Developing Skills at Making Observations

Lessons from Faculty Development and Rater Cognition Research

Eric S. Holmboe

Jennifer R. Kogan
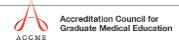
Accreditation Council for
Graduate Medical Education

Perelman
School of Medicine
UNIVERSITY of PENNSYLVANIA

Roadmap

1. Define workplace based assessment and the theories supporting direct observation

2. Identify barriers to high quality, frequent direct observation and solutions to overcome them

Perelman
School of Medicine

Accreditation Council for
Graduate Medical Education   2

Workplace Based Assessment

➢ Assessment of day-to-day practice in the authentic clinical environment

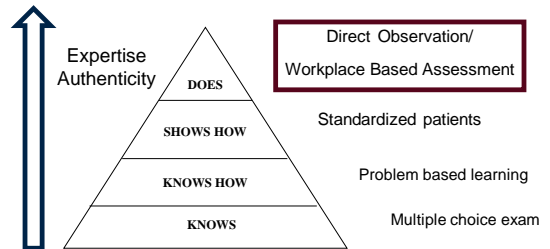➢ Assessment of what doctors actually do in practice

Swanwick T. *Br J Hosp Med.*2009;70:290-3

Perelman
School of Medicine

Accreditation Council for
Graduate Medical Education   3

Miller's Assessment Pyramid

Expertise
Authenticity

Direct Observation/
Workplace Based Assessment

DOES

SHOWS HOW — Standardized patients

KNOWS HOW — Problem based learning

KNOWS — Multiple choice exam

Miller GE. *Acad Med.*1990; 65:S63-7

Perelman
School of Medicine

Accreditation Council for
Graduate Medical Education   4

Theories Supporting Direct Observation

➢ Importance and state of clinical skills
➢ Development of expertise
➢ Role in competency based medical education
➢ Necessity in supervision

Perelman
School of Medicine

Accreditation Council for
Graduate Medical Education   5

Clinical Skills Matter

➢ History leads to diagnosis > 80% of the time
➢ Even in era of technology
➢ Required to avoid unnecessary testing
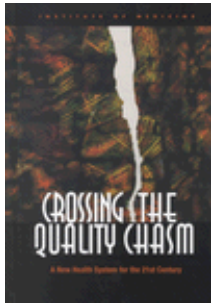➢ Faulty data gathering common source of diagnostic errors

Hampton JR et al. *BMJ* 1975; 2(5969):486-9
Peterson MC et al. *West J Med.* 1992; 156(2):163-5
Graber, M et al. *Acad Med.* 2002;77(10):981-92

Perelman
School of Medicine

Accreditation Council for
Graduate Medical Education   6

High Quality Care

- Timely
- Efficient
- Equitable
- Safe
- Effective
- Patient Centered

Crossing the Quality Chasm:
A New Health System for the 21st Century
2001

Outcomes of Patient Centered Care

- Improves communication
- Promotes patient involvement in care
  - Increases patient knowledge and self-efficacy
- Creates positive relationships with the provider
- Improves adherence
- Improves well-being
- Improved outcomes
- Decreased costs

Levinson W et al. 2010; *Health Aff* 29: 1310-18
Williams S et al. *Fam Prac.*1998;15:480-92
DiMatteo M. *Patient Educ Counsel.* 2004;55:339-44
Stewart M . *CMAJ.* 1995; 152:1423-33.

State of Clinical Skills

- **Trainees**
  - Wide variability in graduating students' clinical skills measured as MS4s or starting internship
    - History taking
    - Exam

- **Practicing physicians**
  - Variability in physical exam skills
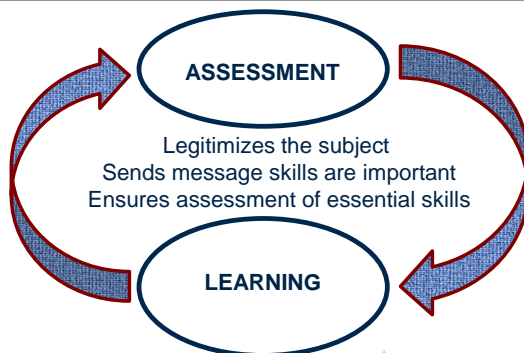 - Missing elements of informed decision making

Stillman. Ann Intern Med.1990; Sachdeva. Arch Surg.1995;
Lypson.Acad Med.2004; Mangione.1997; Braddock.1999

Why the Gap??

- Communication is a sophisticated procedure
  - Needs to be taught and honed throughout one's career

- Skills of patient-centered communication are rarely taught or practiced

Levinson W. *BMJ* Qual Saf 2011;20:823-5

Direct Observation to Assess Core Skills

ASSESSMENT

Legitimizes the subject
Sends message skills are important
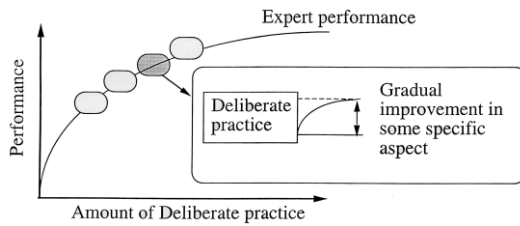Ensures assessment of essential skills

LEARNING

What Do They Have in Common?

Design and Sequencing of Training Activities

* Monitor students' development
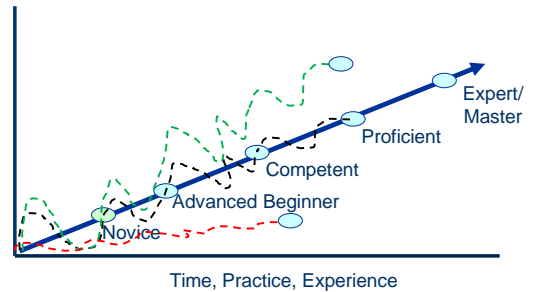* Design and select training tasks for individual students

Professional teachers and coaches

From Anders Ericsson: Used by Permission

Dreyfus & Dreyfus Development Model

Time, Practice, Experience

Dreyfus SE and Dreyfus HL. 1980
Carraccio CL et al. Acad Med 2008;83:761-7

The Role of the Coach

➢ "They observe, they judge, and they guide"

➢ "That one twenty-minute discussion gave me more to consider and work on than I'd had in the past five years"

➢ "Medical practice is largely unseen by anyone who might raise one's sights. I'd had no outside ears and eyes."
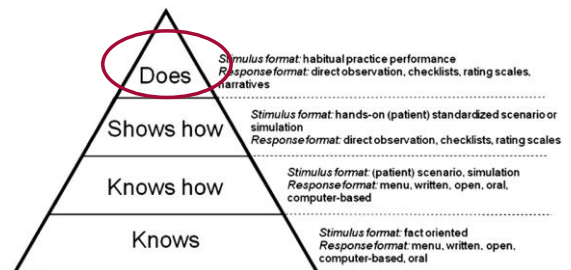
Atul Gawande, New Yorker 10/3/2011

Miller's Assessment Pyramid

Stimulus format: habitual practice performance
Response format: direct observation, checklists, rating scales, narratives — **Does**

Stimulus format: hands-on (patient) standardized scenario or simulation
Response format: direct observation, checklists, rating scales — **Shows how**

Stimulus format: (patient) scenario, simulation
Response format: menu, written, open, oral, computer-based — **Knows how**

Stimulus format: fact oriented
Response format: menu, written, open, computer-based, oral — **Knows**

Van der vleuten CPM et al. *Best Practice & Research Clinical Obstetrics and Gynaecology* . 2010(24) :703–19

In-Training Performance Assessment

➢Assessment in authentic situations

   ➢ Learners' ability to combine knowledge, skills, judgments, attitudes in dealing with realistic problems of professional practice

➢Assessment in day to day practice

   ➢ Enables assessment of a range of essential competencies, some of which cannot be validly assessed otherwise

Govaerts MJB et al. Adv Health Sci Edu. 2007;12:239-60

Observation and Safe Patient Care

Safe, effective patient-centered care

GOAL

Appropriate level of supervision**
**a function of attending competence in context

Trainee performance*
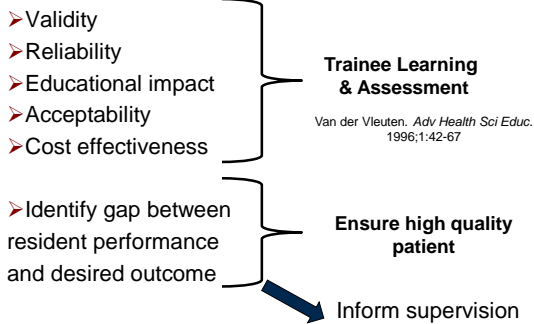* a function of level of competence in context

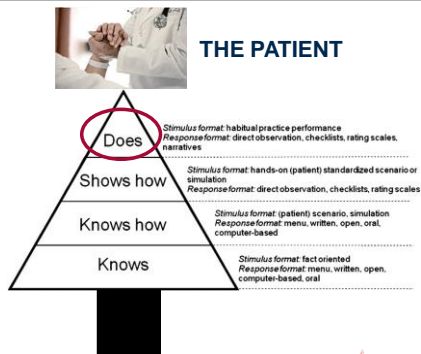Kogan JR et al. *Acad Med*; 2014;89:721-7

Utility Elements of Assessment

- Validity
- Reliability
- Educational impact
- Acceptability
- Cost effectiveness

Trainee Learning & Assessment

Van der Vleuten. *Adv Health Sci Educ.* 1996;1:42-67

- Identify gap between resident performance and desired outcome

Ensure high quality patient

Inform supervision

Perelman School of Medicine
University of Pennsylvania

Accreditation Council for Graduate Medical Education **19**

Assessing Does

THE PATIENT

Does — *Stimulus format:* habitual practice performance
Response format: direct observation, checklists, rating scales, narratives

Shows how — *Stimulus format:* hands-on (patient) standardized scenario or simulation
Response format: direct observation, checklists, rating scales

Knows how — *Stimulus format:* (patient) scenario, simulation
Response format: menu, written, open, oral, computer-based

Knows — *Stimulus format:* fact oriented
Response format: menu, written, open, computer-based, oral

Perelman School of Medicine
University of Pennsylvania

Accreditation Council for Graduate Medical Education **20**

Summary: Reasons for DO

Patient Safety

Assessment

Coaching

Perelman School of Medicine
University of Pennsylvania

Accreditation Council for Graduate Medical Education **21**

Problems with Performance Assessment

Perelman School of Medicine
University of Pennsylvania

Accreditation Council for Graduate Medical Education **22**

Key Issues: Psychometric

- **Multiple studies demonstrating major issues in intra- and inter-rater reliability**
  - Usual response – change the form or tool…
- **Limited evidence for validity**
  - Modest correlations between high-stakes assessments and faculty ratings
- **Lack of discrimination among domains of competence**
  - The "factor analytic" problem

Perelman School of Medicine
University of Pennsylvania

Accreditation Council for Graduate Medical Education **23**
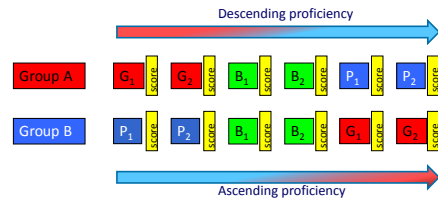

Key Issues: Errors

Multiple forms and types of error:
- **Correlational errors**
  - Halo effect
  - Horn effect
    - Ratings based mostly on *perceived* knowledge and personality
- **Distributional errors**
  - Leniency error ("Doves")
 - Severity error ("Hawks")
  - Central tendency

Perelman School of Medicine
University of Pennsylvania

24

Key Issues: Human Limitations

➢ **Limitation in working memory and mental processing**
➢ **Subconscious processes**
  ▪ Bias and stereotyping
➢ **Cognitive Load**

25

Yeates: Contrast effects

Yeates, P. et al., *Medical education*, 2013, 47(9), pp.910–22.

26

Yeates: Contrast effects

Scores by group for different levels of performance
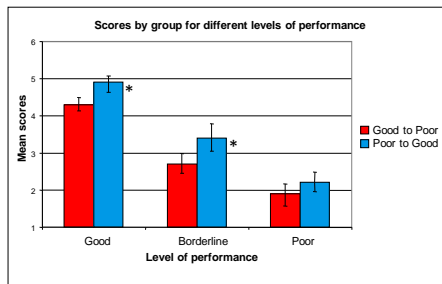
* = p <0.05   Yeates, P. et al., *Medical education*, 2013, 47(9), pp.910–22.

27

Cognitive Load

➢ **There is a limit as to how much you can ask faculty to observe and capture**
  ▪ Clinical units: complex environment
  ▪ Selective attention
➢ **Byrne et. al. (Med Educ 2014)**
  ▪ Average cognitive load for faculty judging OSCE stations *was higher* than anesthesia trainees during induction for routine surgery
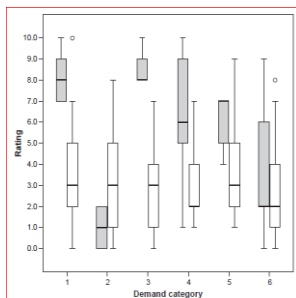   ▪ OSCE had 21-22 items in an 8 minute station

28

Cognitive Load

Figure 3 Comparison of NASA–Task Load Index (NASA-TLX) scores in the study subjects (grey boxes) and trainee anaesthetists (white boxes).

Demand categories:
1 = mental demand
2 = physical demand
3 = temporal demand
4 = performance/success
5 = effort
6 = frustration

Byrne A, Tweed N, Halligan C. A pilot study of the mental workload of OSCE examiners. Med Educ. 2014; 48: 262-67.

29

Key Issues: Individual Effects

➢ **Inference**
➢ **Variability among faculty**
  ▪ Strengths and weaknesses
    ▪ Clinical
    ▪ Educational
    ▪ Assessment
  ▪ Variable frames of reference
➢ **Idiosyncrasy**
➢ **Contextual factors**

30

Page 5

High Level Inference

The Problem with Inference

➢Inferences are not recognized

➢Inferences are rarely validated for accuracy

➢Inferences can be wrong

Types of Inference about Residents

➢**Skills**
  ➢Knowledge
  ➢Competence
  ➢Work-ethic

➢**Prior experiences**
  ➢Familiarity with scenario

➢**Feelings**
  ➢Comfort
  ➢Confidence
  ➢Intentions
  ➢Ownership

➢**Personality**

➢**Culture**

Govaerts MJB et al. Adv Health Sci Educ Theory Pract. 2011;16:151-65
Kogan JR et al. Med Educ. 2011;45:1048-60

Frames of Reference

1	2	3	4	5	6	7	8	9

Unsatisfactory	Satisfactory	Superior

ORDINAL

Below Expectation	At Expectation	Exceeds Expectation

NORMATIVE

????	????	????

GESTALT

Kogan JR et al. Med Educ. 2011;45:1048-60

1	2	3	4	5	6	7	8	9

Not What I Do	Close to What I Do	What I Do (or better)

SELF

Missing evidence based elements	Most evidence based elements	All evidence based elements

BEST PRACTICE

Kogan JR et al. Med Educ. 2011;45:1048-60

Assessors' Own Clinical Skills

Variable and sometimes deficient
➢History taking
➢Physical exam
➢Counseling/shared decision making
➢Patient centered communication

Ramsey PG et al. JAMA 1993;269:1655-60
Paauw DS et al. JAMA 1995;274:1380-2
Vukanovick-Criley JM et al. Arch Intern Med. 2006;166:610-16
Braddock CH 3rd et al. J Gen Intern Med. 1997;12(6):339-45
Kogan JR. et al. Acad Med. 2010;85(S10):S25-8
Levinson W. BMJ Qual Saf 2011;20:823-5
Frankel RM et al. Acad Med 2011;86:445-52

Faculty OSCE Clinical Skills

Competency	Mean (SD)	Range	Generaliz-ability
History Taking	65.5% (9.6%)	34% - 79%	0.80
Physical Exam	78.9% (13.6%)	36% - 100%	0.52
Counseling	77.1% (7.8%)	60% - 93%	0.33
Patient Satisfaction[1]	5.62 (0.48)	4.43 – 6.63	0.60

[1]On 7-point scale   N=44

Kogan JR. et al. Acad Med. 2010;85(10 Suppl):S25-8

Other Factors That May Impact Ratings

➢**Minimal impact of demographics**
  ➢Age, gender, clinical and teaching experience

➢**Faculty's own clinical skills may matter**
  ➢Faculty with higher history and patient satisfaction performance scores provide more stringent ratings.

Kogan JR. et al. Acad Med. 2010;85(10 Suppl):S25-8

Idiosyncrasy: What if…

Low inter-rater reliability was found to come from experts forming different and/or conflicting, yet equivalently relevant, interpretations?

Assessment as "Saturation"

➢**Narrative just as, or perhaps more, meaningful as ratings through numbers**
➢**Sampling of multiple idiosyncratic, yet meaningful, observations can lead to robust information "saturation"**
➢**Part of rationale for the inclusion of group process in the accreditation system**

Contextual Factors

➢Encounter complexity
➢Resident characteristics
➢Institutional culture

Kogan JR et al. Med Educ. 2011;45:1048-60
Kogan JR et al. Med Educ. 2012;46:201-15

Direct Observation: A Conceptual Model
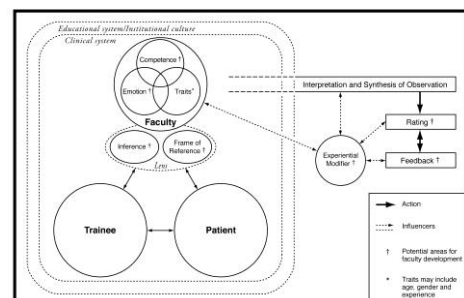
Kogan JR, et al. Med Educ. 2011

Solution: Faculty Development

➢Performance dimension training
➢Synthesis to final judgment

Kogan JR et al. JAMA. 2009;302:1316-26
Holmboe ES et al. *Ann Intern Med.* 2004;140:874-81.
Cook DA et al. *JGIM.* 2009;24:74-9.
Donato AA et al. *Med Educ.* 2008;42:1234-40.

Performance Dimension Training

Identify specific dimensions of a competency
in behavioral terms

⬇

Discuss the criteria and qualifications required for
each dimension of that competency

⬇

Develop a **SHARED MENTAL MODEL**

⬇

Achieve evidence-based standardization
and calibration

Holmboe ES ABIM 2010

Benefits of Performance Dimension Training

➢**Direct observation**
  ➢↑ Standardized, broad, systematic approach
  ➢↑ Attentiveness to interpersonal/communication skills
➢**Feedback**
  ➢↑ Breadth of skills discussed using more granular vocab
  ➢↑ Self-efficacy giving specific, constructive feedback
  ➢↑ Ability to deconstruct holistic assessments
➢**Faculty clinical skills**
  ➢ Acquisition of new knowledge

Kogan JR et al. Faculty Experience of Direct Observation Training for Workplace Based
Assessment, submitted 2014

Synthesis to Judgment

➢Goal: Improve the quality and accuracy of the
educational "judgment" using a compare and contrast
process

Steps: Synthesis to Judgment

➢Review vignettes of different performance levels
➢Judge using behaviorally-based frameworks
  (e.g. evidence based frame of reference)
➢Trainer provides feedback on assessment
  accuracy
➢Discuss discrepancies between scripted
  performance and participants' assessments

Satisfactory Compared to What?

Compared to	Frame of reference
What I do	Self
What resident at similar PGY level does	Normative
Readiness for independent practice	Criterion referenced

Kogan JR et al. Med Educ.

What is Needed by the Patient

Novice — Advanced Beginner — Competent — Proficient — Expert/Master

Time, Practice, Experience

Dreyfus SE and Dreyfus HL. A 1980
Carraccio CL et al. Acad Med 2008;83:761-7

Perelman School of Medicine University of Pennsylvania

Accreditation Council for Graduate Medical Education   **49**

Entrustment as Assessment Construct

➢Cognitively aligned scale resonates with raters' experience
➢Increases discrimination
➢Reduces disagreement
➢Reduces # assessments for good reliability (generalizability coefficient 0.7)
  ➢Mini-cex: 6->3
➢Decreases assessor workload approx 50%

Crossley J et al. Med Educ 2011;45:560-9.

Perelman School of Medicine University of Pennsylvania

Accreditation Council for Graduate Medical Education   **50**

Shifting to Entrustment as FoR

➢Define competency based assessment
➢Define competence
➢Performance dimension training
➢Peer support/group discussions
➢Social and cultural factors

Kogan JR et al. Faculty Experience of Direct Observation Training for Workplace Based Assessment, submitted 2014

Perelman School of Medicine University of Pennsylvania

Accreditation Council for Graduate Medical Education   **51**

Lessons in Rater Cognition

➢ **Assessment (rater cognition) is a complex process**
  ▪ Training can help, but will not solve "all problems"
    ▪ Clarity on outcomes
    ▪ Shared mental models
    ▪ Own clinical skills matter
  ▪ Sampling remains essential
    ▪ Multiple raters in multiple setting
  ▪ Not all variation is bad, but not limitless
    ▪ Variation is a bounded condition

Perelman
© 2014 Accreditation Council for Graduate Medical Education   **52**

Questions

Perelman School of Medicine University of Pennsylvania

Accreditation Council for Graduate Medical Education   **53**

Page 9