Summative assessment using simulation

Teodor Grantcharov MD, PhD, FACS
University of Toronto
Outline

‣ why is assessment important?
‣ what are we assessing?
‣ what tools to use and what setting?
‣ how to set standards?
‣ examples from GS
‣ what are the consequences?
Why is assessment important?

- critical for any educational intervention
- competency based education
- move from fixed time to fixed educational outcome
- selection, progression, certification & re-certification
- address the “failure to fail” culture in surgery
What are we assessing?

- skill, competency, proficiency
- knowledge/decision-making, abilities/skills, attitudes
- technical vs non-technical
Assessment tools

- methodologies not clearly described
- limited efforts to establish standards
- not sufficient evidence for summative assessment

<table>
<thead>
<tr>
<th>Assessment method</th>
<th>Psychometric properties</th>
<th>Standard Setting</th>
<th>Studies**</th>
</tr>
</thead>
<tbody>
<tr>
<td>1) Likert scales</td>
<td>IRR, IAR, IC, CV, CCV, COV</td>
<td>ROC</td>
<td>33, 38-43</td>
</tr>
<tr>
<td></td>
<td>IRR, IC, CV</td>
<td>Not documented</td>
<td>44-46</td>
</tr>
<tr>
<td></td>
<td>IRR, IAR, IC, CV, CCV, COV, FV, V₀</td>
<td>Modified Angoff, CGM</td>
<td>35, 36, 47-71</td>
</tr>
<tr>
<td>2) Benchmarks</td>
<td>IRR, R², IC, CV, CCV, FV</td>
<td>ROC</td>
<td>31, 72-98</td>
</tr>
<tr>
<td></td>
<td>IRR, R₀, CV, CCV</td>
<td>Not documented</td>
<td>99-101</td>
</tr>
<tr>
<td>3) Binary outcomes</td>
<td>IRR, IC, CV, FV</td>
<td>Not documented</td>
<td>102-108</td>
</tr>
<tr>
<td></td>
<td>R², CV, V₀</td>
<td>Not documented</td>
<td>109-112</td>
</tr>
<tr>
<td>4) Novel tools</td>
<td>IRR, R₀, IC, CV</td>
<td>Not documented</td>
<td>113-115</td>
</tr>
<tr>
<td></td>
<td>IRR, CV</td>
<td>Not documented</td>
<td>116</td>
</tr>
<tr>
<td>5) Surrogate Outcomes</td>
<td>Not documented</td>
<td>Not documented</td>
<td>117</td>
</tr>
</tbody>
</table>


**: Study reference within each category
**: This study performed both a modified Angoff methodology as the primary standard setting approach and the contrasting groups methodology for confirmation.
Standard setting

- previous work has focused on different frameworks of validity
- well-established in oral and written examinations
- insufficiently applied in surgical education
- passing score
Contrasting group method

- define 2 groups
- passing score is at the intersection of the distributions
- used for FLS
Borderline group method

- passing score at the level of the borderline candidate
- passing score set as the mean score of all borderline performances on each station
- requires a larger sample size
- OSCE, OSATS
Angoff method

- judges review each item of the exam
- probability of a “borderline candidate” to answer correctly
- probabilities are averaged to determine a passing score
COSATS & GOSATS

- expert consensus on the content
- development of simulated tasks
- development of passing scores

OSATS for in-training assessment

- 513 PGY 1 surgical residents at UofT
- multi-station sim. based assessment using the OSATS
- 3 standard setting methodologies
- pass/fail scores established

**Table 2.** Consistency of the pass/fail status of general surgery residents across three standard setting methodologies using a compensatory model

<table>
<thead>
<tr>
<th>Pass/ fail status</th>
<th>Number of residents</th>
<th>Percent (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Consistent pass‡</td>
<td>106</td>
<td>79.7</td>
</tr>
<tr>
<td>Consistent fail**</td>
<td>21</td>
<td>15.8</td>
</tr>
<tr>
<td>Inconsistent status</td>
<td>6</td>
<td>4.5</td>
</tr>
<tr>
<td>Total</td>
<td>133</td>
<td>100</td>
</tr>
</tbody>
</table>

*pass based on overall exam score  
‡ pass status for all three methodologies  
** fail status for all three methodologies

De Montrbun S, Satterthwaite L, Leung S, Grantcharov T. ACS 2014
Assessment for practicing surgeons: OR blackbox

- record audio/video/metadata from the OR
- capture & analyze data, never been available before
- understand factors that may influence outcomes
- develop targeted educational interventions/policy changes based on real-life events and crisis scenarios

courtesy St. Michael's Hospital
Performance analysis in the OR

Procedural steps

Technical error rating tool

Non-technical skills

- Environmental factors
- OR attendance
- Patient and occupational safety
- Device software

Performance analysis in the OR

Procedural Step
- Bougie
- TRANSIT
- END Procedure
- OTHER STEPS

Technical error rating tool
- Surgical tasks
  - Abdominal Access
  - Retractors
  - Energy
  - Grasping/Dissection
  - Cutting/Stapling
  - Clipping
  - Suturing
  - Suction
  - Other

Technical Error Event
- Error modes
  - Force/Distance
  - Wrong Orientation
  - Inadequate Visualization

Rectification of Event
- NOPE - Internal

Non-technical skills
- Situational Awareness
  - Gathering information
  - Understanding information
  - Projecting and anticipating future state
- Decision making
  - Considering Options
  - Selecting and communicating options
  - Implementing and reviewing decisions
- Communication and teamwork
  - Exchanging information
  - Establishing a shared understanding
  - Co-ordinating team activities
- Leadership
  - Setting and maintaining standards
  - Supporting others
  - Coping with pressure

Person Involved
- Rating
  - Surgeon 1
  - Surgeon 2
  - Fellow
  - Trainee Senior
  - Trainee Junior

Procedural Step
- Before Procedure
- During Procedure
- After Procedure
- WAITING
Adverse events

pilot study: 54 procedures, 66 adverse events in 38 of them

- bleeding: 60%
- thermal injury: 9%
- serosal tear: 9%
- devascularization: 8%
- suture events: 6%
- soft tissue injury: 5%
- stapler failure: 2%

Bonrath E, Gordon L & Grantcharov T + own unpublished data
Are we ready for high-stakes assessment?

- selection
- in-training progression
- certification, re-certification