Standard Setting for Competencies: The Sum of the Parts or…. Some of the Parts

Judy A. Shea, PhD
Professor of Medicine
Associate Dean of Assessment

November 7, 2019
Setting the Stage

- In an era of competency based assessment
  - Milestones, EPAS

- Systems of assessment
  - Make us look broadly at performance
  - Value formative and summative assessments

- Have examples of different approaches
  - Clinical Competency Committees
  - Portfolio based programs e.g., Cleveland Clinic Lerner COM
  - Programmatic assessment
  - Suggest conjunctive versus compensatory decisions

- Lots of literature from 60s-90s on standard setting
  - Largely confined to static assessment

- Literature largely silent on how to set standards for systems of assessment that ‘pass’ learners on to the next level

- Need a framework to pull it together
Objectives

- Walk through where we are now
- Ask a lot of questions
- Outline 4 key issues
- And 4 possible frameworks
- End with unanswered questions
Principles of standards

- Transparency
- Consistency
- Involve experts
- Judgments
- Arbitrary
We have a course
We actually have lots of courses
..in each “year”
..in each “year”
..in each “year”
What do we know about these courses?

- They have different doses (e.g., length)
- They differentially target competencies
Three courses and six competencies

Course 1

Course 2

Course 3

C1  C2  C3  C4  C5  C6

Penn Medicine
What do we know about these courses?

- They have different doses (e.g., length)
- They differentially target competencies
- Courses and competencies have varying “weights” based on:
  - Dose
  - “Importance”
  - “Match”
- Most will have some type of assessment
  - Might (hopefully) relate to blueprint of competencies
- Misalignment between competencies taught and competencies assessed
Competencies .... Courses versus Assessments
What do we know about these courses?

- They have different doses (e.g., length)
- They differentially target competencies
- Courses and competencies have varying “weights” based on:
  - Dose
  - “Importance”
  - “Match”
- Most will have some type of assessment
  - Might (hopefully) relate to blueprint of competencies
- Misalignment between competencies taught and competencies assessed
  - Need to keep standard setting task focused on what is assessed
  - Curriculum and scoring are not part of standard setting
- Most assessments will have a standard
  - Implicit
  - Explicit
Issue 1

Over time/courses/units

- We have multiple assessments per competency
- (And they are probably assessed with different methods)
Issue 2

Over time/courses/units.

- We have multiple assessments per competency
- Usually we do not have a ‘grade’ per competency within each assessment
When we actually look at students we see..

Student 1:  
Course 1 – P  
Course 2 – 84  
Course 3 – H  
Course 4 – HP

Student 2:  
Course 1 – P  
Course 2 – 93  
Course 3 – P  
Course 4 – H

Student 3:  
Course 1 - F  
Course 2 - 80  
Course 3 - HP  
Course 4 –P

Single grade – NOT one for each competency
Static look in time
Issue 3

Over time/courses/units.

- We have multiple assessments per competency
- Usually we do not have a grade per competency within each assessment

- **Not all assessments have grades**
  - Qualitative/quantitative dilemma
  - And maybe we don’t value all competencies equally
  - At the very least we don’t assess all competencies equally
Over time/courses/units.

- We have multiple assessments per competency
- Usually we do not have a grade per competency within each assessment
- Not all assessments have grades
  - And maybe we don’t value all competencies equally
  - At the very least we don’t assess all competencies equally
- The amount of data we have over time grows
Assessment 1

Test 1

- C1
- C2
- C3
- C4
- C5
- C6
Decision 1

Test 1  Decision 1
C1  C2  C3  C4  C5  C6
Decision 2

Decision 1
Test 2
Decision 2

C1
C2
C3
C4
C5
C6

C1
C2
C3
C4
C5
C6
Assessment 3

Test 3
C1
C2
C3
C4
C5
C6
Issue 4

Over time/courses/units.

- We have multiple assessments per competency
- Usually we do not have a grade per competency within each assessment
- Not all assessments have grades
  - And maybe we don’t value all competencies equally
  - At the very least we don’t assess all competencies equally
- The amount of data we have over time grows
  - Pluses:
    - Increase reliability
  - Possible negatives
    - Reverse earlier decisions
    - Reweight assessments
    - Compensatory within a competency?
  - Define decision points
  - Address how if past performance influences current/future decisions
Issue 4

Over time/courses/units.

- We have multiple assessments per competency
- Usually we do not have a grade per competency within each assessment
- Not all assessments have grades
  - And maybe we don’t value all competencies equally
  - At the very least we don’t assess all competencies equally
- The amount of data we have over time grows
  - Pluses:
    - Increase reliability
  - Possible negatives
    - Reverse earlier decisions
    - Reweight assessments
    - Compensatory within a competency?
  - Define decision points
  - Address how if past performance influences current/future decisions
- We need a framework for setting up our approach
As we approach the task we ask some questions

- Do all assessments count equally?
  - no, they never have

- Do later assessments count more than earlier assessments?
  - probably except when early assessments are a hard stop point?

- Is a decision a one time thing?
  - no BUT more enthusiasm for a ‘final’ send off

- How to (pre) define a hard stop/final stamp of approval?
  - How do we set the ‘bar’?
  - How do combine ‘bars’ within competencies and across assessments?
When we think about options…

- Spectrum of quantitative/formulaic to global/CCC
- Reminder: we are already doing it
  - AOA, MSPE bottom line
- Some framework options
  - The whole is the sum of the parts
  - The whole is some of the parts
  - The whole is more than the parts
  - The whole is different than the parts
Option 1: The whole is the sum of the parts

<table>
<thead>
<tr>
<th></th>
<th>C1</th>
<th>C2</th>
<th>C3</th>
<th>C4</th>
<th>C5</th>
<th>C6</th>
<th>Decision</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exam 1 possible points</td>
<td>18</td>
<td>10</td>
<td>10</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exam 2 possible points</td>
<td>8</td>
<td>4</td>
<td>1</td>
<td>6</td>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exam 3 possible points</td>
<td>12</td>
<td>2</td>
<td>2</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Possible points total</strong></td>
<td><strong>38</strong></td>
<td><strong>16</strong></td>
<td><strong>11</strong></td>
<td><strong>8</strong></td>
<td><strong>5</strong></td>
<td><strong>9</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Standard (70%)</strong></td>
<td><strong>26</strong></td>
<td><strong>11</strong></td>
<td><strong>7</strong></td>
<td><strong>5</strong></td>
<td><strong>3</strong></td>
<td><strong>6</strong></td>
<td></td>
</tr>
<tr>
<td>Student 1 points w/in competencies</td>
<td>26</td>
<td>14</td>
<td>9</td>
<td>7</td>
<td>4</td>
<td>8</td>
<td>Good!</td>
</tr>
<tr>
<td>Student 2 points w/in competencies</td>
<td>36</td>
<td>15</td>
<td>9</td>
<td>7</td>
<td>2</td>
<td>5</td>
<td>Nope!</td>
</tr>
<tr>
<td>Student 3 points w/in competencies</td>
<td>24</td>
<td>15</td>
<td>10</td>
<td>8</td>
<td>4</td>
<td>8</td>
<td>Nope!</td>
</tr>
</tbody>
</table>
The whole is the sum of the parts

- Looking within competency
- NOT compensatory among competencies
- Might be easy to implement – have to force all courses to use numbers/templates…for each competency
  - E.g., final grade is on a 1-10 scale or 1/100 or F/P/HP/H
- Could be transparent
- Questions remain:
  - Times and timing/intervals
- Early strong performance can make up for later poor performance and vice versa…
- Bottom line: clean but unsatisfying
Option #2: The whole is some of the parts

<table>
<thead>
<tr>
<th></th>
<th>C1</th>
<th>C2</th>
<th>C3</th>
<th>C4</th>
<th>C5</th>
<th>C6</th>
<th>Decision</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exam 1 possible points</td>
<td>18</td>
<td>10</td>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exam 2 possible points</td>
<td>8</td>
<td>4</td>
<td>1</td>
<td>6</td>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exam 3 possible points</td>
<td>12</td>
<td>2</td>
<td></td>
<td>2</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Possible points total</strong></td>
<td><strong>38</strong></td>
<td><strong>16</strong></td>
<td><strong>11</strong></td>
<td><strong>8</strong></td>
<td><strong>5</strong></td>
<td><strong>9</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Standard (70%)</strong></td>
<td><strong>26</strong></td>
<td><strong>11</strong></td>
<td><strong>7</strong></td>
<td><strong>5</strong></td>
<td><strong>3</strong></td>
<td><strong>6</strong></td>
<td></td>
</tr>
<tr>
<td>Student 1 points w/in competencies</td>
<td>26</td>
<td>14</td>
<td>9</td>
<td>7</td>
<td>4</td>
<td>8</td>
<td>Good!</td>
</tr>
<tr>
<td>Student 2 points w/in competencies</td>
<td>36</td>
<td>15</td>
<td>9</td>
<td>7</td>
<td>2</td>
<td>5</td>
<td>Good!</td>
</tr>
<tr>
<td>Student 3 points w/in competencies</td>
<td><strong>24</strong></td>
<td>15</td>
<td>10</td>
<td>8</td>
<td>4</td>
<td>8</td>
<td>Nope!</td>
</tr>
</tbody>
</table>
The whole is some of the parts

- Who decides what counts?
- Mixed messaging - hidden curriculum
- But maybe fairer
  - Not all assessments are high quality
### Option 3: The whole is more than the parts

<table>
<thead>
<tr>
<th></th>
<th>C1</th>
<th>C2</th>
<th>C3</th>
<th>C4</th>
<th>C5</th>
<th>C6</th>
<th>Decision</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exam 1 possible points</td>
<td>18</td>
<td>10</td>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td>5</td>
</tr>
<tr>
<td>Exam 2 possible points</td>
<td>8</td>
<td>4</td>
<td>1</td>
<td>6</td>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exam 3 possible points</td>
<td>12</td>
<td>2</td>
<td></td>
<td>2</td>
<td></td>
<td>4</td>
<td></td>
</tr>
<tr>
<td><strong>Possible points total</strong></td>
<td><strong>38</strong></td>
<td><strong>16</strong></td>
<td><strong>11</strong></td>
<td><strong>8</strong></td>
<td><strong>5</strong></td>
<td><strong>9</strong></td>
<td></td>
</tr>
<tr>
<td>Standard (70%)</td>
<td>26</td>
<td>11</td>
<td>7</td>
<td>5</td>
<td>3</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>Student 1 points w/in competencies</td>
<td>26</td>
<td>14</td>
<td>9</td>
<td>7</td>
<td>4</td>
<td>8</td>
<td>Stay tuned</td>
</tr>
<tr>
<td>Student 2 points w/in competencies</td>
<td>36</td>
<td>15</td>
<td>9</td>
<td>7</td>
<td>2</td>
<td>5</td>
<td>Stay tuned</td>
</tr>
<tr>
<td>Student 3 points w/in competencies</td>
<td><strong>24</strong></td>
<td>15</td>
<td>10</td>
<td>8</td>
<td>4</td>
<td>8</td>
<td>Stay tuned</td>
</tr>
</tbody>
</table>
The whole is more than the parts

- **What is more?**
  - Other skills – e.g., teamwork, community volunteerism
  - Preceptors’ comments
  - Peers’ evaluations
  - Research projects/scholarly pursuits
  - But then why not a competency?

- **How do we integrate “other” experiences/knowledge with assessed competencies?**
Option 4: The whole is different than the parts

<table>
<thead>
<tr>
<th></th>
<th>C1</th>
<th>C2</th>
<th>C3</th>
<th>C4</th>
<th>C5</th>
<th>C6</th>
<th>Decision</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exam 1 possible points</td>
<td>18</td>
<td>5</td>
<td>5</td>
<td>10</td>
<td>8</td>
<td>5</td>
<td>18</td>
</tr>
<tr>
<td>Exam 2 possible points</td>
<td>14</td>
<td>5</td>
<td>5</td>
<td>10</td>
<td>8</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Exam 3 possible points</td>
<td>12</td>
<td>2</td>
<td>4</td>
<td>3</td>
<td>8</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>Possible points total</td>
<td>5</td>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Standard (70%)</td>
<td>3</td>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Student 1 points w/in competencies</td>
<td>4</td>
<td>8</td>
<td>4</td>
<td>8</td>
<td>7</td>
<td></td>
<td>Good!</td>
</tr>
<tr>
<td>Student 2 points w/in competencies</td>
<td>2</td>
<td>5</td>
<td>2</td>
<td>5</td>
<td>7</td>
<td></td>
<td>Nope!</td>
</tr>
<tr>
<td>Student 3 points w/in competencies</td>
<td>24</td>
<td>15</td>
<td>10</td>
<td>8</td>
<td>4</td>
<td>8</td>
<td>Nope!</td>
</tr>
</tbody>
</table>
The whole is different than the parts

- What changes? And why?
- Use all assessments for formative feedback but add multiple ‘cumulative’ exams that assess all competencies
  - Downside is that we lose a lot of information
  - Perhaps not so palatable to learners
    - External
    - High stress
  - But is some ways/places we already do it….
- Possibly not the answer as schools have accountability
Other issues

- Curricula/pathways become customized
- Different careers require different (weighting of) competencies
- Grades are moving to pass/fail
- Integrated performances
Questions

- Portfolios/holistic scoring with rubrics
  - How do we get from rubrics to standards?
- How do we deal with trajectories?
  - Assume upward growth
- Variability over time - what is (lack of) reliability versus growth?
- When is knowledge too late to be helpful?
- Am I making this harder than it needs to be?