Constructing Rubrics for Performance-based Assessments

Jordan Holmes, MA, MRT(N), CHSE
Manager, Centre for Learning, Innovation and Simulation

jholmes@michener.ca  @jordanholmes
Objectives

At the end of this presentation, you will be able to:

1. Describe the fundamental principles of performance-based assessment
2. Describe how to construct rubrics that accurately and fairly assess student performance
3. Consider when to implement procedural checklists versus global rating scales
4. Describe how to establish a pass score for a performance-based assessment
What is Performance-based Assessment?

• The appraisal of a learner’s proficiency in demonstrating a particular competency or set of competencies
  – May include knowledge, skills, judgement: Observable criteria
  – Time-bound

• Two general purposes:
  – Formative: assessment for learning
  – Summative: assessment of learning
Miller’s Pyramid of Assessment

## Example – Venipuncture Checklist

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Washed hands</td>
<td>Complete/Not Complete</td>
</tr>
<tr>
<td>2 Checked that all necessary equipment is available and ready to use</td>
<td>Complete/Not Complete</td>
</tr>
<tr>
<td>3 Put on gloves</td>
<td>Complete/Not Complete</td>
</tr>
<tr>
<td>4 ...</td>
<td></td>
</tr>
</tbody>
</table>
Checklists – Strengths and Limitations

Strengths

• “Recipe” – easy for rater to follow, intuitive
• Rater can be less familiar with clinical task
• Provides template for formative feedback

Limitations

• Binary (e.g., complete/not complete) ratings may result in a loss of information
• May reward thoroughness at the expense of expertise

Do Checklists Capture Expertise?

### Global Rating Scales (GRSs)

#### Global Rating Scale of Operative Performance

Please circle the number corresponding to the candidate's performance in each category, irrespective of training level.

<table>
<thead>
<tr>
<th>Respect for Tissue:</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequently used unnecessary force on tissue or caused damage by inappropriate use of instruments</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Careful handling of tissue but occasionally caused inadvertent damage</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Consistently handled tissue appropriately with minimal damage</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Time and Motion:</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Many unnecessary moves</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Efficient time/motion but some unnecessary moves</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clear economy of movement and maximum efficiency</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Instrument Handling:</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Repeatedly makes tentative or awkward moves with instruments by inappropriate use of instruments</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Competent use of instruments but occasionally appeared stiff or awkward</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fluid moves with instruments and no awkwardness</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Knowledge of Instruments:</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequently asked for wrong instrument or used inappropriate instrument</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Knew names of most instruments and used appropriate instrument</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Obviously familiar with the instruments and their names</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Checklists vs. GRSs

A recent systematic review (Ilgen et al., 2015) examined the validity evidence of checklists vs. GRSs in simulation-based assessment of health professionals:

- Inter-rater reliability similar between checklists and GRSs
- Inter-item and inter-station reliability higher for GRSs
- Separate tasks require separate checklists – single GRS can be used across tasks
- GRSs may better capture nuanced elements of expertise
- Checklists appropriate for well-defined procedures with clear best-practice (e.g., venipuncture)
GRS Construction

• Where possible, start with published scales in the literature that have demonstrated to pass basic psychometric properties (ANTS, BARS, OSATS, OCAT, mini-CEX, etc.)
• Scale should be a collaboration – developed and validated by multiple content experts
• Well defined anchors on rating scale (e.g., what does “exceeds expectations” vs. “meets expectations” look like for each criterion?)
Setting a Pass Score

• Most frequent method for high-stakes:
  – Angoff: expert judges indicate probability that a “borderline” student would accomplish each item – total test pass score set to average of item probabilities

• Examinees are allowed to miss a fixed percentage of items, **regardless of item importance**...

• Is this a defensible practice in a competency-based program?
Safety-anchored Pass Score

• Yudkowsky et al. (2014): identify what procedural items are essential to patient/clinician safety or the outcome of the procedure
• Set a conjunctive pass score, i.e., all essential items plus a percentage of non-essential items
• Students challenge exam as many times as it takes until they are able to demonstrate mastery – deliberate practice and feedback in-between
Questions?

jholmes@michener.ca

@jordanholmes on Twitter

