Rubrics and Standardization

Caren M. Stalburg, MD MA
Clinical Assistant Professor
Obstetrics and Gynecology and Medical Education

Unless otherwise noted, this material is made available under the terms of the
Creative Commons Attribution Share Alike-3.0 License: http://creativecommons.org/licenses/by-sa/3.0/
Intended Learning Outcomes

• Understand how to create scoring rubrics
• Become familiar with standard setting techniques
SCORE = performance on the assessment

STANDARD = acceptable score to indicate desired level of performance
## Scoring Rubrics

<table>
<thead>
<tr>
<th>Assessed domain</th>
<th>Descriptor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Performance</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Historical component</th>
<th>Novice</th>
<th>Intermediate</th>
<th>Competent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Obtains most relevant patient history, missing key components</td>
<td>o o o</td>
<td>o o o</td>
<td>o o o</td>
</tr>
<tr>
<td>Most of relevant patient history, most key elements, not all</td>
<td>1 2 3</td>
<td>4 5 6</td>
<td>7 8 9</td>
</tr>
<tr>
<td>Includes all relevant components of patient history</td>
<td>1 2 3</td>
<td>4 5 6</td>
<td>7 8 9</td>
</tr>
</tbody>
</table>

**Score**

- **Novice:** 1-3
- **Intermediate:** 4-6
- **Competent:** 7-9
# VERBAL SKILLS

<table>
<thead>
<tr>
<th></th>
<th>5</th>
<th>4</th>
<th>3</th>
<th>2</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clarity</td>
<td>Clear, easy to understand</td>
<td>Occasionally difficult to understand</td>
<td>Audience must put forth effort to</td>
<td>Unclear, difficult to understand</td>
<td>Loses audience’s attention due to lack of clarity</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>listen, poor pronunciation</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Evaluation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Volume</td>
<td>Easy to hear, doesn’t overpower audio equipment</td>
<td>Overall appropriate, some sentences trail off or are hard to hear</td>
<td>Generally audible, often hard to hear</td>
<td>Difficult to hear, poorly positioned audio equipment</td>
<td>Generally inaudible, not using audio equipment</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Evaluation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Clarity**
- **5**: Clear, easy to understand
- **4**: Occasionally difficult to understand
- **3**: Audience must put forth effort to listen, poor pronunciation
- **2**: Unclear, difficult to understand
- **1**: Loses audience’s attention due to lack of clarity

**Volume**
- **5**: Easy to hear, doesn’t overpower audio equipment
- **4**: Overall appropriate, some sentences trail off or are hard to hear
- **3**: Generally audible, often hard to hear
- **2**: Difficult to hear, poorly positioned audio equipment
- **1**: Generally inaudible, not using audio equipment
Standards

• What do the scores ‘mean’
• Thoughtful judgment by experts
  – Content of the assessment/exam
  – Purpose of the exam: stakes?
  – Criteria that can be explained and justified
  – Understand the learners or group being tested
Cut-off scores

- The number below which performance is deemed unacceptable.
  - Can have significant ramifications for the individuals and for you.

- **Relative standards** = Norm-referenced, based on the performance of a group
  - Exam mean is a set as a C, bottom 10th percentile fails. If the mean is 60%, people still pass

- **Absolute standards** = Criterion based
  - Independent of group performance. If learner gets 70% of exam questions correct, then that demonstrates they have mastered enough of the material to have an adequate performance. If none of the class scores 70% then no one passes.
Angoff Method: test centered

• How will a borderline student perform?
• Panel of expert judges each question
• Gives an estimate of the chance that a BORDERLINE student would answer correctly
• Average the estimates of the raters per item
• Average estimate for each item is then averaged across the exam providing a cut-off score
Hofstee method

• Judges make 4 judgments:
  – lowest & highest acceptable passing scores
  – lowest & highest acceptable fail rates

• Passing score = mean of 4 judgments plotted against cumulative score distribution
Hofstee

Student performance curve

Pass-fail cut point

Highest and lowest fail rates

Fail Rate

Number of items correct
Cut-points

• May have unique ramifications
  – Too many passes
  – False-positives: you pass those who shouldn’t
  – Practical aspects of administering a test
  – Dependent on ‘rectangle’ and ‘curve’ of student performance line actually intersecting….
Standard setting

- Difficult to do
- The higher the stakes, the more intentional the process should be
- Defensible, reproducible, contextual
- May need to be revisited over time