Enhancing Assessment Through Direct Observation

Katy Bartlett, MD
Suzette Caudle, MD
Lisa Martin, MD, MPH
Gwen McIntosh, MD, MPH

COMSEP/APPD Meeting
April 10, 2013
We have no financial disclosures to discuss.
Objectives

- List the strengths and practical applications of various assessment tools.
- Define the critical features of good assessment tools.
- State the rationale for including direct observation in clinical education.
- Recognize opportunities for direct observation assessment in your training environment.
- Identify useful direct observation assessment tools for use in your clerkship or residency program.
Warm-Up

How Observant are You?
Introduction

- In 2000 the ACGME changed the way in which residency programs are evaluated:

  - Educational Opportunities Provided -1999
  - Six Core Competencies 2000-2013
  - Milestones 2013-

- Patient Care, Interpersonal and Communication Skills and Professionalism competencies cannot be assessed by written exam.
Types of Assessments
Types of Assessments

- Written Exams
- OSCEs
- Checklists
- Oral Exams
- Simulations
- Standardized Patients
- Video Reviews
- Chart Reviews
- Global Rating Scales
- Portfolios
- 360° Evaluations
Types of Assessments

- **Written Exams**
  - May be standardized from external source (e.g. NBME) or written by local faculty members
  - May be multiple choice or short answer/essay

- **Oral Exams**
  - Face-to-face learner-evaluator encounters
  - Can be used to determine if learners can withstand stress
Global Rating Scale
- Typical clinical evaluation format
- Usually based on rater’s memory rather than utilizing direct observation

Checklists
- “Yes/No” format
- Often used to assess skill at a clinical procedure
● Chart Reviews
  ● Teacher/learner discussions based on progress notes from patient charts

● Video Reviews
  ● Learner-faculty review and critique of videotaped encounters of learner with patients

● Standardized Patients (SPs)
  ● Laypersons trained to present patient problems in a uniform fashion and to assess learner’s performance.
Simulations
- May involve models, mannequins or more dynamic computer-based or virtual approximations of clinical encounters
- Can be used for individual or team assessments – formative or summative

Objective Structured Clinical Examinations (OSCEs)
- Learners complete a series of stations where they aim to show proficiency with clinical material.
- May involve standardized patients, simulations or pencil/paper tasks.
- **360° Evaluations**
  - Utilizes data from a variety of sources, *e.g.*, self, peers, faculty, patients, nursing staff

- **Portfolios**
  - Tangible, cumulative record of clinical, scholarly and professional accomplishments
  - May include publications, patient logs, procedure logs, records of teaching activities, etc.
Miller’s Pyramid

Model of Competence

- Knows
- Knows how
- Shows how
- Does

Behaviour – skills/attitudes
Cognition – Knowledge

Professional Authenticity

Miller GE: The assessment of clinical skills/ performance
Academic Medicine (Supplement) 1990, 65: S63-S7
It’s Your Turn…

In your small group, consider the following for each type of assessment:

- Where is each useful as an assessment tool?
- What are advantages of each?
- Are there any problems or limitations for each assessment?
- Would each assessment help you to place a learner on Miller’s Pyramid? If so, what part?
Properties of Assessments

VALIDITY, RELIABILITY AND UTILITY
Case Presentation: Observing a learner in the ED

A Toddler with First Seizure

Thanks to Drs. Dan Schumacher and Brad Benson from the Milestones Working Group for writing and filming this video, respectively, and making it available for public use.
What competency can we evaluate based on this video?

Patient Care: Making informed diagnostic and therapeutic decisions (clinical judgment)
Mini-CLEX Direct Observation Tool

Mini-Clinical Evaluation Exercise (CEX)

Evaluator: ____________________________ Date: ____________________________

Resident: ____________________________  ○ R-1  ○ R-2  ○ R-3

Patient Problem/Dx: ____________________________

Setting:  ○ Ambulatory  ○ In-patient  ○ ED  ○ Other ____________________________

Patient:  Age: ________  Sex: ________  ○ New  ○ Follow-up

Complexity:  ○ Low  ○ Moderate  ○ High

Focus:  ○ Data Gathering  ○ Diagnosis  ○ Therapy  ○ Counseling

1. Medical Interviewing Skills (○ Not observed)

<table>
<thead>
<tr>
<th>1 2 3</th>
<th>4 5 6</th>
<th>7 8 9</th>
</tr>
</thead>
<tbody>
<tr>
<td>UNSATISFACTORY</td>
<td>SATISFACTORY</td>
<td>SUPERIOR</td>
</tr>
</tbody>
</table>

2. Physical Examination Skills (○ Not observed)

<table>
<thead>
<tr>
<th>1 2 3</th>
<th>4 5 6</th>
<th>7 8 9</th>
</tr>
</thead>
<tbody>
<tr>
<td>UNSATISFACTORY</td>
<td>SATISFACTORY</td>
<td>SUPERIOR</td>
</tr>
</tbody>
</table>

3. Humanistic Qualities/Professionalism

<table>
<thead>
<tr>
<th>1 2 3</th>
<th>4 5 6</th>
<th>7 8 9</th>
</tr>
</thead>
<tbody>
<tr>
<td>UNSATISFACTORY</td>
<td>SATISFACTORY</td>
<td>SUPERIOR</td>
</tr>
</tbody>
</table>

4. Clinical Judgment (○ Not observed)

<table>
<thead>
<tr>
<th>1 2 3</th>
<th>4 5 6</th>
<th>7 8 9</th>
</tr>
</thead>
<tbody>
<tr>
<td>UNSATISFACTORY</td>
<td>SATISFACTORY</td>
<td>SUPERIOR</td>
</tr>
</tbody>
</table>
Assessment with behavioral anchors:

| Sub-competency: Make informed diagnostic and therapeutic decisions that result in optimal clinical judgment. | Behavioral Anchors |
|---|---|---|---|---|
| | 1 | 2 | 3 | 4 |
| Regurgitates history and physical and then looks to supervisor for synthesis and plan | Jumps from information gathering to broad evaluation without focused differential | Synthesizes information to allow a working diagnosis and differential diagnosis that informs the evaluation and management plan | Early directed hypothesis testing and ability to discriminate between features leads to unifying diagnosis, and effective/efficient work-up and plan |

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
</table>
What is validity?

- Validity: the evidence presented to support the meaning assigned to the results of an assessment.
- Without evidence of validity, assessments have no meaning.

- E.g.: USMLE: validity of this test is the evidence presented to support that it is a good measure of a student’s knowledge of the basic sciences.

Downing, SM. Medical Education 2003; 37: 830-37.
The Components of Validity

- Five components of validity:
  - Content: Is the assessment testing what is should be testing or what was intended to be taught?
    - Blueprinting and adequate/representative sampling
  - Response Process: quality control of scoring process
  - Internal Structure: inter-item correlations and item-total correlation; reliability
  - Relationship to other variables: Does the assessment correlate well with other measures of performance?
  - Consequences: reasonableness of pass/fail cut score
- Not enough to say an assessment has “face validity”

Downing, SM. Medical Education 2003; 37: 830-37.
Reliability

- The reproducibility of assessment results over time or occasions.
  - Within one assessor (intra-rater reliability)
  - Between independent assessors (inter-relator reliability)
- Quantification of consistency
- Reported as percent agreement, the kappa statistic or generalizability theory.
- Any “real-life” assessment will have lower reliability because of uncontrolled variables and lack of standardization.

Downing SM. *Medical Education* 2004; 38: 1006-1012.
How much reliability is enough?

- **High-stakes**: coefficient of $\geq 0.9$
  - E.g. licensure or certification examinations
- **Moderate-stakes**: 0.8-0.89
  - E.g. End-of-year or end-of-course summative exam or OSCE
- **Low-stakes**: 0.7-0.79
  - Formative assessments

Downing SM. *Medical Education* 2004; 38: 1006-1012.
Validity of Direct Observation

- Spectrum of content: in the clinical setting, patients do not always match the learning objectives targeted
  - Encourage faculty to choose patients that address certain content areas (e.g., newborn, toddler, school-aged, teen, etc.)
  - Provide enough exposure to varied situations (e.g. disease and health, gamut of difficult communication situations)
  - Recognize that heterogeneous setting is where we need learners to perform
- Response Process: Need to ensure standardized raters with well-designed rating forms to remove systematic error (bias) from the measurements.

Validity of Direct Observation

- **Internal Structure:** reliability of the scores and generalizability
  - Variance in scores between and within faculty raters
  - The generalizability of the cases to the larger domain
- **Relationship to other variables:** may not need to correlate with written exams
  - Direct observation has correlated well to OSCEs, patient write-ups and clerkship scores; seems to test distinct constructs from written exams.
- **Consequences:** Usually a moderate-stakes assessment; not the sole source of a clerkship grade; often formative

## Threats to Validity

<table>
<thead>
<tr>
<th>Threats to Validity</th>
<th>How to address them</th>
</tr>
</thead>
<tbody>
<tr>
<td>Too few observations of clinical behavior</td>
<td>Multiple observations for each learner</td>
</tr>
<tr>
<td>Too few independent raters</td>
<td>Multiple raters</td>
</tr>
<tr>
<td>Low reliability of ratings</td>
<td>Rater training; reliable tool</td>
</tr>
<tr>
<td>Inappropriate rating items</td>
<td>Ensure tool matches content being tested</td>
</tr>
<tr>
<td>Low generalizability</td>
<td>Ensure appropriate spectrum of patients</td>
</tr>
<tr>
<td>Incomplete observations</td>
<td>Avoid intrusions and interruptions</td>
</tr>
<tr>
<td>Rationale for passing score not well justified</td>
<td>Appropriate weight of DO in grading/competency assessment</td>
</tr>
<tr>
<td>Rater bias</td>
<td>Tools with behavioral anchors</td>
</tr>
<tr>
<td>Systematic rater error</td>
<td>Rater training</td>
</tr>
</tbody>
</table>

Downing SM and Haladyna TM. *Medical Education* 2004; 38: 327-333.
Common Rater Errors

- Halo
- Generosity
- Central Tendency
- Hawk/Dove
- Stereotyping
- Perception

- Rater Drift
- Recency
- Uninformed
- Hawthorne
- Generic
- Development

Adapted from slides by Lisa Howley, PhD
Utility

- Combines concepts of reliability and validity with feasibility:

\[
\text{Utility} = \text{reliability} \times \text{validity} \times \text{acceptability/practicality} \times \text{cost} \times \text{educational impact}
\]

- The most reliable and valid tools may be cost-prohibitive or infeasible to implement.

Burke, A. “Measurement Principles in Medical Education.” Assessment in Graduate Medical Education: A Primer for Pediatric Program Directors. Chapel Hill: ABP, 2011.
Multiple assessments

○ “Because physicians do not perform consistently from task to task, broad sampling across cases is essential to assess clinical competence reliably. This observation ... challenges the traditional approach to clinical competence testing, whereby the competence of individuals is assessed based on a single case, namely the case observed by the assessor.”

Hicks, PJ et al. Journal of Graduate Medical Education 2010.
Direct Observation
Direct Observation

- What are the barriers to direct observation in your program?
Direct Observation

- What are the barriers?
  - Lack of time
    - Faculty production pressures
    - Resident duty hours (less face time with faculty)
  - Reluctance
    - Learner feels uneasy being observed
    - Faculty feel ill prepared to do observation
  - Tradition
    - Decades of taking learner’s word about hx, PE
Direct Observation

- What are you doing?
- What are your tools?
Direct Observation

**Miller’s Pyramid**

- **Does**
- **Shows how**
- **Knows how**
- **Knows**

---

**Observation**

*Learner observed in practice:*
Mini CEX, Skills checklist, 360 evals

*Learner observed in simulated environment:*
OSCE, Sim Lab, Standardized patients
Direct Observation Tools

- OSCE
- SCO and CSCO
- Mini CEX
- CEX
- Checklists
- Observation cards
- And many others...
Direct Observation Tools

- Many tools exist but...

  *There is no “Holy Grail”*

- Pick a few* that work for your program
  - Train faculty
  - Train learners
  - Implement consistently across program
Practical tips

- Room set up for successful observation

Adapted from E. Holmboe
Practical tips

- Holmboe’s 4 simple rules
  1. Correct positioning
  2. Minimize external interruptions
  3. Avoid intrusions: don’t interrupt
  4. Be prepared: know your goals, use tools

Adapted from E. Holmboe
Practical tips

- Prepare learner for observation
- Acknowledge discomfort on both sides
- Seek multiple observations over time

Adapted from E. Holmboe
Direct Observation Tools

- Use “real time” in actual clinic setting (no extra time carve out)
- Use across multiple clinical environments
- Short tools --> brief faculty training
- Formative assessment

↓

Reduces learner anxiety about observation
Reduces faculty reluctance to observe learner
Recognizing Opportunities for Direct Observation

Opportunities for Direct Observation abound.....
We just have to recognize them...
And capitalize on them...
Opportunities for Direct Observation

Small Group Exercise (12 min)
- Each table has a competency domain
- Consider the entire scope of your training environment
- Discuss and scribe opportunities to use DO to assess learners w/in your assigned competency domain

Large Group Exercise (12 min)
- Report out opportunities to entire group
Wrap-Up Thoughts on Observing the Competencies

○ Think **Multiples**:  
  ● multiple observations  
  ● by multiple observers  
  ● in multiple settings

○ Think **Efficient/Easy**

○ Think **Small** periods  
  of focused time  
  ● Goal setting

Think MES to avoid a MESS!
What direct assessment tools might be most useful for your program?

Practical application
Milestones and Direct Observation
Pediatrics Milestone Project

Milestones represent a sequence of

- narrative descriptions of
- observable behaviors at
- advancing levels of development
- across the continuum of education, training and practice
Contribution of the Milestones

- Take the abstract language of the competencies and translate it into narrative descriptions of behaviors
- Address the continuum from novice to expert
- Provide a learning roadmap
Terminology

<table>
<thead>
<tr>
<th>ACGME</th>
<th>International</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Competency</td>
<td>Domain of Competence</td>
<td>Patient Care</td>
</tr>
<tr>
<td>Sub-competency*</td>
<td>Competency</td>
<td>Gather Essential and Accurate Information</td>
</tr>
<tr>
<td>Milestones</td>
<td>Milestones</td>
<td>Behavioral Anchor</td>
</tr>
</tbody>
</table>

* Sub-competencies are often colloquially referred to as “milestones.”
  We will use ACGME terms.
The Good Doctor: PUTTING IT ALL TOGETHER

EPAs

COMPETENCIES

SUB-COMPETENCIES

MILESTONES
## Competency: Interpersonal & Communication Skills

<table>
<thead>
<tr>
<th>Sub-Competency</th>
<th>Developmental Milestones / Anchors</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Novice</td>
</tr>
</tbody>
</table>
| Communicate effectively with patients, families and the public, as appropriate, across a broad range of socioeconomic and cultural backgrounds | - Uses a **standard medical interview template** to prompt all questions  
- **Does not vary the approach** based on a patient’s unique physical, cultural, socioeconomic or situational needs. May feel intimidated or **uncomfortable asking personal questions of patients** | - Uses the medical interview to **establish rapport** and focus on information exchange relevant to a patient’s or family’s primary concerns  
- **Identifies physical, cultural, psychological and social barriers to communication** but often has difficulty managing them  
- Begins to use **nonjudgmental questioning scripts** in response to sensitive situations | - Uses the interview to effectively establish rapport  
- **Able to mitigate** physical, cultural, psychological and social barriers in most situations  
- Verbal and nonverbal communication skills promote trust, respect and understanding  
- Develops scripts to approach most difficult communication scenarios | - Uses communication to establish and maintain a **therapeutic alliance**  
- Sees beyond stereotypes and works to **tailor communication to the individual**  
- A wealth of experience has led to development of scripts for the gamut of difficult communication scenarios  
- **Able to adjust scripts ad hoc** for specific encounters | - Connects with patients and families in an **authentic manner** that fosters a trusting and loyal relationship  
- Effectively educates patients, families, and the public as part of all communication  
- **Intuitively handles the gamut of difficult communication scenarios** with grace and humility |
ICS1. Communicate effectively with patients, families, and the public, as appropriate, across a broad range of socioeconomic and cultural backgrounds

<table>
<thead>
<tr>
<th>Not yet Assessable</th>
<th>Level 1</th>
<th>Level 2</th>
<th>Level 3</th>
<th>Level 4</th>
<th>Level 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Uses standard medical interview template to prompt all questions; does not vary the approach based on a patient’s unique physical, cultural, socioeconomic, or situational needs; may feel intimidated or uncomfortable asking personal questions of patients</td>
<td>Uses the medical interview to establish rapport and focus on information exchange relevant to a patient’s or family’s primary concerns; identifies physical, cultural, psychological, and social barriers to communication, but often has difficulty managing them; begins to use non-judgmental questioning scripts in response to sensitive situations</td>
<td>Uses the interview to effectively establish rapport; is able to mitigate physical, cultural, psychological, and social barriers in most situations; verbal and non-verbal communication skills promote trust, respect, and understanding; develops scripts to approach most difficult communication scenarios</td>
<td>Uses communication to establish and maintain a therapeutic alliance; sees beyond stereotypes and works to tailor communication to the individual; a wealth of experience has led to development of scripts for the gamut of difficult communication scenarios; is able to adjust scripts ad hoc for specific encounters</td>
<td>Connects with patients and families in an authentic manner that fosters a trusting and loyal relationship; effectively educates patients, families, and the public as part of all communication; intuitively handles the gamut of difficult communication scenarios with grace and humility</td>
<td></td>
</tr>
</tbody>
</table>

Comments:
Take Home Points

- In the era of competency-based medical education, it is no longer enough to prove that our learners “know”; we must show that they “do”.
- Direct observation is a critical tool for proving that learners “do”.
- Direct observations may be more successful with correct positioning, minimal interruptions and intrusions, and preparation of rater and learner.
Take Home Points (continued)

- Real-world singular observations will never be perfectly valid for high-stakes assessment.
- Behavioral anchors and rater training are important methods to increase validity of observations.
- Multiple observations by multiple observers over time are key for showing competence and informing Milestones.
Take Home Materials


- Sample packet of direct observation tools

- Worksheet with pros and cons of different assessment tools. *Updated version will be posted on web.*

- Worksheet on Opportunities for direct observation. *Updated version will be posted on the web.*

- Slides from presentation on the web.