

Standard Setting for NGSS

Presentation at the Reidy Interactive Lecture Series (RILS)

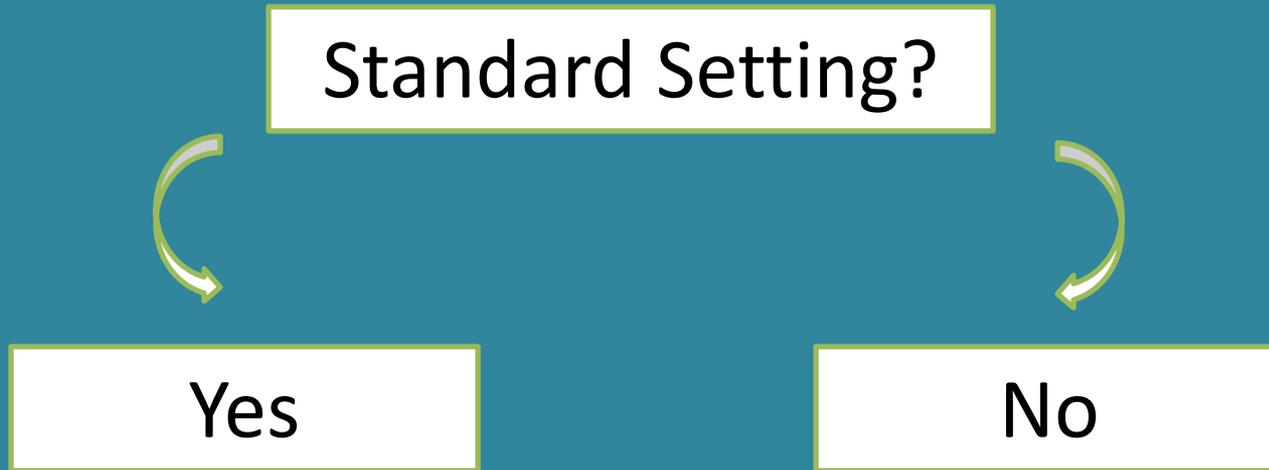
Leslie Keng, Center for Assessment

September 29, 2017

Presentation Outline

- Overview of Standard Setting (maybe)
- Standard Setting Considerations for NGSS
 - Performance Level Descriptors
 - Standard Setting Committee Meetings
 - Use of Empirical Data

Choose Your Own Adventure...



Standard Setting Overview

- A judgmental, value-based process used to establish performance standards for an assessment program.
- The process typically involves a well-defined, legally-defensible approach to obtain cut score recommendations from stakeholders.
- Standard setting for an assessment program usually involves educators and subject matter experts who know the test-taking population and assessed content.

Standard Setting Overview

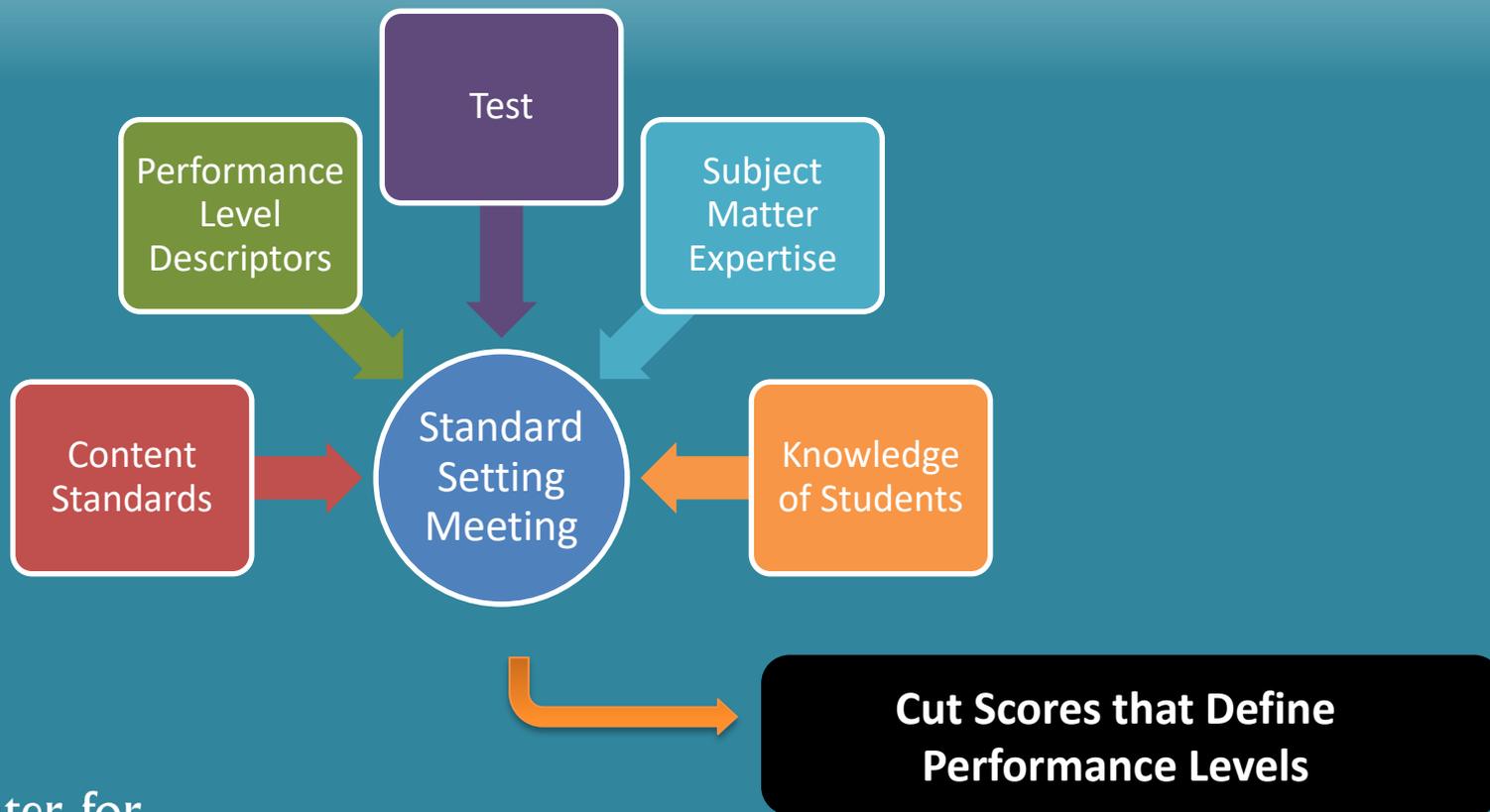
- A judgmental, value-based process used to establish performance standards for an assessment program
- The process typically involves a well-defined, legally-defensible process that results in a standard that is recorded
- Standard setting for an assessment program usually involves educators and subject matter experts who know the test-taking population and assessed content.

Gives meaning to test scores

Standard Setting Concepts

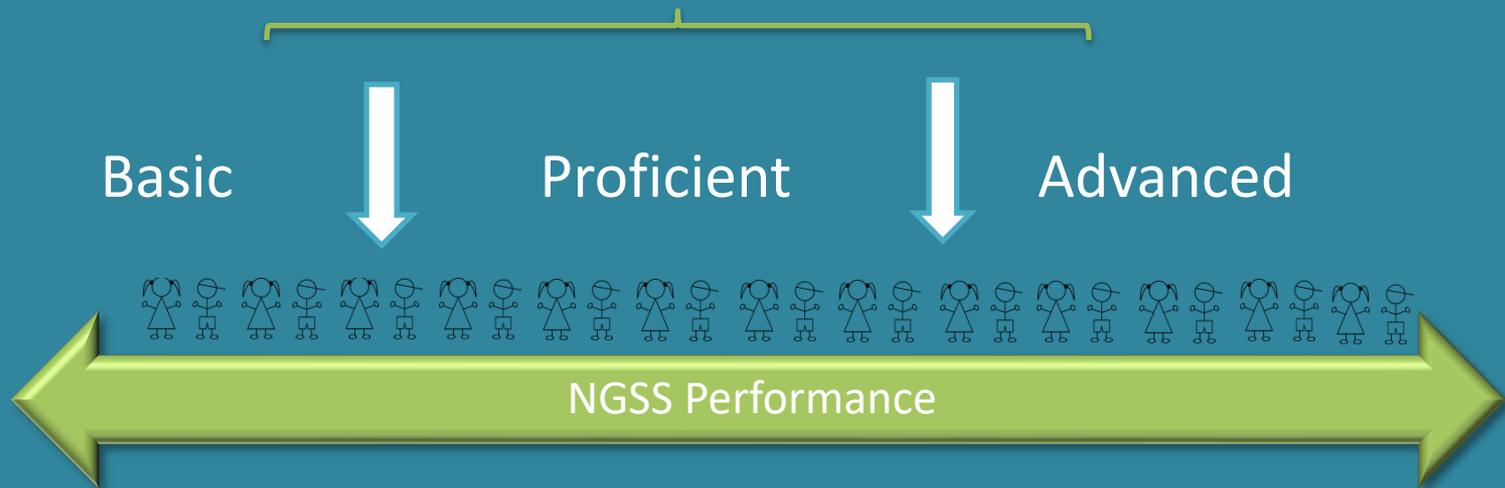
- Performance Level Descriptors (PLDs)
 - Statements that describe the expected knowledge, skills and abilities (KSAs) of students in each performance level for a given assessment.
- Performance Standards (or Cut Scores)
 - Points on the score scale that define the performance levels for a specific assessment.
- Standard Setting Committee
 - A group of educators and subject matter experts that convene and follow an established process (standard setting procedure) to recommend performance standards.

Standard Setting Overview



Standard Setting Overview

Cut Scores (Operationalization of PLDs)



Standard Setting for NGSS

- Distinguishing features of NGSS
 - Representation of science as three interwoven dimensions: *Science and Engineering Practices (SEP)*, *Disciplinary Core Ideas (DCI)* and *Cross-cutting Concepts (CCC)*
 - *Performance expectations (PEs)*, which provide examples of assessment targets that integrate the three dimensions.
- Implications for establishing performance levels for NGSS assessments
 - PLDs, standard setting committee meetings, use of empirical data

Acknowledgement

- Some of the ideas described in this brief are based on the NGSS standard setting process implemented in
 - Oklahoma
 - Washington D.C.
- More information is in the presentation, *Operationalizing NGSS Assessments: Performance Level Descriptors, Alignment Studies, and Standard Setting* (NCSA, 2017).
 - <https://ccsso.confex.com/ccsso/2017/webprogram/Session4997.html>

Performance Level Descriptors (PLDs)

- PLDs are statements that describe the expected knowledge, skills and abilities (KSAs) of students in each performance level for a given assessment.
 - PLDs could be used to inform item development, in the standard setting process, and/or as part of score reports.
- NGSS performance expectations (PEs) serve a similar purpose and could be used as the basis for PLDs.

NGSS PEs vs. PLDs

- PEs are written at a more generic level and are usually not specific enough to differentiate between students in 3 to 4 different performance levels.
 - 2-PS1-2: “Analyze data obtained from testing different materials to determine which materials have the properties that are best suited for an intended purpose.”
 - What is expected of “Proficient” students? “Advanced” students? “Basic” or “Below Basic” students?

NGSS PEs to PLDs

- Breadth – PLDs should distinguish the *amount* of things student should know or be able to do in the PE.
 - For 2-PS-1-2, the PLDs should show a progression across performance levels in the amount of “different materials” tested to obtain data, and in the number of “properties” of materials to identify.
- Depth – PLDs should describe the *level of cognitive complexity* at which students can apply knowledge.
 - For 2-PS-1-2, the PLDs should show a progression across levels that differentiate the types or complexity of data to be analyzed and the intended purposes of the different materials.

NGSS PEs vs. PLDs

- Accuracy – PLDs should differentiate how accurately or *consistently* students can demonstrate the PE.
 - Include clauses such as “most of the time”, “sometimes”, “seldom”
- Level of support – PLDs should describe how *independently* students can demonstrate the PE.
 - Can be based on features in the test items, such as graphs, charts, or interactive features, or test accommodations such as manipulatives or guided instructions.
 - Particularly important for PLDs written for NGSS assessments taken by SWD or ELLs.

Right-Sizing PLDs

- NGSS defines
 - 8 SEPs × 39+ DCIs × 7 CCCs = 2,184 potential PEs!!!
- Need to organizing the PEs into manageable but logical groupings or categories PLDs for the PLD development and standard setting committees.
 - Oklahoma: bundled PLDs by SEPs
 - DC: organized PLDs into 4 categories: conceptual understanding, performances, application, and communication.

Standard Setting Committee

- Consider including representatives from
 - Institutions of higher education (e.g., professors from the science or engineering department)
 - Professional organizations (e.g., research scientists or engineering practitioners)

Standard Setting Meeting Flow

- Experience the NGSS assessments
- Review and discuss PLDs
- Generate descriptors for borderline students
- Engage in rounds of judgments and feedback
- Participate in vertical articulation

Standard Setting Method

- Item-centric (bookmark, Angoff, ID matching) vs. student-centric (body of work) procedures
 - OK used bookmark and DC used modified Angoff (yes/no)
- Use of ordered “item” booklet (OIB)
 - Is it appropriate to “break up” NGSS tasks from the same set or cluster and spreading them across an OIB?

Standard Setting Procedure

Potential approaches

1. Present with each item the associated stimulus from the cluster
2. Treat each cluster as indivisible and make it the basic unit on which the committee members make their judgments
3. Use a procedures that does not require an OIB (such as Angoff or body of work)

Use of Empirical Data

- External validity evidence for performance standards
 - Support claims of college and career readiness for cut scores on next generation assessments
- Potential sources of external validity data for NGSS
 - Nationally-recognized tests for college admissions or college credits (e.g., science tests for ACT, SAT subject tests, AP or IB)
 - The NAEP science test (note: used by OK)
 - Assessment data from other K-12 science assessments.
 - Longitudinal college-level performance data

Use of Empirical Data

- Prior to the standard setting meetings
 - Inform reasonable (or “policy”) ranges for the cut scores
- During the standard setting meetings
 - Provide context for the recommended cut scores to help committee members evaluate the reasonableness of their cut scores
- After the standard setting meetings
 - Use as validity evidence in technical documentation and in the communication of the NGSS performance standards to the stakeholders

Center for Assessment
www.nciea.org



Leslie Keng
(lkeng@nciea.org)