

Sandy Garrett  
State Superintendent of  
Public Instruction  
Oklahoma State Department of Education

---

## BUILDING A TRANSPARENT ASSESSMENT SYSTEM FOR THE OKLAHOMA CORE CURRICULUM TESTS (OCCT)

Joyce DeFehr  
Executive Director, State Testing  
Office of Accountability and Assessments

# Why Should You Know About Test Development

- Understanding test development may help answer questions from students, parents, teachers, and policy makers.
- The information may help guide development of local benchmark assessments.
- It will also help to design tests that will answer the appropriate questions.

# Test, Test, Test

□ Every assessment should have a purpose.

Formative  
Assessments

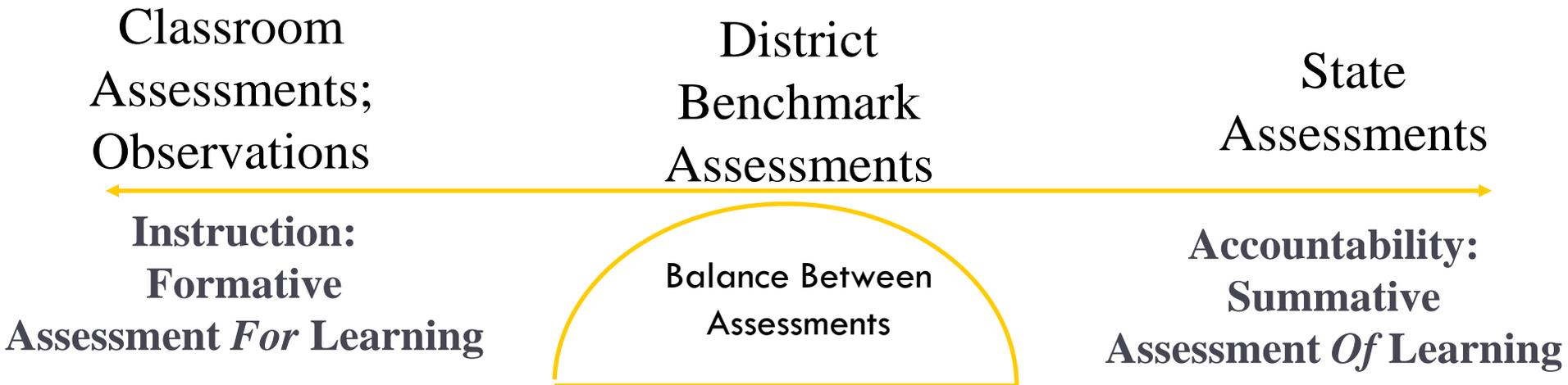
- Classroom assessments
- Teacher observations
- Benchmark assessments (9 week tests)
- Summative assessments (state tests)

# Stakeholders' Questions?

- Student → Am I doing better?
- Teacher → What can I do differently?
- Principal → Are my teachers adjusting appropriately?
- Community → Are our tax dollars being well spent?
- District → Are principals effectively leading?
- Region → Where is teacher training needed most?
- State → Which schools are effective?
- “Feds” → Is “Reading First” working?

# Balanced Assessment System

- **Can a single assessment answer all of these questions?  
Probably not!**
- **A balanced assessment system is one that includes multiple measures selected to meet specific purposes.**



# Overview of State Test Development

The state determines the following:

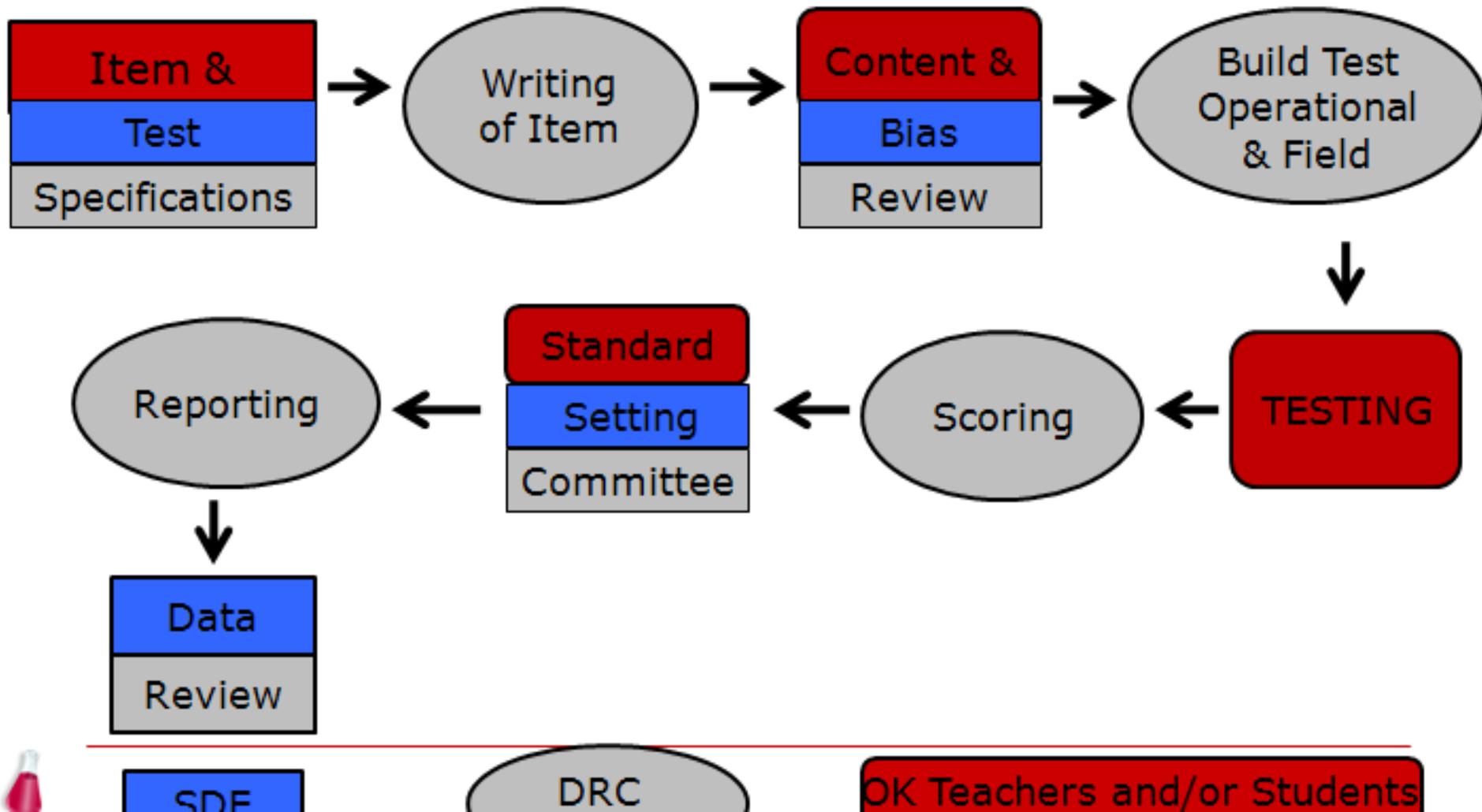
- What content areas and grade levels will be assessed?
- What is the design of the assessment?
- What are the timelines?
- How will the assessment be scored and reported?

# Overview of Test Development

7

- Oklahoma educators are involved.
- Committees composed of classroom teachers, curriculum specialists, administrators, and other educational stakeholders.
- Committees represent the state geographically, ethnically, by gender, and by type and size of school district.

# OCCT TEST DEVELOPMENT CYCLE



# Test Development

- **Development of Standards** – Educational stakeholders develop State mandated curriculum for a particular subject that aligns with good classroom instruction and student learning.
- **Test and Item Specifications** – Educator committees assist in the development of measurement specifications for each objective. These specifications outline the eligible test content and test item formats and include sample items.

# Test Development

- **Blueprints** – With educator input, the length of the test and number of items measuring each objective are determined.
- **Item Development** – Items are developed based on the objectives and measurement specifications. The item development process includes many phases.
- **Building the Test** – The test is built to match the blueprint as well as measurement statistics.

# Item Development Cycle

## Phase I: Item Writing

- Assessment specialists with the contracted testing vendor begin the initial writing of items developed to meet item specifications.
- Internal item reviews are conducted by assessment specialists with the State and the testing vendor to ensure that items meet standards and specifications, are free from bias and error, and are fair and accurate.

# Item Development Cycle

## Phase II: Content and Bias Review

- Oklahoma educators from the specific content area/grade level check that items correctly measure the *PASS* standards and objectives.
- They review items for ethnic, gender, religious, or cultural bias as well as subject and grade appropriateness.
- Assessment specialists edit and revise items according to feedback.

# Item Development

- Bias is not
  - The mere presence of a score difference between groups, or
  - Necessarily present if scores reflect real differences in achievement.

# Item Development

- An item is unbiased
  - When the probability of successfully answering the item is the same for equally able examinees, regardless of their subgroup membership, and
  - When it does not result in an advantage or disadvantage for a particular group of students for reasons unrelated to the knowledge or skill being assessed.

# Item Development

## Reviewing Test Items

- The item measures the standard and the objective assigned.
- Information in the item is true and accurate.
- Difficulty level of the item is appropriate.
- Reading level of the item is appropriate.
- Depth of Knowledge (DOK) is appropriate.

# Item Development

- The item should
  - ▣ Be clearly stated;
  - ▣ Use precise and concise language;
  - ▣ Avoid the use of negatives;
  - ▣ Avoid the use of ill-defined terms or ambiguous words;
  - ▣ Be free from grammatical errors; and
  - ▣ Not provide clues to the correct answer.

# Item Development

- Answer choices should
  - ▣ Be plausible;
  - ▣ Be similar in specificity, complexity, and structure;
  - ▣ Be similar in length, or two long and two short;
  - ▣ Avoid repeating words or phrases that could be in the stem; and
  - ▣ Avoid the use of absolutes.

# Item Development

---

- The correct answer should
  - ▣ Be true and accurate; and
  - ▣ Be clearly correct to those who have the knowledge.

# Item Development

- Distractors should
  - Be attractive to those who do not have the knowledge or skills,
  - Represent students' common misconceptions or mistakes, and
  - Be clearly incorrect.

# Item Development

- DOK is not the same as difficulty.
- It is the complexity vs. the difficulty of a question.
- DOK focuses on how deeply a student has to know the process/content.

# Item Development

---

## Depth of Knowledge

- Recall – Level 1
- Skills and Concepts – Level 2
- Strategic Thinking – Level 3

# Item Development

## Depth of Knowledge 1

- Requires students to demonstrate a rote response, use a well-known formula, follow a simple procedure (recipe), or perform a clearly defined series of steps.
- A “simple” procedure is well defined and typically involves only one step.

*Key Words: identify, recall, recognize, use, and measure*

# Item Development

## Depth of Knowledge 2

- It requires students to make some decisions as to how to approach the question or problem.
- Requires deeper knowledge than just giving a definition such as explaining how or why; it implies more than one step.

*Key Words: compare, contrast, organize, estimate, summarize, classify, infer, and summarize*

# Item Development

## Depth of Knowledge 3

- Requires deep understanding as exhibited through planning, using evidence, and more demanding cognitive reasoning.
- Draw conclusions from multiple or complex stimuli.
- Recognize and explain misconceptions.
- Make connections across time and place to explain a concept.
- Analyze and synthesize information from multiple sources.

# Item Development Cycle

## Phase III: Operational Test Construction and Field Testing

- Assessment specialists select items for the operational test.
- Item selection is balanced to meet alignment, blueprint, and statistical criteria.
- Each year, ten (10) field test questions are placed throughout the Geography test.
- Oklahoma students participate in the spring assessments.

# Item Development Cycle

## Phase IV: Data Analysis of Field Test Items

- Psychometricians from DRC calibrate items and provide statistical data.
- Assessment specialists from DRC and the State review how well the items performed on the field test.

# Item Development Cycle

---

## Phase IV: Data Analysis of Operational Test Items

- Psychometricians analyze items to determine if the items “behave” predictably.
- Items are calibrated with equating procedures to produce scale scores and assessment results.

# Standard Setting

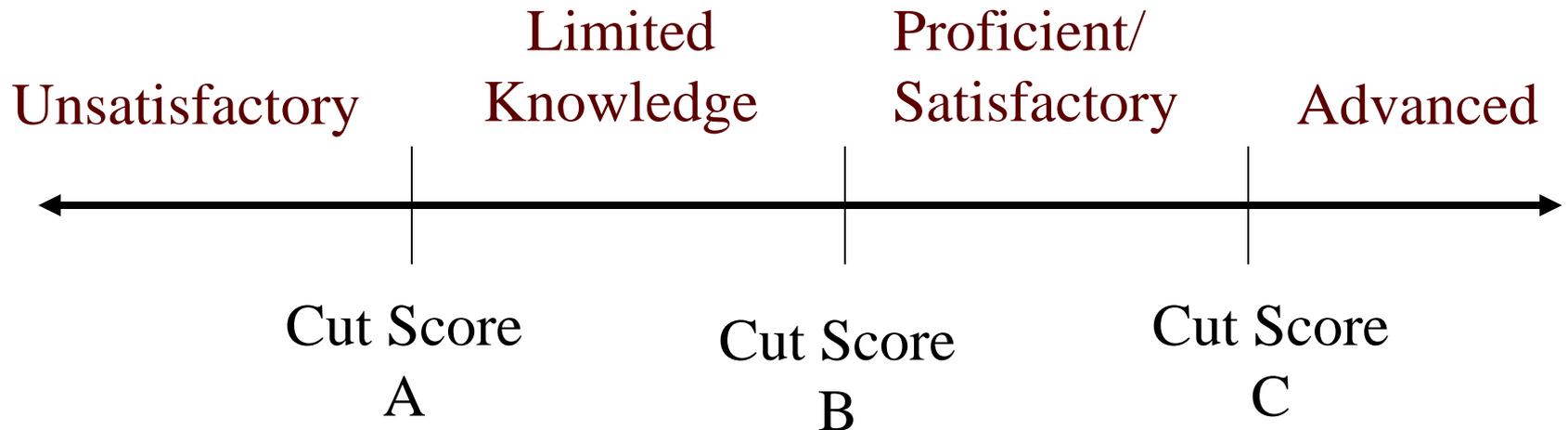
## Phase V: Setting Standards or Cut Scores

- Standard Setting is the process that allows experts to make judgments about the content that a student should know and be able to do in order to be classified in a specific performance level.
- Cut scores are necessary for the categorization of student test scores into the four performance levels utilized in Oklahoma:
  - Advanced,
  - Proficient/Satisfactory,
  - Limited Knowledge, and
  - Unsatisfactory.

# Standard Setting

## What is standard setting?

- It is the minimum test score a student must earn to be considered at a certain performance level.
- Three cut scores result in four levels of performance.



# When Is Standard Setting Necessary?

---

- New Tests
- Curriculum Updates
- Blueprint Changes
- Performance Level Descriptor Changes

# Basic Vocabulary

- **Content Standards:** The content and skills that students are expected to know and be able to do
- **Performance Levels** (Achievement Levels, Performance Categories): Labels for levels of student achievement (e.g., Limited Knowledge, Proficient/Satisfactory, and Advanced)
- **Performance Level Descriptors** (PLDs): Descriptions of the competencies associated with each level of achievement
- **Cut scores** (Standards): Scores on an assessment that separate one level of achievement from another.

# Standard Setting

- The standard setting panelists determine cut score recommendations based on test content, performance standards, and test item difficulty, NOT based on number of test items answered correctly.
- Performance Level Descriptors (PLDs) are developed to describe the amount of knowledge and skills required to achieve at the specific performance level.
- PLDs are linked directly to the *Priority Academic Student Skills (PASS)*.

# Standard Setting

## Selection Criteria for Panel Members

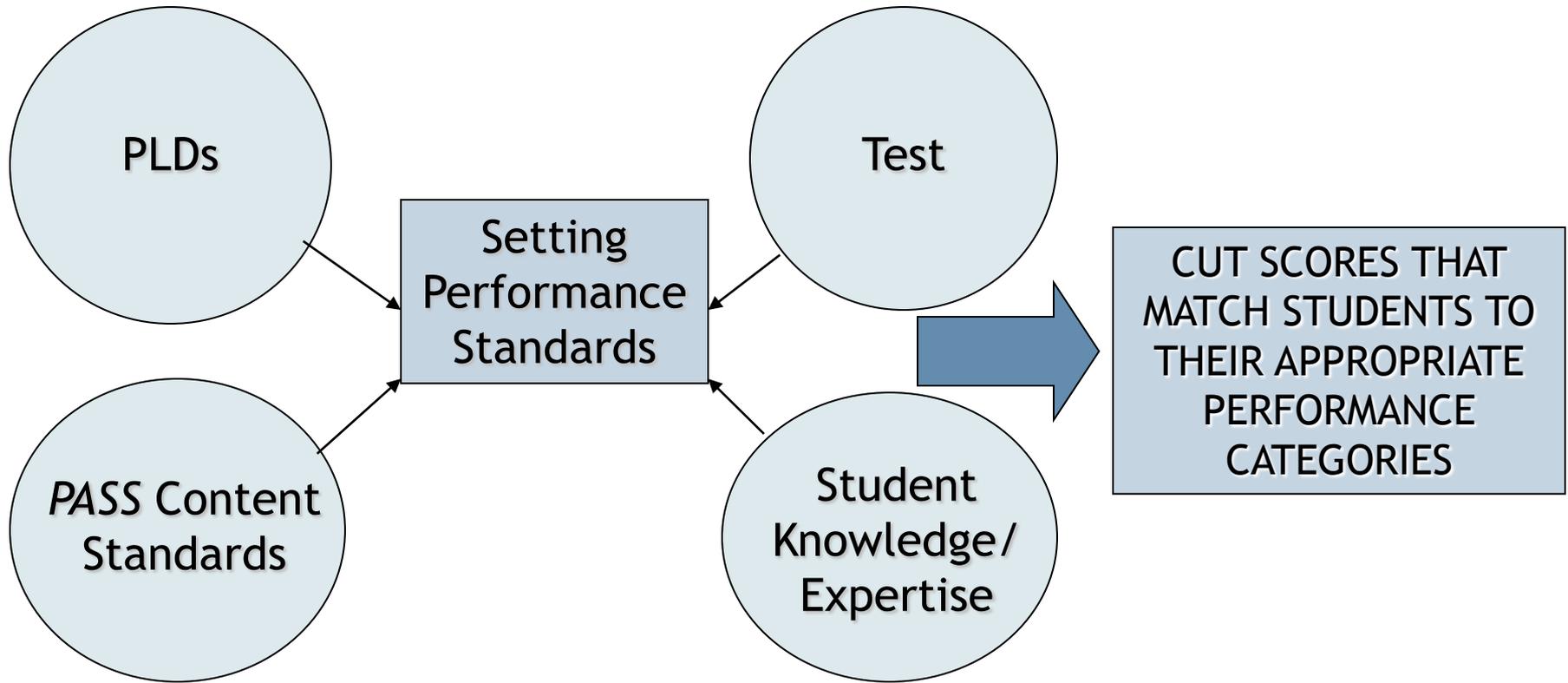
- Educational Stakeholders
  - ✓ Teachers and Curriculum Experts
  - ✓ Higher Education
  - ✓ Business leaders
  - ✓ Legislators or representatives from Educational Agencies
  - ✓ School Board Members
- Knowledgeable in tested content
- Familiar with tested populations at multiple grade levels
- Diverse backgrounds, demographics, and experience
  - ✓ Gender
  - ✓ Years of experience
  - ✓ Region
  - ✓ Ethnicity
  - ✓ Student population

# Standard Setting

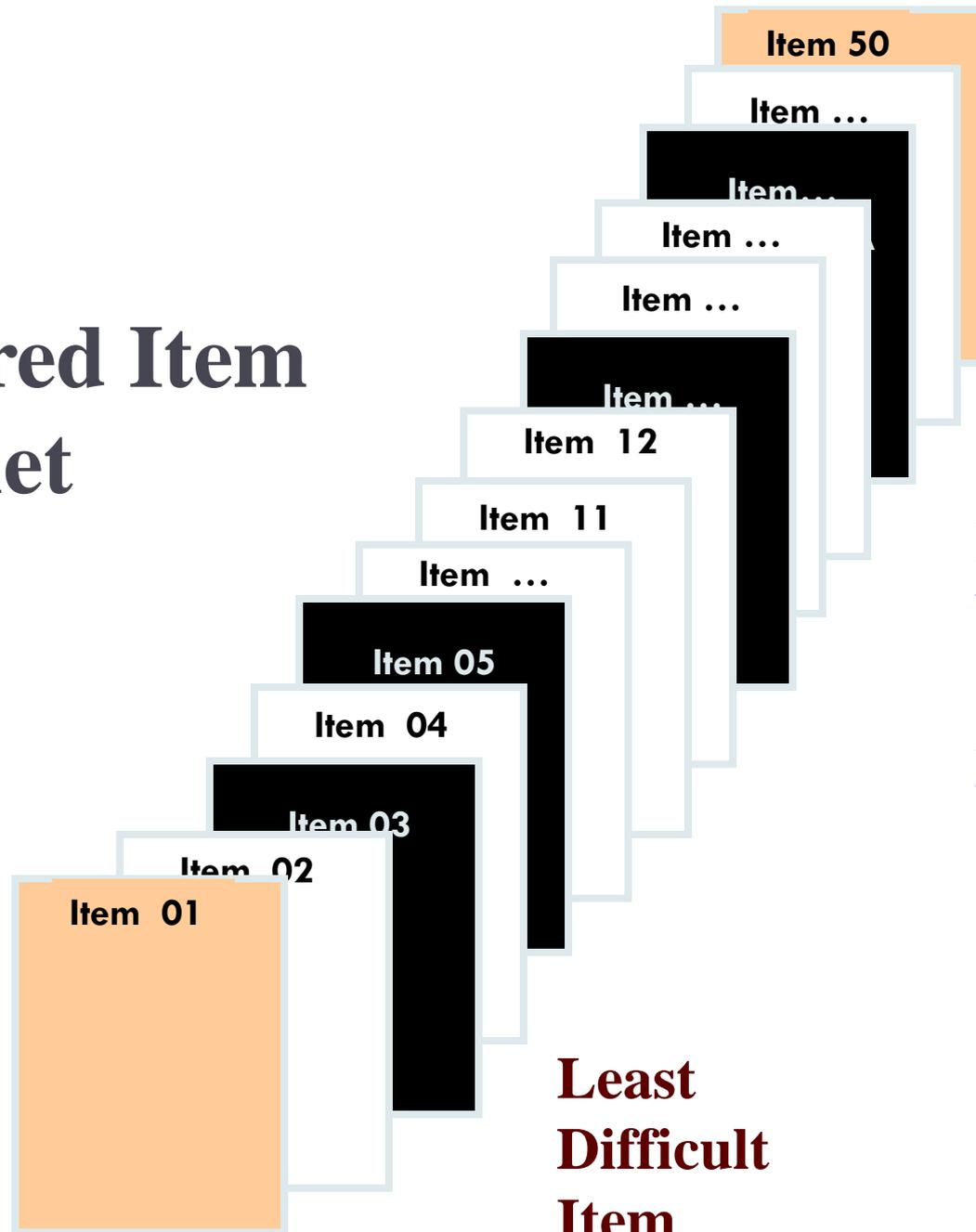
## Role of the Panel

- To understand the purpose of the tests and cut scores
- To be familiar with the PASS content standards and objectives
- To translate the minimum that students should know and be able to do at each level into a cut score
- To recommend cut scores to the Oklahoma State Department of Education (OSDE) that articulate well with the current OCCT assessments

# Setting Performance Standards



# Ordered Item Booklet

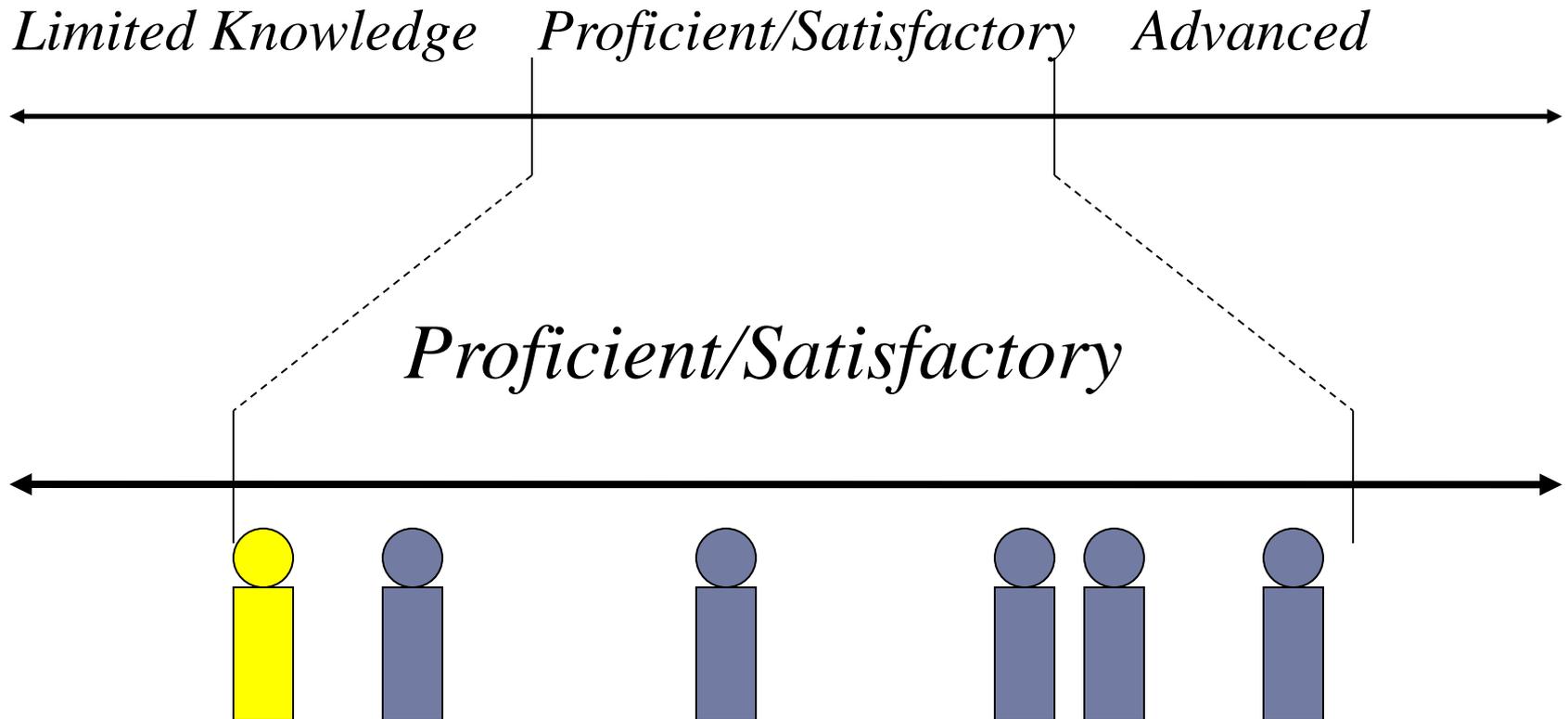


**Most  
Difficult  
Item**

Difficulty  
based on  
student  
performance

**Least  
Difficult  
Item**

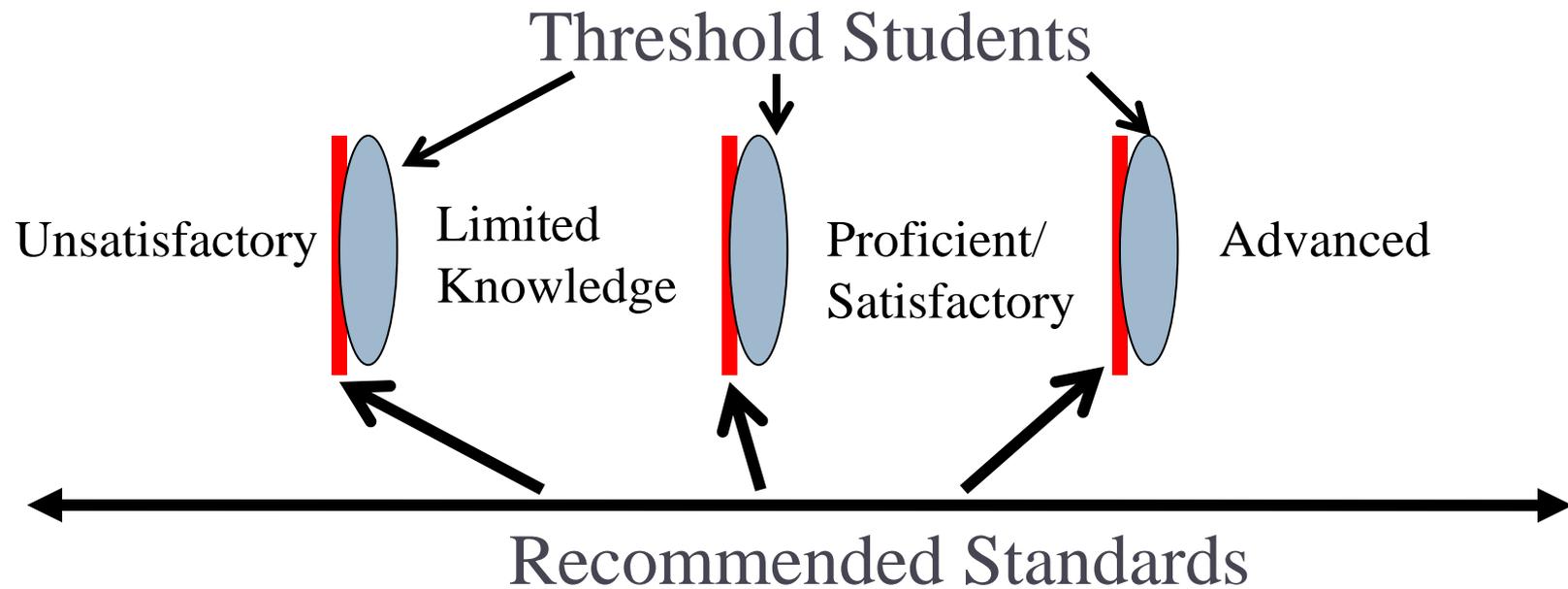
# Who Is the Target Student in Standard Setting?



# Threshold Students

## The Threshold-Student:

- Borderline or minimally qualified student in terms of performance



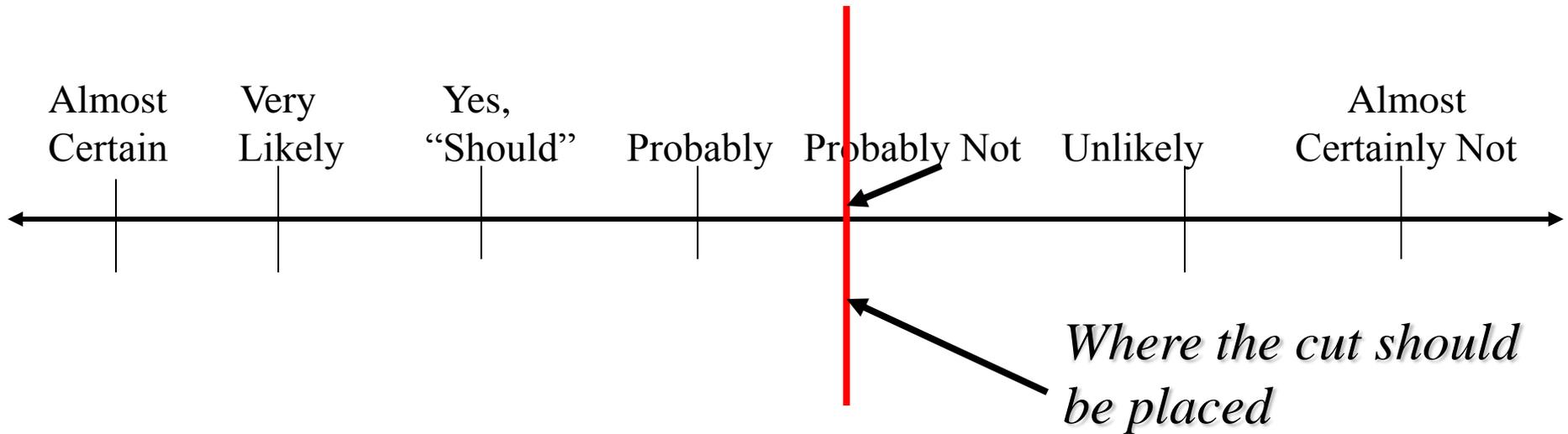
# The Bookmark Task

Place a bookmark on the first item in the ordered item booklet where the knowledge and skills measured by that item exceed the knowledge and skills of a student who just meets the performance level description.



# Degree of Confidence that a “Threshold” Student Should Answer an Item Correctly

Item Position in the Ordered Item Booklet:



# The Process

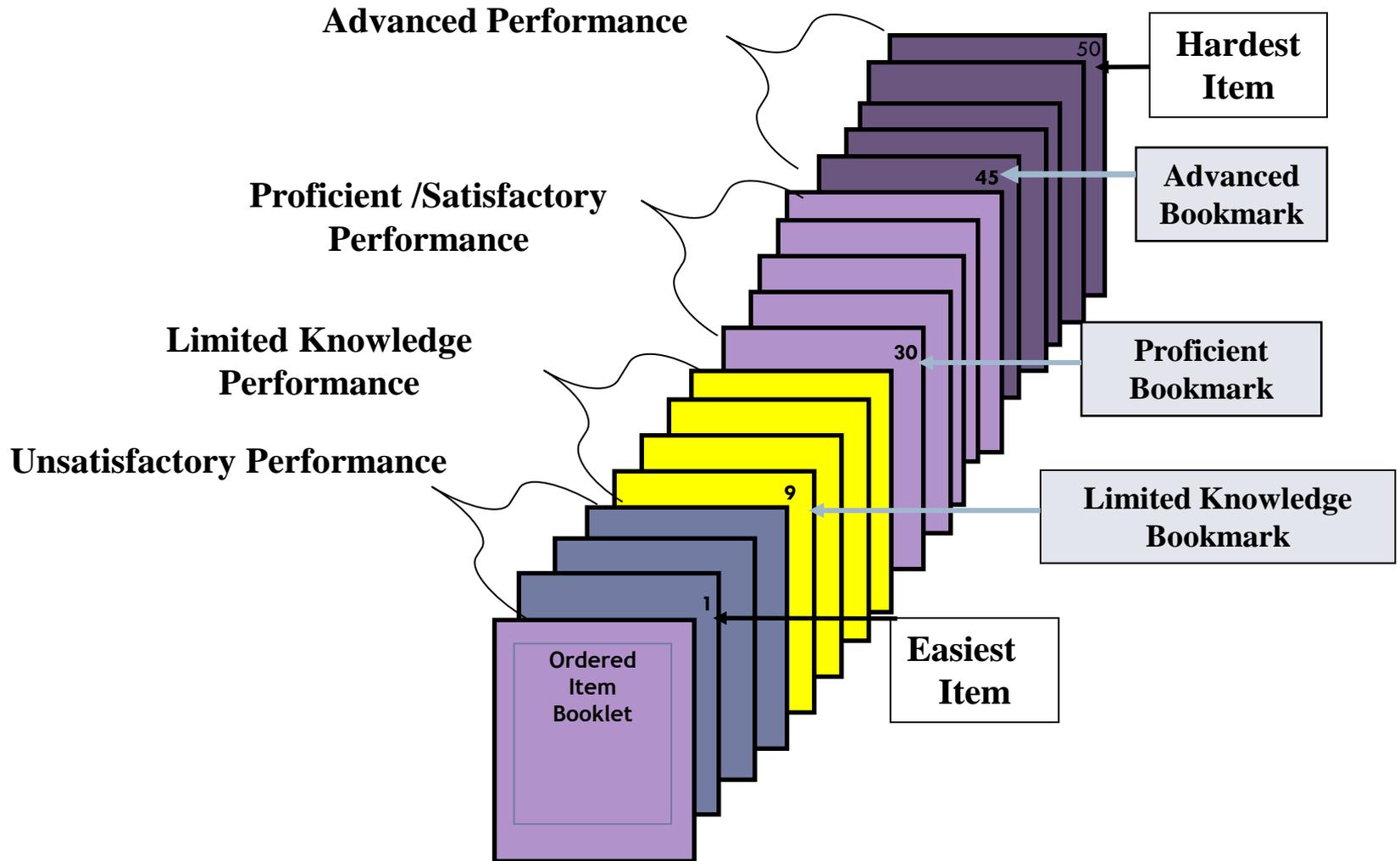
---

- Panelists receive training.
- Discuss and take the test.
- Review content standards and PLDs.
- Discuss threshold student at each level.
- Practice bookmark placement.

# The Process

- Review the Ordered Item Booklet, discussing each item with the other panelists at the table.
- Final review of task.
- Round 1: Make first bookmark judgment.
- Round 2: Receive and discuss feedback with other panelists at the table; make second bookmark judgment.
- Round 3: Receive and discuss feedback with all panelists tables; receive student impact data; make final bookmark judgment.

# End Product of Item Mapping Rounds



# After Standard Setting

What happens after the panel is finished?

- Tables are created showing the panel's recommended cut scores based on the median cut score as well as other possible cut scores within a defined interval.
- The OSDE reviews these recommendations with the State Board of Education, who adopts cut scores.
- Score reports are sent to districts.

# Item Review Committees

- ❑ Must be nominated or recommended by District/Site Administrator or Curriculum Director
- ❑ Form is available on the secure SDE site or upon request from SDE by an Administrator.
- ❑ Must be received by the end of December 2009, through the secure Administrators site, faxed, or mailed