SURGICAL EDUCATION IN THE 21ST CENTURY:
OPTIMAL FEEDBACK & ASSESSMENT

Eva Aagaard, MD
Director, Academy of Medical Educators
Dr. Aagaard does not have any relevant commercial financial relationships to report.

Dr. Aagaard does not intend to reference off-label/unapproved uses of products or devices in this presentation.
Objectives

• Recognize the importance of feedback and assessment in learner development

• Compare and contrast feedback and assessment

• Describe effective assessment strategies in surgical education

• Describe how to give effective feedback

• Discuss the future of assessment in surgical education
Why Should You Care?

• To Err is Human 1998
  • 43,000-98,000 preventable deaths each year in US hospitals

• Utah Colorado Medical Practice Study 2000
  • ½ of all adverse events in surgical patients were preventable

• Experience Alone ≠ Skill
ACGME “Competencies”

Goals of the Outcome Project
- Developing competence as a physician
  - Patient Care
  - Interpersonal & Communication Skills
  - Professionalism
  - Practice-based Learning & Improvement
  - Systems-based Practice
  - Medical Knowledge
  - Interpersonal & Communication Skills

ACGME website
Derstine, 2006
Phase 3 (7/06-7/11)

- Full INTEGRATION of the competencies and their assessment
- Use resident performance as the basis for improvement
- Use external measures to verify resident and program performance levels
HAVE WE ACHIEVED THIS GOAL?

Why Not?
Challenges

• Clinical performance is context-specific
  • Requires multiple observations in different situations
• Raters vary- “Hawk and Dove”
  • Requires multiple raters
• Standardization of environment is challenging
  • Competence in simulation ≠ competence in REAL world
• Lack of expertise/ infrastructure to support implementation
• $$$$$$$$$$$- “unfunded mandate”

But, there is hope
DEFINITIONS

What is Assessment?
How is it Different from Feedback?
Types of Assessment

- **Summative**
  - High stakes
  - Generally standardized
  - Goal is to grade relative to peers or a gold standard
  - Determine ability to progress to next phase

- **Formative = Feedback**
  - Low stakes
  - Can be formal or informal
  - Goal is to help people improve or grow
Key Point

- NOT the TOOL itself
- The way the information from the tool is being USED
Characteristics of Effective Assessments

- Feasible
- Valid
- Reliable
- Important
Reliability & Validity

RELIABILITY
• Consistency or reproducibility of measurements
  • Intrarater/interrater: measurements are the same when repeated by same/different person
  • Test/retest reliability: measurements are the same when repeated at different times

VALIDITY
• Do results represent what they claim to?

Source: Kern et al.
Face Validity

- Degree to which instrument seems to measure what it is supposed to (aka surface/content validity)

= ANGRY
Criterion Validity

- **Concurrent validity**: Results from new instrument are the same as another proven instrument
  - Example: Pass rate of students on in-house MCQ test is same as on shelf exam

- **Predictive validity**: Instrument predicts individual’s performance on specific abilities
  - Example: Residents who pass cardiology curriculum post-test are more likely to prescribe beta blockers when treating post-MI patients

Source: Kern et al.
Construct Validity

- Instrument performs as expected when used in groups with or without the attribute being measured
  - Example: Test of surgical skills attainment
What About Importance?

Miller’s Pyramid of Clinical Competence

- Knows
- Knows how
- Shows how
- Does

Action
Performance
Competence
Knowledge

Miller. Acad Med 1990
Technical Skills

Non-Technical Skills

- Medical Knowledge
- Decision-Making
- Interpersonal & Communication Skills
- Teamwork
- Leadership
- Professionalism
WHAT ASSESSMENT METHODS ARE YOU USING?

Medical Students
Residents
Maintenance of Certification
NON-TECHNICAL ASSESSMENT TOOLS

Outside of the OR
Inside the OR
Multi-Source Feedback

- Professionalism
- Interpersonal skills
- Communication skills
- Teamwork
- Leadership

Violato, BMJ 2003

- Floor Nurse
- Patients
- Peers
- Colleagues
Mini Clinical Evaluation Exercise (Mini-CEX)

1. Medical Interviewing Skills (O Not Observed)

1 2 3 4 5 6 7 8 9

UNSATISFACTORY SATISFACTORY SUPERIOR

2. Physical Examination Skills (O Not Observed)

1 2 3 4 5 6 7 8 9

UNSATISFACTORY SATISFACTORY SUPERIOR

Kogan, JAMA 2009
Inside the OR----
Non-Technical Skills for Surgeons (NOTSS)

- Developed using task-analysis
- 4-category behavioral rating system
  - Situational awareness
  - Decision-making
  - Communication and team-work
  - Leadership

The Non-Technical Skills for Surgeons (NOTSS) System Handbook v1.2

Structuring observation, rating and feedback of surgeons’ behaviours in the operating theatre
<table>
<thead>
<tr>
<th>Category</th>
<th>Element</th>
<th>Element rating*</th>
<th>Feedback on performance and debriefing notes</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Situation awareness</strong></td>
<td>Gathering information</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Understanding information</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Projecting and anticipating future state</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Decision-making</strong></td>
<td>Considering options</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Selecting and communicating option</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Implementing and reviewing decisions</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Communication and teamwork</strong></td>
<td>Exchanging information</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Establishing a shared understanding</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Coordinating team activities</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Leadership</strong></td>
<td>Setting and maintaining standards</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Supporting others</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Coping with pressure</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* 1 Poor; 2 Marginal; 3 Acceptable; 4 Good; NA Not applicable

1 Poor  Performance endangered or potentially endangered patient safety; serious remediation is required
2 Marginal Performance indicated cause for concern; considerable improvement is needed
3 Acceptable Performance was of a satisfactory standard but could be improved
4 Good Performance was of a consistently high standard, enhancing patient safety; it could be used as a positive example for others
NA Not applicable
Prospective Observational Study of NOTSS in the OR

- Anesthetists, scrub nurses, independent trained assessors provided 715 assessments of 404 cases across 6 specialties and 85 trainees
- Evaluated:
  - Feasibility
  - Validity
  - Reliability
- Real world setting; minimally trained assessors

Crossley et al, BJS, 2011
Feasible

- Provides a common language (77%)
- Added too much time (9%)
- Provides useful feedback (69%)
- Easy to use
  - Cognitive skills (54%)
  - Interpersonal skills (75%)
Valid

• Concurrent validity
  • Correlated with technical skill assessment, especially decision-making domain

• Construct validity
  • Performance correlated with specialist training level and years of UK surgical training
Table 7. D-study: reliability modeling

<table>
<thead>
<tr>
<th>No. of cases</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.35</td>
<td>0.48</td>
<td>0.55</td>
</tr>
<tr>
<td>2</td>
<td>0.57</td>
<td>0.69</td>
<td>0.74</td>
</tr>
<tr>
<td>3</td>
<td>0.68</td>
<td>0.78</td>
<td>0.82</td>
</tr>
<tr>
<td>4</td>
<td>0.75</td>
<td>0.83</td>
<td>0.86</td>
</tr>
<tr>
<td>5</td>
<td>0.79</td>
<td>0.86</td>
<td>0.89</td>
</tr>
<tr>
<td>6</td>
<td>0.82</td>
<td>0.88</td>
<td>0.90</td>
</tr>
<tr>
<td>7</td>
<td>0.85</td>
<td>0.90</td>
<td>0.92</td>
</tr>
<tr>
<td>8</td>
<td>0.86</td>
<td>0.91</td>
<td>0.93</td>
</tr>
</tbody>
</table>

1. Different assessor(s) were assumed for each case.

Crossley et al, BJS, 2011
TECHNICAL SKILLS ASSESSMENT

Global Scales
Procedure Specific Scales
Motion Analysis
Global Rating Scales

- Measure components of effective surgical performance common to many procedures
- Objective Structured Assessment of Technical Skill (OSATS)
  - 7 performance items
    - Respect for tissue
    - Time and motion
    - Instrument handling
    - Knowledge of instrument
    - Flow of operation
    - Use of assistants
    - Knowledge of specific procedure
  - 5 point likert scale with anchors
## OSATS

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Respect for Tissue</strong></td>
<td>Frequently used unnecessarily force on tissue or caused damage by inappropriate use</td>
<td>Careful handling of tissue but occasionally caused inadvertent damage</td>
<td>Consistently handled tissues appropriately with minimal damage</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Time and Motion</strong></td>
<td>Many unnecessary moves</td>
<td>Efficient time motion but some unnecessary moves</td>
<td>Clear economy of movement and maximum efficiency</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*American Journal of Surgery, 180, Reznick R. Testing technical skill via an innovative “Bench Station” examination, 226-30, 1997*
OSATS

- 26 studies
  - 19 in sim lab
  - 7 in OR
- Construct validity (18 studies)
- Internal consistency $>0.8$ (12 studies)
- Inter-rater reliability $>0.8$ (10 studies)
- No high quality evidence in the OR

Van Hove, BJS 2010
Global Operative Assessment of Laparoscopic Skills (GOALS)

Depth perception
1. Constantly overshoots target, wide swings, slow to correct
3. Some overshooting or missing of target, but quick to correct
5. Accurately directs instruments in the correct plane to target

Bimanual dexterity
1. Uses only one hand, ignores nondominant hand, poor coordination between hands
3. Uses both hands, but does not optimize interaction between hands
5. Expertly uses both hands in a complimentary manner to provide optimal exposure

Efficiency
1. Uncertain, inefficient efforts; many tentative movements; constantly changing focus or persisting without progress
3. Slow, but planned movements are reasonably organized
5. Confident, efficient and safe conduct, maintains focus on task until it is better performed by way of an alternative approach
GOALS

• 4 studies
• Construct validity
• Predictive and concurrent validity with FLS simulator (r=0.77)
• High inter-rater reliability

Van Hove, BJS 2010
Procedure-Specific Scales

- Break-down specific elements of a procedure
  - Each component assessed independently
    - Raw score
    - Error score
    - Final procedure score
- Most commonly used for and assessed in laparoscopic cholecystectomy
  - Construct validity
  - High inter-rater reliability (>0.8)

Sarker JSLS 2006
Eubanks, J am Coll Surg 1999
Skills Lab Measurements

• Motion Analysis
  • Time
  • #of movements
  • Depth perception
  • Movement quality
    • Path length
    • Smoothness
    • Angular area & volume

• Construct validity
• Very high reliability
Challenges

• Most studies of insufficient quality
  • Un-blinded
  • Not studied in real world settings
  • Lack clear cut-off values for P/F determination

Van Hove, BJS, 2011
What Does This All Mean?

• These tools are all sufficient for FORMATIVE assessment
• Insufficient evidence for HIGH stakes SUMMATIVE assessment
• Exceptions
  • VR simulator (MIST VR)
  • LapSIM
  • FLS manual skills test
Closing the loop

FEEDBACK
Assessment is Not Enough
The Groundwork

• Set clear objectives **upfront**
  • What specific behaviors do you expect?
  • When will you give feedback?
  • SET AN APPOINTMENT
• When will you reassess performance and reset goals?

*Adapted from Ende, JAMA 1983*
Prepare

• Organize your thoughts and observations
• Where:
  • In private
• When:
  • As close to event as possible
  • Learner is ready to hear it
  • Not being paged
  • Not distracted
  • Not distraught
Use the “F Word”

• Describe the Purpose
• Label it FEEDBACK

Sostok, Acad Med 2002
Conduct the Meeting

☐ ASK
  ☐ Elicit self-reflection

☐ TELL
  ☐ Give both REINFORCING and CORRECTIVE feedback
  ☐ Be SPECIFIC and use non-judgmental language
    ☐ Behaviors not personality
    ☐ Objective, observable and MODIFIABLE
  ☐ Provide suggestions for how to improve

☐ ASK
  ☐ Own suggestions for improvement plan
  ☐ Elicit understanding of feedback- CLOSE THE LOOP
Feedback Tips

Focus feedback on behavior, not personality:

“You are aloof and uncaring and cannot be bothered to communicate even simple instructions.”

“I am not sure the patient understood your instructions. Let’s discuss ways you can improve your communication skills.”

(Alguire et al., 2008)
Feedback Tips

Focus feedback on MODIFIABLE behavior:

“Your accent is so pronounced that no one can understand you.”

“Patients sometimes have trouble understanding you. I recommend that you speak more slowly and frequently check to see if you are being understood.”

(Alguire et al., 2008)
REMEMBER....

- Limit constructive feedback
Feedback Tips

“You seem to know nothing about caring for patients with diabetes, including when to initiate insulin therapy, how to monitor therapy, therapeutic goals or how to follow up on treatment. You also lack confidence when talking with patients and give confusing instructions, and your notes are disorganized.”

“I can see that you never started a patient on insulin before. Read about initiating insulin therapy in patients with type 2 diabetes and we can discuss this patient again in the morning.”

(Alguire et al., 2008)
DOES ANY OF THIS MAKE A DIFFERENCE?

Evidence that Feedback Works.....
Does Feedback Improve Surgical Skills?

- 33 2\textsuperscript{nd} and 3\textsuperscript{rd} year Med Students at Southern Illinois University
- All given same instruction on 2-handed surgical knot-tying
- Randomized to \textbf{specific feedback} or \textbf{general compliments}
- Videotaped performing the skill before and after feedback
- Student skills scored by 3 blinded observers
- Student satisfaction with instruction assessed via 7-point global rating scale

\textit{Boehler, Medical Education 2006}
Mean Performance Scores

- Specific Feedback: P<0.001
- General Compliments: P=0.2

Total Possible Ratings 0-32; IRR >0.8
Mean Satisfaction Scores

7-Point Likert Scale (1=Very Poor; 7= Truly Exceptional)

P=0.005
Take Home Points

• Specific feedback can improve performance on procedural skills
  • FEEDBACK □ LEARNING

• Satisfaction with feedback is a poor marker of the QUALITY of feedback
  • SATISFACTION ≠ QUALITY

• Combination of specific feedback and compliments may be best option
THE FUTURE OF FEEDBACK & ASSESSMENT
The Simulated Operating Suite (SOS)- the future is now

- Fully functional OR
- Real instruments/materials
- Anesthetic simulator
- Laparoscopic simulator
- Standardized actors as OR team
Near Infrared Spectroscopy (NIRS)

- Uses cerebral perfusion patterns as a surrogate for cortical activation
  - Prefrontal cortical behavior during surgical knot-tying varies with yrs of experience
  - Neuropastic changes occur with practicing surgical tasks

Eye Tracking

- May
  - Identify thought processes/ image features of expert surgeons
  - Identify inefficient or atypical behavior
Summary & Key Points

• We have moral and social imperative to assess our learners

• There are a variety of tools available
  • Most lack sufficient accuracy to be used alone for high stakes assessment

• Solution:
  • Multiple formative assessments combined with global ratings by informed assessors

• Assessment without feedback is not enough to improve performance
THANK YOU

Questions