GUIDANCE IN THE STANDARDS FOR CLASSROOM ASSESSMENT PRACTICES TO SUPPORT INSTRUCTIONAL DECISIONS: USEFUL OR IRRELEVANT?

Steve Ferrara, Measured Progress; K. Maxey-Moore, Denver Public Schools; Susan M. Brookhart, Duquesne University

April 14, 2018

In S. Brookhart (Organizer), Classroom Assessment and Educational Measurement, a coordinated session conducted at the annual meeting of the National Council on Measurement in Education, New York
Overview

• Line of reasoning ➔ POC ➔ goal of this session
• The standards we selected for the POC
• Illustration of translated standards using 1.11, *Content Oriented Evidence*
• Illustration of the POC test for 1.11
Line of reasoning

• ~Disregard psychometric formulations of fundamental measurement concepts in the *Standards* like reliability and validity...
  – Re-conceptualize them for classroom assessment (e.g., Brookhart, 2003; Moss, 2003; Smith, 2003; Shepard, 2006)
• “Classroom teachers would benefit from reading the *Standards*” (Plake & Wise, 2014, p. 6)
• “The standards are written in our technical language... [not teacher language, and]...a translation of relevant *Standards* into standards for classroom assessment practice could be valuable” (Ferrara, 2014, p. 25)
• Report on proof-of-concept tests of this claim
Comments on line of reasoning

• We’re aware of the *Classroom Assessment Standards* (2015) and recommendations in the various widely used textbooks on classroom assessment for teachers

• We wanted to conduct the POC on the *Standards*, independent of everything else
Procedures

• Give the standard, verbatim
• Translate it for classroom assessment purposes
• Evaluate each translated standard in terms of relevance for several classroom assessment practices
  – Teacher-made or selected unit tests
  – Teacher inferences about student understanding...
  – SLOs aligned with competencies used for high stakes decisions
  – Commercial formative assessment products
Standards we selected for the POC

• Content-oriented evidence (1.11)
• Evidence regarding cognitive processes (1.12)
• Reliability of inferences about student learning and learning needs: Decision consistency (2.16)
• Fairness: Minimizing threats from construct-irrelevant requirements (3.2)
• Others in the book chapter
ILLUSTRATION OF TRANSLATION, USING STANDARD 1.11, CONTENT-ORIENTED EVIDENCE
Content-oriented evidence (1.11)

• When the rationale for test score interpretation for a given score use rests in part on the appropriateness of test content...

• the procedures followed in specifying and generating test content should be described and justified...

• with reference to the intended population to be tested and the construct the test is intended to measure or the domain it is intended to represent...

• If the definition of the content sampled incorporates criteria such as importance, frequency, or criticality, these criteria should also be clearly explained and justified.
Interpretation

• Sentence 1 parsed:
  – Appropriateness of test content
  – Procedures to specify and generate test content
  – For the intended tested population
  – The construct to be measured or domain that is represented

• Sentence 2:
  – The references to “important, frequency, or criticality” are relevant to, for example, justifying the score point weights assigned to various instructional objectives, based on their importance and/or their emphasis in instruction
Translation

• Standard 1.11 focuses on appropriateness of test content for examinees, test development procedures, and the construct or content domain represented by the test. The content appropriateness criterion in this standard is particularly relevant to classroom assessment practices. The learning outcomes targeted in any classroom assessment should be those outcomes that have been taught and only those outcomes that were taught. In addition, the ways in which the outcomes are assessed should be consistent with intended learning outcomes.

• Summary: *Alignment, in both content area and weight, between the content and format of a classroom assessment activity and the knowledge and skills targeted in learning outcomes*
Rationale for translating

• Making learning outcomes clear to students is fundamental to teaching and learning.
• It is equally fundamental to designing tests, including classroom assessments.
• Teachers can form useful inferences about student learning and learning needs when they’ve made learning outcomes clear and assess those outcomes in appropriate ways.
POC TEST USING HYPOTHESESIZED TEACHER INFERENCE QUESTIONS
Definitions

• Classroom assessment to support instructional decisions and student learning
• Teacher-made or selected unit tests
• Teacher inferences about student understanding of skills, concepts, and procedures during instruction
• SLOs aligned with competencies uses for high stakes decisions
• Commercial formative assessment products: Interim and benchmark assessments
## (1.11) Content-oriented evidence

<table>
<thead>
<tr>
<th>Translations of the Standards to:</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Teacher-made or Selected Unit Tests</td>
<td>Teacher Inferences about Student Understanding of Skills, Concepts, and Procedures during Instruction</td>
</tr>
</tbody>
</table>

**Content-Oriented Evidence (standard 1.11)**

*Alignment, in both content area and weight, between the content and format of a classroom assessment activity and the knowledge and skills targeted in learning outcomes*

- Do the items on this test reflect all of the important learning outcomes? Do the proportions of items reflect the weight of the learning outcomes in the content domain?
- Have I gathered evidence that students have or have not mastered each of the learning outcomes in this lesson?
- Is the evidence gathered to illustrate mastery of each SLO aligned with each SLO?
- Do the items on this assessment cover all learning outcomes adequately? Are the item response demands consistent with the content standards targeted during the period of instruction prior to test administration?
(1.12) Evidence of cognitive processes

<table>
<thead>
<tr>
<th>Translations of the Standards to:</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Teacher-made or Selected Unit Tests</td>
<td>Teacher Inferences about Student Understanding of Skills, Concepts, and Procedures during Instruction</td>
</tr>
</tbody>
</table>

Evidence Regarding Cognitive Processes (standard 1.12)

The cognitive processes required to respond to an item should be consistent with the cognitive processes specified in the learning outcome that is target by item.

Do the items on the test reflect the cognitive process outlined in the content standard? Are levels of questioning scaffolded throughout instruction, going from lower level cognitive processes to higher levels, to meet the level outlined in the content standard? Are evidences of mastery aligned to cognitive processes outlined in SLOs? Are the cognitive demands of items (e.g., DOK levels) aligned to the content standard the item targets and the corresponding Achievement Level Descriptor?
(2.16) Score reliability: Decision Consistency

<table>
<thead>
<tr>
<th>Translations of the Standards to:</th>
<th>Score Reliability: Decision Consistency (standard 2.16)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Teacher-made or Selected Unit Tests</td>
<td>Sufficiency of information about the learning outcomes for students so that my decisions about providing or not providing additional instruction on the outcomes is accurate and in the best interest of the students</td>
</tr>
<tr>
<td>Teacher Inferences about Student Understanding of Skills, Concepts, and Procedures during Instruction</td>
<td>Does this unit test score accurately distinguish students who have and have not mastered the instructional content?</td>
</tr>
<tr>
<td>SLOs Aligned with Competencies Use for High Stakes Decisions</td>
<td>Did these students master the targeted learning outcomes? What additional instruction do they need? Have I accurately identified those students who do and do not need additional instruction?</td>
</tr>
<tr>
<td>Commercial Formative Assessment Products: Interim and Benchmark Assessments</td>
<td>Is the assessment data relevant to each SLO adequately rigorous, when aggregated, to support accurate identification of students who have and have not met graduation readiness decisions?</td>
</tr>
<tr>
<td></td>
<td>Does information from this assessment accurately distinguish students who are and are not progressing toward end of school year proficiency standards?</td>
</tr>
</tbody>
</table>
(3.2) Fairness: Removing construct-irrelevant requirements

<table>
<thead>
<tr>
<th>Translations of the Standards to:</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Teacher-made or Selected Unit Tests</strong></td>
</tr>
</tbody>
</table>

**Fairness (standard 3.2)**

*Classroom assessment information about student learning focuses on student mastery of learning outcomes while minimizing impacts of construct-irrelevant features in the assessment activity that may obscure inferences about student knowledge and skill.*

Are the items on this test written to minimize barriers and maximize flexibility for all (or most) learners? Conversely, do the items on this unit test contain impediments to some students that may obscure what they know and can do in relation to the targeted learning outcomes? Conversely, when I require students to explain understanding of mathematics, science, and social studies concepts, do I remove the influence of expression limitations and grammatical errors from my inference about conceptual understanding? In my observation of student performance or evaluation of student products, am I accounting for impediments in the assessment activity that I may not be able to remove and that may obscure what some students know and can do?

Is any SLO or evidence associated with an SLO influenced by impediments that obscure what students know and can do? Do the items on this assessment product contain impediments to some students that may obscure what they know and can do in relation to the targeted learning outcomes?
Discussion and conclusions

1. Identifying which standards may be relevant to classroom assessment practices and then translating those standards is challenging.

All translations are debatable.

The challenge comes from the complexity of the standards themselves and the difficulty of shifting from psychometric thinking and applications to classroom formative assessment practices and thinking.
Discussion and conclusions

2. Some standards are clearly relevant to classroom assessment (e.g., 1.11, 1.12, 2.16, and 3.12), some less so or not at all (e.g., Chapter 5: Score, Scales, Norms, Score Linking, and Cut Scores).

3. The organization of the Standards—for example, the chapter topics and clusters—suggests a framework for selecting topics for organizing standards for classroom formative assessment practice.

Perhaps many of the Standards could be translated as standards for classroom formative assessment practice.
Discussion and conclusions

4. The *Standards* are not as comprehensive as the *Classroom Assessment Standards*.

The classroom standards cover **Foundations** (e.g., Assessment Purpose, Student Engagement in Assessment), **Uses** (e.g., Analysis of Student Performance, Effective Feedback), and **Quality** (e.g., Cultural and Linguistic Diversity, Reliability and Validity), some which are not addressed in the psychometric standards.

5. Conclusion 4 raises the question, Should we continue translating the *Standards*? Or ~disregard them for the *Classroom Assessment Standards*, as others have proposed (see the discussion in the introduction to this paper)?

That will require an evaluation of the comprehensiveness of the classroom standards in relation to relevant psychometric standards.
Discussion and conclusions

6. Translating the standards for teachers is a start—only a start
Definitions

• Classroom assessment to support instructional decisions and student learning
  – CA is a broad and evolving conceptualization of a process that teachers and students use in collecting, evaluating, and using evidence of student learning for a variety of purposes, including diagnosing student strengths and weaknesses, monitoring student progress toward meeting desired levels of proficiency, assigning grades, and providing feedback to parents (McMillan, 2013, p. 4)
Definitions

• Teacher-made or selected unit tests
  – Teacher-made: Assess acquisition of knowledge and skills after a unit of instruction and make an entry in a course grade book
  – Teacher-selected: Teacher-selected unit tests typically accompany curriculum materials such as textbooks.
  – Intended SIUs: When teachers administer unit tests, they typically intend to make summative inferences about what students have learned, know, and can do at the conclusion of the instructional unit.
Definitions

• Teacher inferences about student understanding of skills, concepts, and procedures during instruction
  – Intended SIUs: When teachers pose questions during instruction, they typically intend to make formative inferences about the degree to which students have learned intended learning outcomes and what additional instruction and learning they may need.
Definitions

• SLOs aligned with competencies uses for high stakes decisions
  – In states and school systems, Student Learning Objectives (SLOs) may be course-long learning objectives set by teachers to identify and monitor student progress along a learning progression towards critical learning outcomes.
  – Intended SIUs: By implementing SLOs, school systems seek to maximize student growth by making effective teaching and assessment practices a part of every teacher’s planning.
Definitions

• Commercial formative assessment products: Interim and benchmark assessments
  – Interim formative assessment products offered to schools and school districts by commercial vendors typically are parallel test forms that cover the content standards for an entire school year and can be administered at multiple times throughout the school year (e.g., fall, winter, and spring).
  – Benchmark formative assessments typically are non-parallel test forms that cover a selected portion of the content standards for an entire school year (e.g., quarter 1) and are intended to be administered at a specified point in the school year, consistent with the scope and sequence of a curriculum.
  – Intended SIUs: When teachers use student performance and achievement growth information from interim and benchmark assessments, they typically intend to make formative inferences about which students may be on track to reach a Proficient standard by the end of the school year and which may need additional instructional support beyond the standard classroom curriculum to get on track to reach that standard.
References


Thanks!

Steve Ferrara
ferrara.steve@measuredprogress.org
603-749-9102

Kristen Maxey-Moore
kristen_maxey-moore@dpsk12.org

Sue Brookhart
susanbrookhart@bresnan.net