



American Board  
of Internal Medicine

---

***“You can observe a lot just by  
watchin’.”***

**Yogi Berra**

# Workshop Objectives

- Review current state of
  - Physician clinical skills
  - Faculty evaluation skills
- Understand the importance of direct observation *by faculty* for assessment of clinical skills

# Workshop Objectives

- Discuss practical strategies for *focused* direct observation
- Review rater training methods
  - Direct Observation of Competence (DOC) training

# Workshop Elements

- Mini-Lectures:
  - State of clinical skills
  - Quality of faculty ratings
- Direct observation exercises
  - Performance dimension exercise
  - Videotape review and rating

# Basic Premises

Accurate evaluation important:

- Summative: “high stakes” decisions
  - Competent versus not competent
- Formative: feedback & professional development
- Professional obligation
  - Public accountability
  - Credible professional self-regulation
  - Faculty observation part of this obligation

# Key Basic Clinical Skills

- Medical interviewing
- Physical examinations
- Informed decision making/counseling
- Clinical judgment/reasoning
- Reflective practice
  - Self-directed learning
  - Professional growth and improvement
  - Learn from errors

# Are Clinical Skills Important?

*Where do clinical skills fall into the hierarchy of physician competencies and mastery in an era of advanced technology?*

# Importance of Sound Clinical Skills

- Diagnostic errors
  - Bordage: Inaccurate/ incomplete medical interview one of leading causes
  - Graber: Faulty synthesis common factor in adverse events (hospital setting)
- Patient satisfaction
  - Higher with better communication skills
- Patient self-care and activation
  - Better adherence and outcomes associated with better physician communication skills



# Diagnosis and Medical Interview

- Hampton (*BMJ*, 1975):
  - Medical interview: 82%
  - Physical exam: 9%
  - Laboratory: 9%
    - Study findings replicated by Peterson (1992)
- Kirch (*Medicine*, 1996)
  - Autopsy study over four decades
  - “Conclusive” information for diagnosis



# Results: Kirch and Scaffi

Diagnostic Procedure	Application Rate	Conclusive Information	Misleading Information
History	96%	73%	<1%
Physical examination	95%	62%	2%
Standard lab tests*	90%	22%	2%
Imaging techniques	72%	35%	7%
Electrocardiogram	71%	23%	4%
Microbiological tests	17%	18%	3%
Histology and cytology	8%	28%	2%

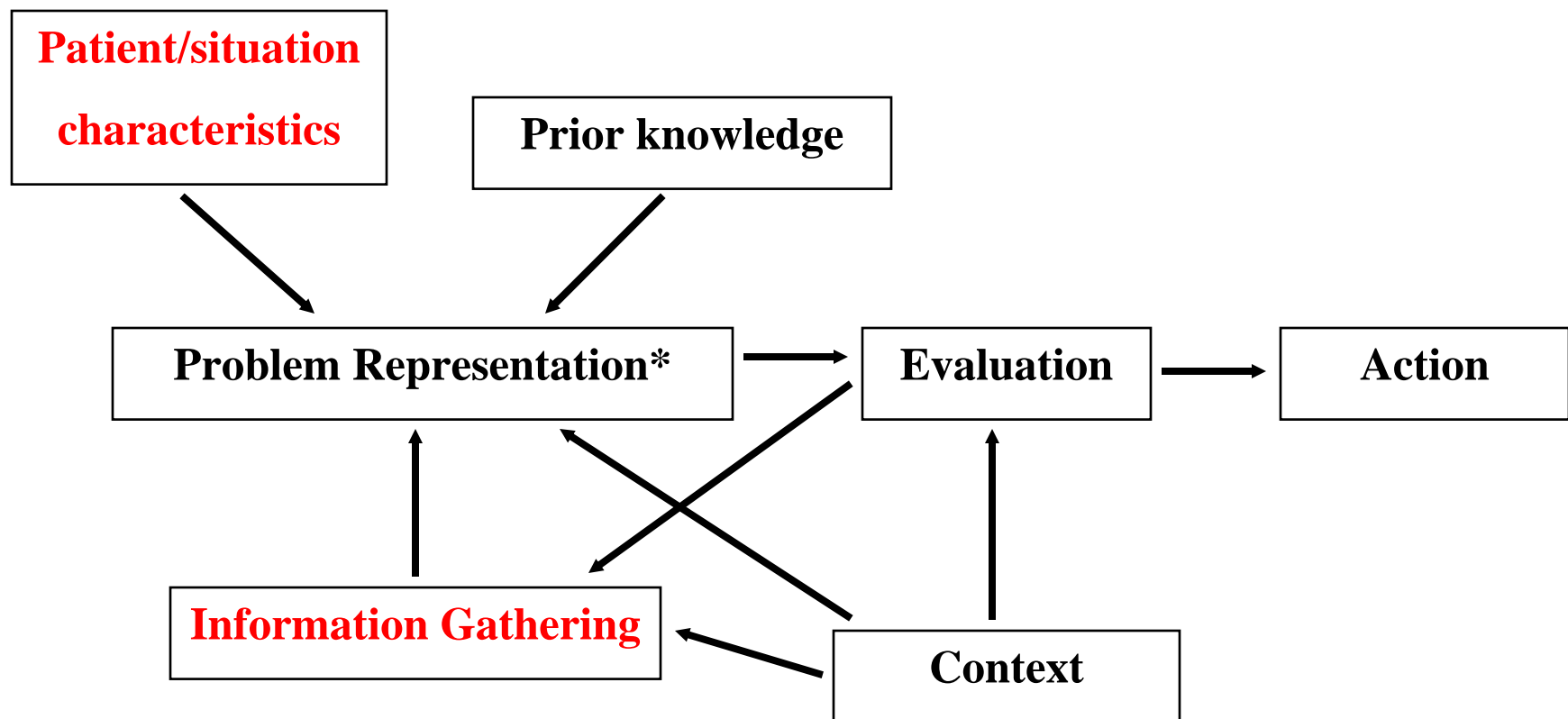
\* Serum chemistry, blood cell count, urinalysis.



American Board  
of Internal Medicine

1.800.441.ABIM | [www.abim.org](http://www.abim.org)

# Clinical Reasoning: A Primer



Gruppen and Frohna, International Handbook on Research, 2002

# Clinical Skills: Trainees

- Stillman (*Ann Intern Med*; 1990)
  - Wide variability in MS4 clinical skills near graduation
- Sachdeva (*Arch Surg*; 1995)
  - Wide variability in surgery intern skills
- Lypson (*Acad Med*; 2004)
  - Wide variability of basic skills *needed to start* internship among new interns at University of Michigan on “entry” OSCE

# State of Clinical Skills: Trainees

## Vukanovic-Criley (*Arch Intern Med*; 2006)

- Study of cardiac examination skills
  - Computer-based, 50 question exam
    - Integrated visual and auditory skills
  - Med students (Y1-4), FP and IM residents, full time faculty, volunteer faculty, cardiac fellows (N=860)
- Results
  - No improvement after MS3 year except cardiac fellows
    - Range of mean scores:
      - 58.5% (MS3) – 60.2% (Fac)
      - Cardiac fellows: 71.75%

# Clinical Skills: Practicing MDs

- Ramsey (*Am J Med*; 1998)
  - Incomplete history-taking / preventive health screening
- Braddock (*JAMA*; 1999)
  - Internal Medicine, Family Medicine, Surgery
  - 1058 patient visits: only 9% fulfilled core elements of informed decision making

# Importance of Faculty

- Reilly Study (*Lancet*, 2003)
  - Reviewed 100 consecutive admissions to GIM service
  - Investigated the number of physical exam findings that:
    - Changed patient management
    - Changed diagnosis
    - “Incidental” findings not counted

# Results: Reilly Study

- Faculty detected 26 PExam findings missed by residents that *changed* patient diagnosis and/or management
  - Essentially *1 in every 4* admissions to the General Medicine teaching service

# Importance of Faculty

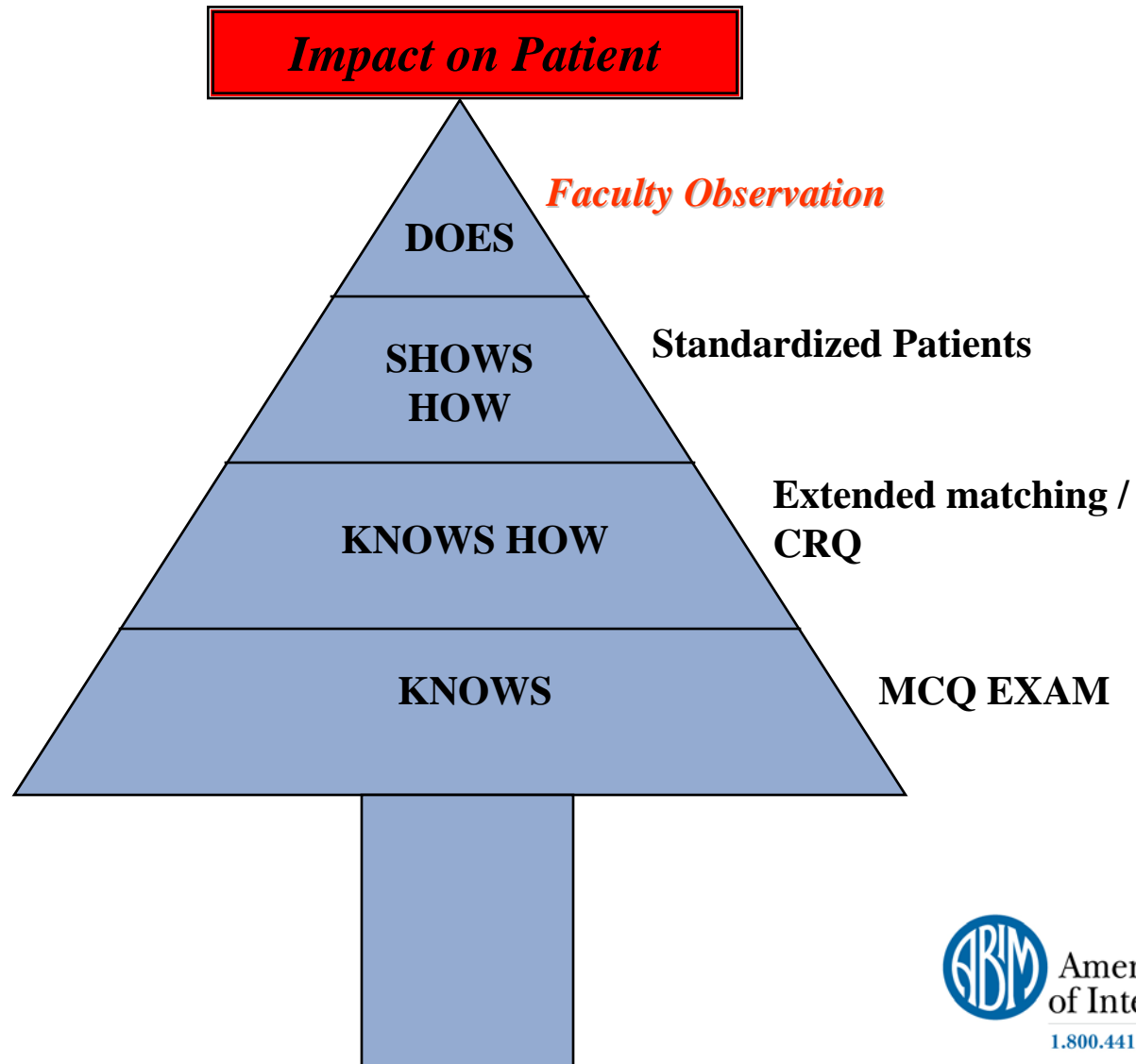
- Wisconsin and USUHS Clinic Studies
  - Faculty assessment disagreed with that of house officer in up to 30% of patients
  - “Severity” of discordance varied between residents and between studies
- What has been your own experience?

# Effective Supervision

AMEE Guide (Kilminster; *Med Teach*, 2007)

- Empirical evidence supports:
  - Direct supervision helps trainees gain skills faster, and behavior changes more quickly
  - Quality of relationship affects effectiveness of supervision
    - Continuity and reflection
  - Self supervision not effective
  - Better supervision associated with improved patient safety and quality of care.

# Miller's Assessment Pyramid



# Clinical Skills: Themes

- Deficiencies exist across continuum
- Not detected by other evaluation methods
  - Performance of basic clinic skills does not correlate with performance in other dimensions of competence

# Clinical Skills: Themes

- Students and Residents:
  - Aware of importance
  - Recognize under-emphasis in training
  - *Desire* better supervision
- Without detection deficiencies in clinical skills cannot be corrected
- Patient safety and quality of care at stake

# Faculty Observation / Rating Skills

Rating scale studies:

- Significant rater errors in use of scales
  - Halo effect (correlational)
    - Ratings based mostly on *perceived* knowledge and personality
  - Leniency error (distributional)
    - Failure to lower ends of scale (“Doves”)

# Faculty Observation / Rating Skills

- Herbers (1989) / Noel (1992)
  - Structured > open-ended form in detection of errors on a videotaped scripted to be marginal
  - Brief training video not effective
  - Increased accuracy  $\neq$  discriminative ability
- Kroboth (1992)
  - Poor inter-rater reliability
  - Rater training ineffective

# Faculty as Raters – Key Issues

- Lack of observation of the actual skill(s) being assessed
  - Monthly evaluation form
- Faculty ratings:
  - Lack reliability (both intra- and inter-rater)
  - Often inaccurate
  - Possess weak validity
- Content and context specificity
  - Many faculty lack competence in the very skill(s) they are asked to assess

# Faculty as Raters - Solutions

- Step 1: Getting faculty to observe
  - *Required* by “accrediting” bodies
    - US (ACGME) and UK (Foundation program)
  - Focused observations are logistically possible
    - 5 to 10 minute observations *are* valuable
  - Build on faculty “surprises”
    - The “You will not believe what I saw today” experience
  - Provide “usable tool”

# Usable Tools

- MiniCEX
  - Simple, portable, reliable 7 category form
  - Evidenced-based; 3 large scale reliability and feasibility studies (2 US; 1 in UK)
    - Core assessment tool: UK Foundation program
- Checklists for medical interview
  - Calgary-Cambridge
  - SEGUE
  - Kalamazoo Consensus

# Opportunities: Inpatient

- Hospital ward
  - Communication and PExam on daily rounds
  - Discharge instructions to a patient
  - Informed consent discussions
  - *Part* of admission work-up
- Emergency Department
  - Excellent for acute care: multiple interactions
  - Admission work-ups

# Logistics: Outpatient Clinic

- One mini-CEX per trainee per day per week
  - One attending observes portion of first visit of the day
  - Minimizes disruption of clinic
  - Perform over course of *academic year*
  - *Easy to obtain 6-8 Mini-CEX's per year per trainee*

# The Patient Encounter

- Sampling “parts” of the encounter:

<b>INTERVIEW</b>	<b>PHYSICAL EXAM</b>	<b>COUNSELING</b>
------------------	--------------------------	-------------------

Key Message: EMBED Observation in what you are already doing as part of clinical care and supervision

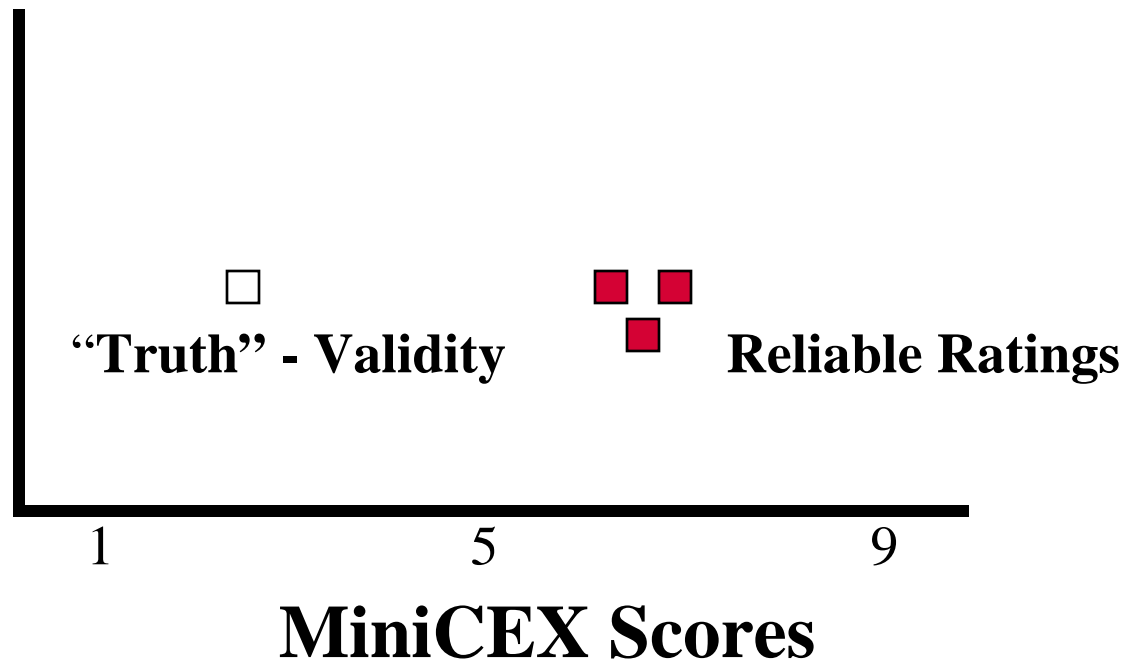
# Faculty as Raters - Solutions

- Step 2: Improving reliability
  - Multiple brief observations
  - Perform over time: outpatient setting allows for longitudinal observation
  - Involve multiple faculty
  - MiniCEX: sufficient reliability for pass/fail determinations after just 4 observations

# Faculty as Raters - Solutions

- Step 3: Improve accuracy and validity
  - Most difficult step
  - Improved with structured rating forms
  - Can be improved with rater training, but:
    - Brief training interventions (e.g. 15-30 minutes) do not work
    - Will need longitudinal approaches to FD
      - “Drip” versus “bolus” method

# Reliability versus Validity



# Does Faculty Training Work?

## Performance Appraisal Literature:

- Can reduce rating errors
- Can improve discriminative ability
- Can improve accuracy

# Approaches to Faculty Training

Empirically studied training methods:

- Behavioral Observation Training
- Performance Dimension Training
- Frame of Reference Training
- Direct Observation of Competence Training

# Behavioral Observation Training

Three main strategies:

- Increase the amount of “sampling”
  - More observations lead to more accurate evaluations (“practice makes perfect”)
- Use of observational “aides”
  - Behavioral diary to record observed performance.
  - Checklists

# Videotape Exercise

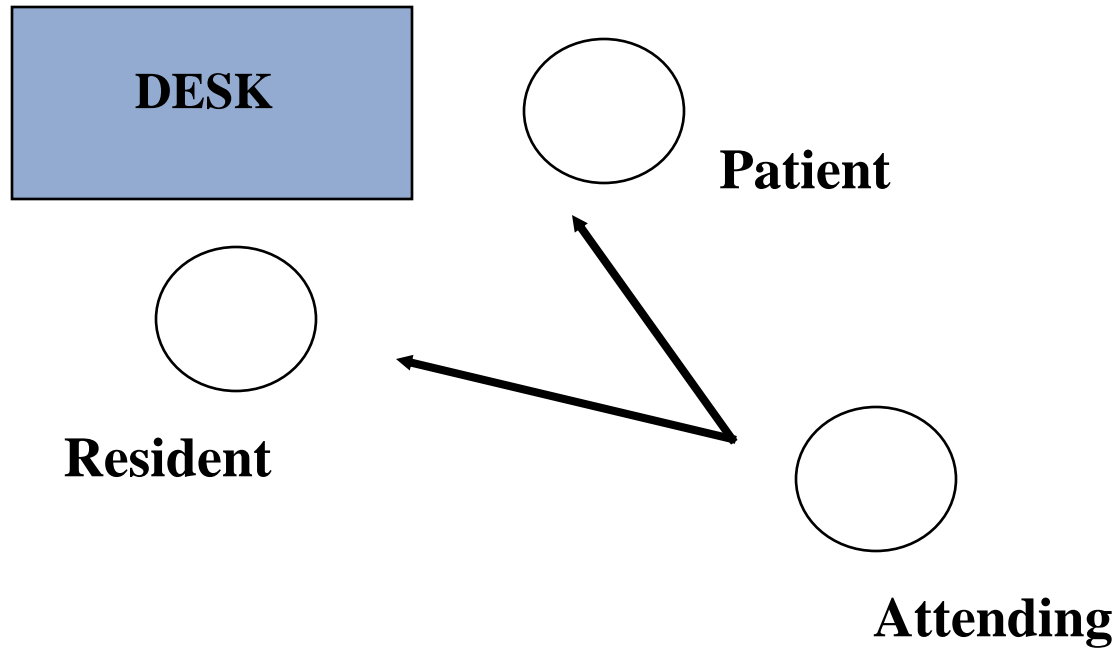
Situation: An attending is performing a miniCEX of a house officer performing a physical exam.

- Questions:
  - How well did this attending evaluate the house officer?
  - How was the resident-patient interaction affected?

# BOT: Structuring the Observation

- Prepare for the observation
  - Faculty: Know what you're looking for
  - Resident: Let them know what to expect
  - Patient: Let them know why you are there
- Minimize intrusiveness – correct positioning
- Minimize interference with the house officer-patient interaction
- Avoid distractions

# Principal of Triangulation



# Performance Dimension Training

- Method to help faculty identify and learn the specific dimensions of a competency
- Involves discussion of the “qualifications” required for each dimension of that competency
- Use specific definitions and criteria of competencies to “calibrate” faculty
  - Goal: get faculty on the “same page”

# PDT Exercise

In your small group, discuss what should be the components of an effective counseling session for a patient starting a new medication

# Frame of Reference Training

- Goal is to improve “judgment” and accuracy

## Steps in FOR training:

1. Raters given descriptions of each dimension - discuss “qualifications” needed for each dimension (PDT)
2. Review of clinical vignettes describing critical incidents of performance: unsatisfactory to average to superior

# Frame of Reference Training

Steps for FoRT (continued):

3. Raters used vignettes to then provide ratings on a behaviorally anchored rating scale (BARS) - think typical eval form
4. Session trainer provides feedback on what “true” ratings should be along with rationale
5. Discussion ensues about discrepancies between trainer’s ratings and the participants’ ratings

# Frame of Reference Training

- Most difficult aspect of FOR:
  - Setting the actual performance standards
  - Reaching agreement and consensus among teaching faculty
- Use Evidence whenever possible
  - E.g. Communication sciences

# DOC Training

- Combination of:
  - Rater error training
  - Performance dimension training
  - Frame of reference training
  - “Live” practice in observation with standardized residents/patients
    - Individual evaluation and feedback
    - Group debrief with Eval and FB

# DoC Training: Evidence

- DoC Training improves:
  - Rater stringency
    - Replicated in Canadian study (unpublished)
    - ↑ Accuracy in identifying unsatisfactory performance
  - Faculty comfort in performing observations

# Direct Observation: Challenges

- Like all skills, requires training and practice
- Faculty “calibration” important
  - Agreeing on “metrics” of performance
  - Faculty comfort with own skills
- Faculty training
  - Brief interventions mostly ineffective

# Observation: Helpful Hints

- Sample “parts” of the visit:
  - History-taking
  - Physical examination
  - Counseling
- Perform longitudinally
  - No need to do it all at once
  - Embed it in what you do
- Agree on performance criteria with faculty

# Summary

- Basic clinical skills are important: so is the need to observe them!
- Observation is a complex skill that requires training and practice
- Direct observation by educators will remain a critical component of both evaluation and feedback

# *Questions*