		**Content Specifications & Evidence Outcomes**		**Content Area & Standard**

Source: Smarter Balance Assessment Consortium (2012)

Evidence-Centered Design (ECD):

- ECD provides a way to:
  - conceptualize all of the elements of assessment design
    - can be used to design new assessments and evaluate current ones
  - develop more standardized performance-based assessments
  - develop assessments that involve complex competencies and complex settings
  - facilitate the job of developing parallel tasks

Evidence-Centered Design

Let's walk through an example using a Colorado standard…

Evidence-Centered Design

Content
Standard

Source: Smarter Balance Assessment Consortium, Introduction to Evidence-Centered Design

Content Area

- What content should be assessed?

- Decide which Colorado Academic Standards and Grade Level Expectations should be assessed.
  - You will align the components of the assessment to this information.
  - For example, make sure the performance tasks you create align with the Grade Level Expectations DOK

Content Area: Reading, Writing, and Communicating

Standard: 1. Oral Expression and Listening

Prepared Graduates:
➤ Use language appropriate for purpose and audience.

Grade Level Expectations: Eleventh Grade

Concepts and skills students master:
 1. Verbal and nonverbal cues impact the intent of communication

Evidence Outcomes	21st Century Skills and Readiness Competencies

Content Area

- Depth of Knowledge (DOK)
  - Given the Grade Level Expectations, what DOK level should the assessment target?
    - Will the student define/list information? (**LEVEL 1**)
    - Will the student compare/categorize? (**LEVEL 2**)
    - Will the student critique/draw conclusions? (**LEVEL 3**)
    - Will the student analyze/synthesize/design? (**LEVEL 4**)

  - One advantage of performance tasks is that they can assess higher levels of DOK

Evidence-Centered Design
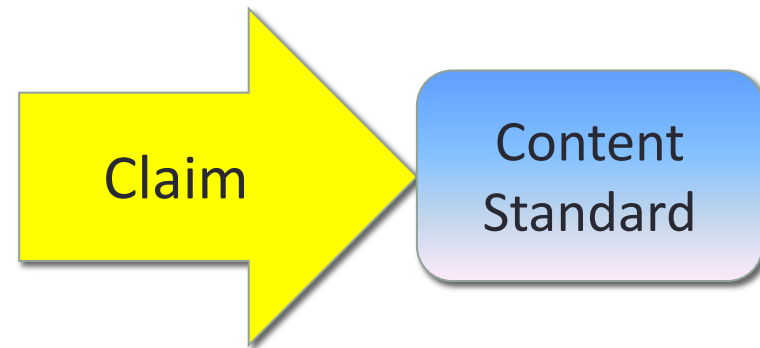
Claim → Content Standard

Source: Smarter Balance Assessment Consortium, Introduction to Evidence-Centered Design

Claims

- Claim: Inferences we make about what a student knows (knowledge) or can do (skills)

- Define the claims that will be made

Content Area: Reading, Writing, and Communicating

Standard: 1. Oral Expression and Listening

Prepared Graduates:
➢ Use language appropriate for purpose and audience.

Grade Level Expectations: Eleventh Grade

Concepts and skills students master:
  1. Verbal and nonverbal cues impact the intent of communication

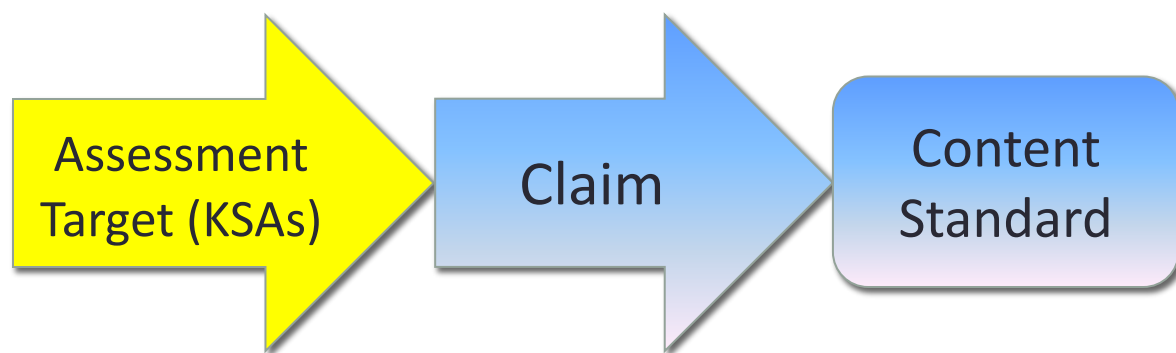| Evidence Outcomes	21st Century Skills and Readiness Competencies
Students can: b. Deliver formal oral presentations for intended purpose and audience, using effective verbal and nonverbal communication.	Inquiry Questions: Relevance and Application: Nature of Reading, Writing, and Communicating:

Evidence-Centered Design

Source: Smarter Balance Assessment Consortium, Introduction to Evidence-Centered Design

Knowledge, Skills, & Other Characteristics

- Based on the content and standards, specify what the assessment is supposed to measure
  - knowledge
  - skills

- Note: These variables are not observable.
  - We infer knowledge/skills from the students' behaviors.

Knowledge, Skills, & Other Characteristics

- Proficiency
  - What does proficiency look like?
  - What can students say/do to show evidence of their proficiencies?
  - What are the focal knowledge/skills that the assessment should address?

| Content Area: Reading, Writing, and Communicating	
Standard: 1. Oral Expression and Listening	
Prepared Graduates: ➤ Use language appropriate for purpose and audience.	
Grade Level Expectations: Eleventh Grade	
Concepts and skills students master: 1. Verbal and nonverbal cues impact the intent of communication	
Evidence Outcomes	**21st Century Skills and Readiness Competencies**
Students can: b. Deliver formal oral presentations for intended purpose and audience, using effective verbal and nonverbal communication.	Inquiry Questions: Relevance and Application: Nature of Reading, Writing, and Communicating:

Skills

Examples:

- Student can effectively demonstrate the appropriate level of <u>verbal</u> communication in a formal speech.

- Student can effectively demonstrate the appropriate level of <u>nonverbal</u> communication in a formal speech.

- Student can successfully deliver a presentation tailored to a particular <u>audience/purpose</u>.

Knowledge

Examples

- Student has knowledge of what constitutes formal versus informal <u>verbal</u> communication.

- Student has knowledge of what constitutes formal versus informal <u>nonverbal</u> communication.

- Student understands the need to tailor presentations <u>for the intended audience/purpose</u>.

Evidence-Centered Design
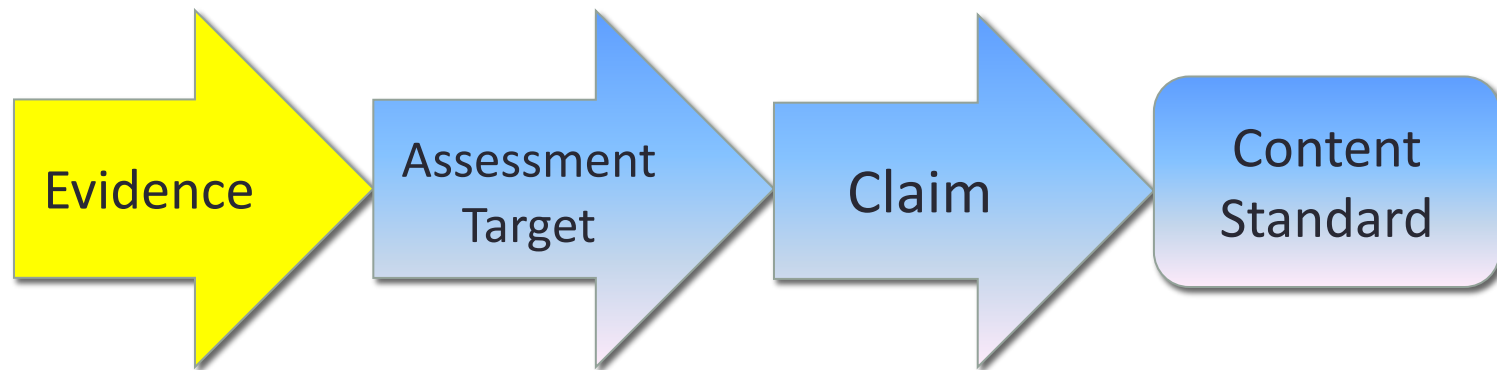
Source: Smarter Balance Assessment Consortium, Introduction to Evidence-Centered Design

Evidence

- What can the student <u>say/do</u> to show evidence of their knowledge and skills?

- What <u>information</u> does the student need to give us through his/her performance?

- What are some of the <u>observable behaviors</u> we are interested in?

Evidence

- How do we collect evidence to support our claims about a student?
  - Performance tasks and all their components
  - Students' responses
  - Rubrics
  - Statistical analyses
  - Score reports
  - Other information

How would we collect evidence of these claims?

Content Area: Reading, Writing, and Communicating	
Standard: 1. Oral Expression and Listening	
Prepared Graduates: ➢ Use language appropriate for purpose and audience.	
Grade Level Expectations: Eleventh Grade	
Concepts and skills students master: 1. Verbal and nonverbal cues impact the intent of communication	
Evidence Outcomes	**21st Century Skills and Readiness Competencies**
Students can: b. Deliver formal oral presentations for intended purpose and audience, using effective verbal and nonverbal communication.	Inquiry Questions: Relevance and Application: Nature of Reading, Writing, and Communicating:

Evidence

- **Performance tasks**
  - Student delivers a formal presentation
- **Students' responses**
  - Watch the presentation and collect evidence about the students' knowledge/skills with regard to verbal communication, nonverbal communication, etc.
- **Rubrics**
  - Scoring criteria in the rubrics would describe specific behaviors that indicate mastery, non-mastery

Collecting Evidence

	Excellent	Good	Fair	Poor	
	3	2	1	0	Weight
Eye Contact	Maintains eye contact across the entire audience for the duration of the presentation	Maintains eye contact across the audience, but looks away (e.g., at floor, at notes, etc.) from time to time	Displays some eye contact with the audience but does not look across the entire audience or frequently looks away	Displays very little or no eye contact with audience; reads from notes (or looks at floor, etc.) throughout the duration of the presentation	X3

Evidence

- We are accustomed to thinking about test 'questions' or test 'items'
- However, performance tasks are more complex
  - Keep this in mind as you develop your performance tasks and scoring criteria!

Evidence

- What kind of evidence should we collect?
- Example: Career & Technical Education welding performance task

Evidence

- **Scoring**
  - What are we evaluating at the task level?
  - What observable variables are we looking for?
  - Collecting evidence/scoring should:
    - Be objective
    - Have reliability among raters
  - As such, scoring rubrics should:
    - Contain clear language (for raters and students)
    - Contain clear scoring criteria
    - Contain clearly defined score categories

Evidence-Centered Design

Performance Task → Evidence → Assessment Target → Claim → Content Standard

Source: Smarter Balance Assessment Consortium, Introduction to Evidence-Centered Design

Tasks

- How will we "capture" the students' performance?
- What are some real-world situations we can specify for the student?
- What tasks allow us to elicit information from the students that relates to students' knowledge, skills, etc.?
  - **Examples:**
    - **an essay**
    - **a speech**
    - **a portfolio**
    - **an experiment**

Tasks

- Questions to consider:
  - How do you describe the tasks?
    - Task context/setting
    - Task difficulty
    - DOK
    - Content of the tasks
  - What will be presented to the student (presentation material)?
    - Instructions
    - Stimulus material, prompt
    - Resources/tools
  - What is the final product that the student will produce?
    - What will the student produce?
    - What will the student do/say?

Tasks

- **Substance**
  - Think about the tasks…what kind of performance will they elicit from the student?
  - What evidence of knowledge/skills will the tasks elicit?
  - What DOK Level will the tasks assess? Does this match the Grade Level Expectations?

Operational Assessment

- Just like every other "layer" of the assessment framework, careful assembly of the operational assessment helps us accumulate **evidence** about the **claims** we make about students.

- In the process of assembling the assessment, we are always thinking back to those student factors that we want to measure:
  - Knowledge
  - Skills
  - Other characteristics

Operational Assessment

- Think about…
  - What the operational assessment looks like
  - How the assessment will look to the students, to the test administrators, evaluators, etc.?
  - How will test administrators interact with the students?
  - What work product will the students submit to the test administrator?

Operational Assessment—Tasks

- Questions to ask:
  - What tasks will the students engage in?
  - What are the task specifications?
  - What will the student do? What will the student produce?

- How do we put the tasks together (i.e., how tasks will relate to on another or depend on one another)?
  - Welding performance task example

Operational Assessment—Universal Design

An example: The Evolution of the Door Knob

Source: www. Austinpremierlocks.com

Operational Assessment—Universal Design

- **Fair and Unbiased**: Is the content fair given the Grade Level Expectations?
  - Linguistic Complexity
    - Number of words
    - Length of words/sentences
    - Ambiguous language
  - Readability
    - Font size
    - White space
    - Charts/graphs/pictures
- **Sensitivity**: Avoid construct irrelevant topics that may be offensive or distracting

Operational Assessment—Universal Design

- Language—things to avoid:
  - Word with multiple meanings:
    - A doctor's practice
    - Example from a (very) bad joke: "A dog gave birth to puppies near the road and was cited for littering."
  - Confusing/distracting language:
    - "Up to a 50% discount or more"
    - "They do what they do due to Congress."

"Our language is funny. A 'fat chance' and a 'slim chance' are the same thing." –J. Gustav White

Operational Assessment—Universal Design

- Do not limit receptive or expressive access unless there is a compelling reason
  - Avoid time limits
  - Do not let motor ability or visual acuity interfere with accuracy of responses

- Can you think of any other ways to enhance access?

Operational Assessment—Accommodations

- When universal design has limitations, identify any accommodations that may be necessary for the assessment

- Several examples (Source: CDE Content Review Tool Training):
  - Linguistic accommodations (for English language learners)
  - Task Presentation Accommodation
  - Task Response Accommodation
  - Setting Accommodation
  - Timing/Scheduling Accommodation
  - Many others…

Operational Assessment—Interactions

- How will the students interact with the test administrator(s)?

- What kind of 'work product' will students present to the test administrator(s)?

Operational Assessment—Gathering Evidence

- What part of the student's performance are we evaluating?
  - Final product?
  - Process?
  - Both final product and process?

- Here, we are gathering evidence of the students' knowledge/skills.
  - Use rubrics to record this evidence.

Operational Assessment—Overall

- Other overall considerations that affect the assessment as a whole:
  - Overall context of the assessment (e.g., computer-based, student completes task at a work station with tools, etc.)
  - What tools are students given before and during the assessment?
  - At what point do students:
    - Begin/end the assessment
    - Receive instructions
  - Administrative constraints (e.g., cost)
  - Testing window
  - Secure administration of the assessment
  - How to collect demographic information on the students?
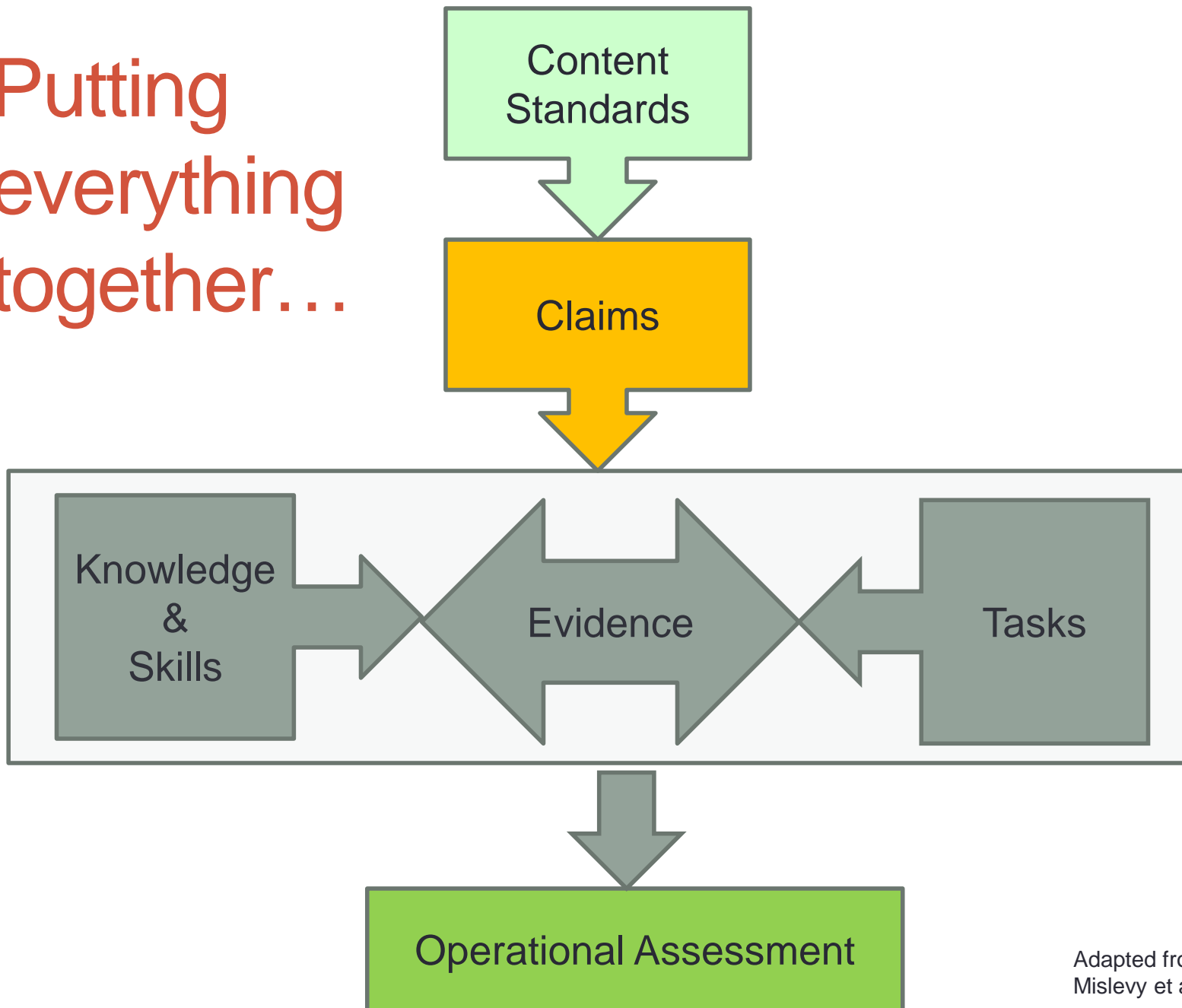
Where to start?!

There are so many elements to think about…

- What knowledge do we wish to measure?

- What skill set do we wish to measure?

- What do we want students to learn how to do?

- For what real-world situations are we interested in preparing students?

- How will we evaluate students' performance?

Use the Performance Assessment Development Template to Create Your Assessment

See Guidelines for the Trials

References

Almond, R.G., Steinberg, L.S., & Mislevy, R.J. (2002). Enhancing the design and delivery of assessment systems: A four process architecture. The Journal of Technology, Learning, and Assessment, 1(5). Retrieved January 28, 2013, from http://ejournals.bc.edu/ojs/index.php/jtla/article/view/1671/1509

"Authentic" assessment at school of the future [Web]. (2012). Retrieved from https://www.youtube.com/watch?v=9l5m66Y607U

Behrens, J. T., Mislevy, R. J., Bauer, M., Williamson, D. M., & Levy, R. (2004). Introduction to evidence centered design and lessons learned from its application in a global e-learning program. *International Journal of Testing*, *4*(4), 295-301.

Colorado Department of Education. (2012, February 01). *Colorado content collaboratives (cccs)*. Retrieved from www.cde.state.co.us/ContentCollaboratives

Linn, R. L., Baker, E. L. & Dunbar, S. B. (1991). Complex, performance-based assessment: expectations and validation criteria. *Educational Researcher*, *20*(8), 15-21.

Messick, S. (1994). The interplay of evidence and consequences in the validation of performance assessments. *Educational Researcher*, *23*(2), 13-23.

Miller, L.A., Mcintire, S.A., & Lovler, R.L. (2011). Foundations of Psychological Testing (3rd ed.), Los Angeles: Sage.

Mislevy, R. (2002). Making Sense of Data From Complex Assessments. *Applied Measurement In Education*, *15*(4), 363-389.

Mislevy, R. J., & Haertel, G. D. (2006). Implications of Evidence-Centered Design for Educational Testing. *Educational Measurement: Issues and Practice*, *25*(4), 6-20.

Mislevy, R.J., & Riconscente, M.M. (July 2005). Evidence-centered assessment design: Layers, structures, and terminology. Menlo Park, CA: SRI International.

Mislevy, R. J., Steinberg, L. S., & Almond, R. G. (2002). Design and analysis in task-based language assessment. *Language testing*, *19*(4), 477-496.

Mislevy, R. J., Steinberg, L. S., & Almond, R. G. (2003). On the structure of educational assessments. Measurement: Interdisciplinary Research and Perspectives, 1, 3–62.

Smarter Balanced Assessment Consortium, (2012). Introduction to Smarter Balanced item and performance task development [PowerPoint slides]. Retrieved from www.smarterbalanced.org

Stecher, B. (2010). Performance Assessment in an era of standards-based educational accountability. Stanford Center for Opportunity Policy in Education, RAND Corporation.