The Evolution of Workplace-Based Assessment

John Norcini, Ph.D.
John.Norcini@gmail.com

Overview

Past	• Empasis on measurement
Present	• Emphasis on learning
Future	• Emphasis on implementation

The past

"That part of eternity with some small fraction of which we have a slight and regrettable acquaintance."

A Bierce

The Devil's Dictionary

- Origin of the mini-CEX
  - From 1936 to 1970, the ABIM required an oral exam
    - Two long cases (inpatients), two examiners
  - Discontinued in 1970 mainly because of feasibility
    - But there were other problems as well…

The origin of the Mini-CEX

Candidates' fate depended in part on which examiners they drew

'Black' Jack Myers, MD

The origin of the Mini-CEX

Candidates' fate depended in part on which patients they drew

The origin of the Mini-CEX

- In 1970, the ABIM delegated the decision about clinical skills to training program directors
  - Required a clinical evaluation exercise (CEX)
- By the late 1980s, trainees and program directors were grumbling
  - Hard to schedule and stressful for trainees

The origin of the Mini-CEX

- By then an understanding of what constituted good assessment was based in research
  - Once is not enough
  - One is not enough
  - Reality matters

Once is not enough

In the CEX, the trainee was evaluated with only one patient and physician performance varies considerably from patient to patient

"The tendency of the casual mind is to pick out or stumble upon a sample which supports or defies its prejudices, and then to make it the representative of a whole class."

Walter Lippmann

One is not enough

In the CEX, the trainee was evaluated by only one examiner and examiners differ in stringency

“Whenever people agree with me I always feel I must be wrong.”

Oscar Wilde

Reality matters

Most real physician-patient encounters are short and focused, often with returning patients, and in an outpatient setting so the task was artificial

“You can avoid reality, but you cannot avoid the consequences of avoiding reality."

Ayn Rand

Mini-CEX

- Assessor
  - Observed a brief patient-trainee encounter
  - Rated it (interviewing, counseling, PE…) and provided feedback
- Multiple encounters and examiners
- By the late 1990s, it was an alternative to the CEX

Methods

Faculty ratings

360º assessment

Chart-stimulated recall

Blinded patient encounters

Direct observation of procedural skills

And more…

Peer assessment

mini-CEX

Clinical encounter cards

Clinical work sampling

Methods

Faculty ratings

360º assessment

Chart-stimulated recall

Blinded patient encounters

Direct observation of procedural skills

And more…

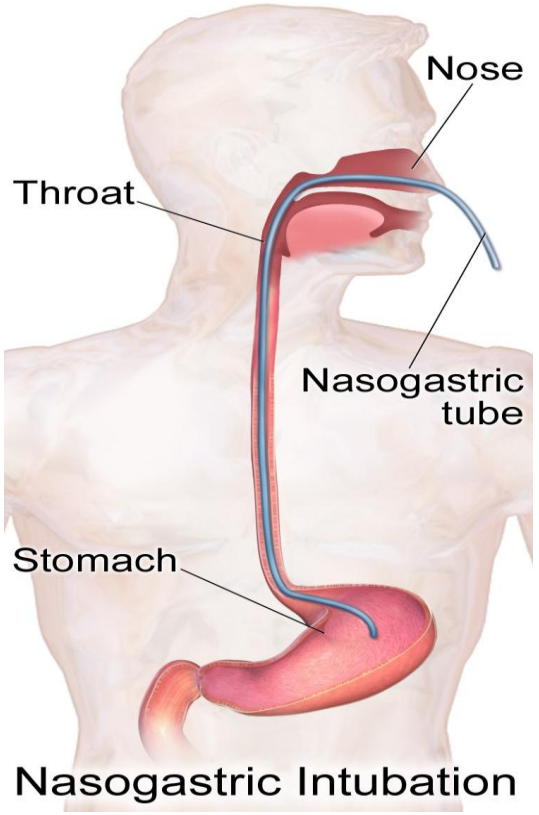
Peer assessment

mini-CEX

Clinical encounter cards

Clinical work sampling

Direct Observation of Procedural Skills



- Assessor observes a patient-trainee encounter
  - Procedure
- Assessor rates along several dimensions
- Assessor provides feedback
- Multiple encounters and examiners

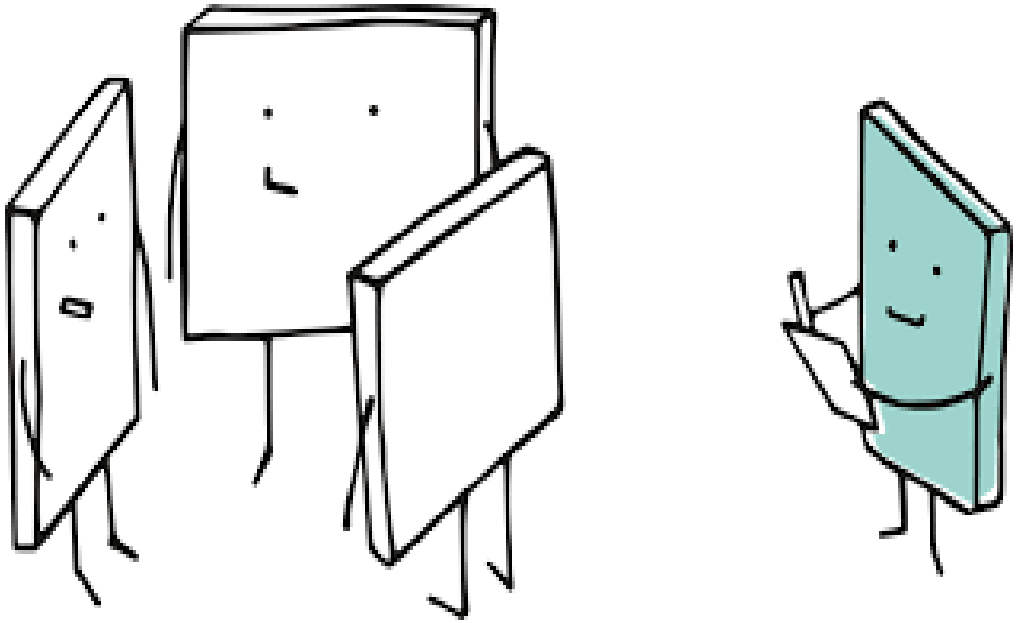
Chart-Stimulated Recall

- Assessor reviews a patient record where the trainee made notes
- Discussion around note quality and clinical reasoning
- Assessor rates along several dimensions
- Multiple encounters and examiners

Multisource Feedback

- Trainee nominates assessors and self-rates
- Assesses clinical and generic skills
  - Collated centrally
- Trainee given self-ratings, assessor ratings, mean ratings, and comments

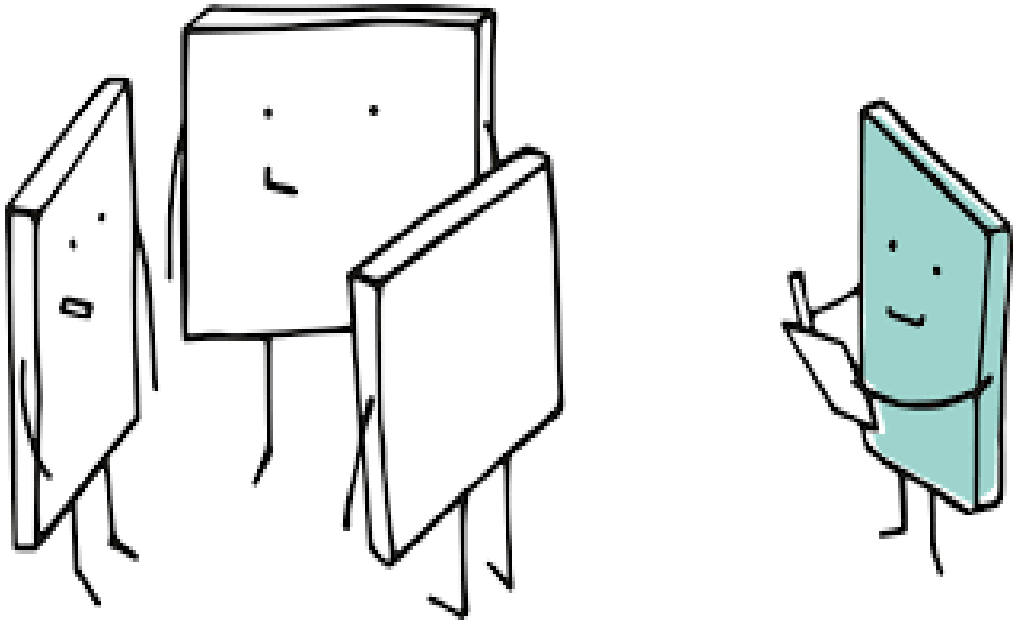
Common features: The basis for judgment



Trainee is observed in workplace activities in one of two ways, either over time (indirect-informal-unidirectional)

- Based on observation over multiple occasions
- Sometimes the behavior is not actually observed
- It is subject to general impressions

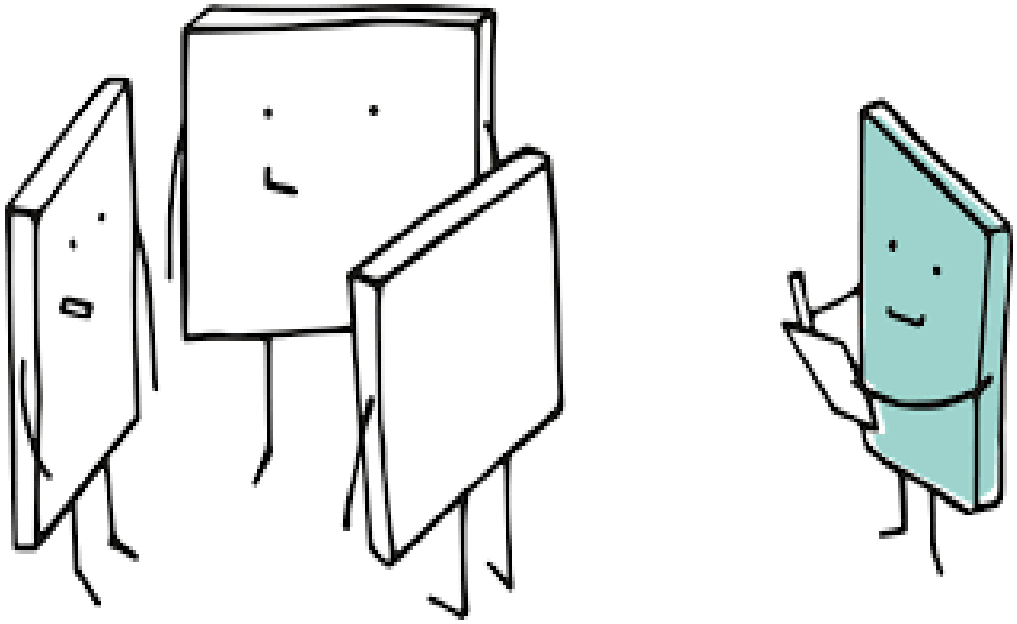
Common features: The basis for judgment



…or in specific encounters (direct-formal-bidirectional)

- Less subject to general impressions
- Multiple encounters are needed so it is time-consuming

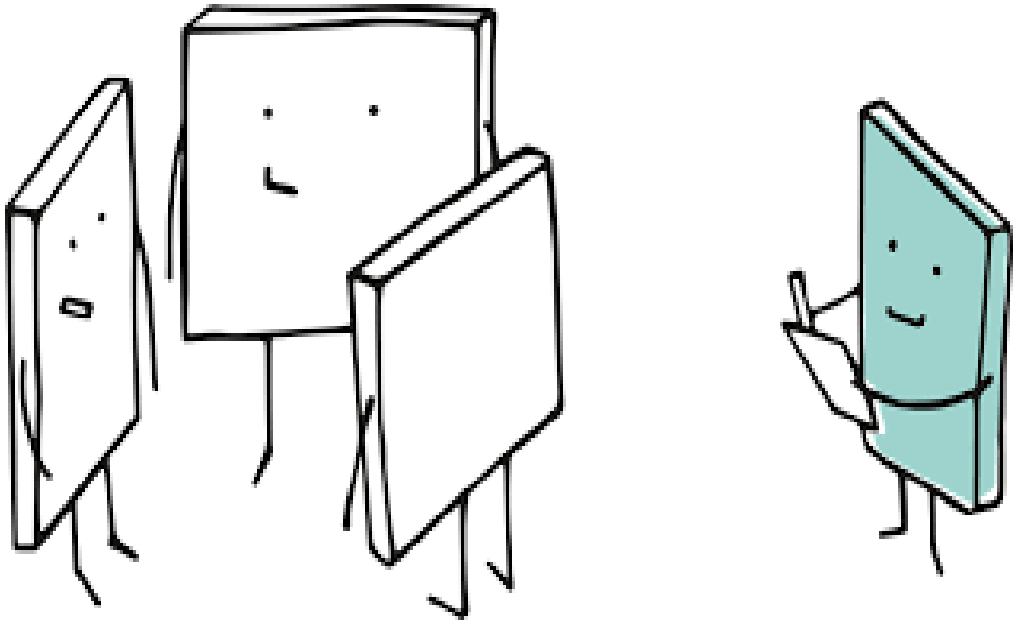
Common features: Capturing the performance



Observer judges the performance in one of two ways, either in terms of occurrence (checklist)

- Simply note if a behavior has occurred
- Non-experts can use checklists
- Provides guidance for feedback
- Misses some of the subtleties

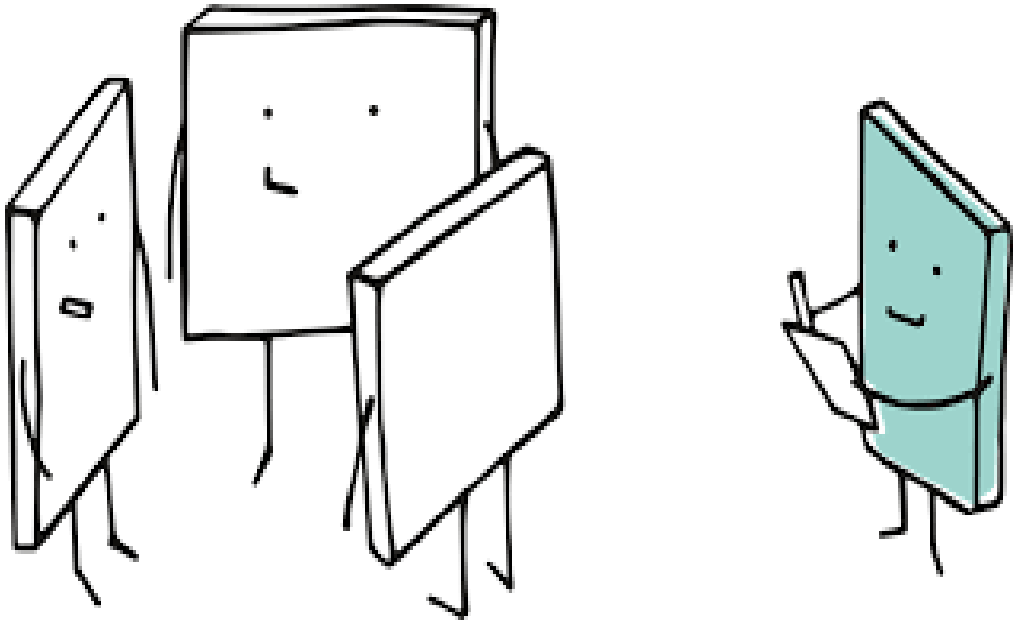
Common features: Capturing the performance



…or in terms of quality (global ratings)

- Evaluate the quality of a performance
- Requires an expert but allows use of judgment
- Highly correlated with checklists but slightly more valid and slightly less reliable

Common features



- After their assessment, the observer provides feedback
- Observer is the assessment device
  - Critical role in their quality
  - Faculty development is essential

The present

"That part of eternity dividing the domain of disappointment from the realm of hope"

A Bierce
The Devil's Dictionary

- As the mini-CEX was introduced, research about learning in the workplace increased
  - Observation was infrequent
  - Feedback was infrequent
  - Feedback has a major effect on learning
  - Spacing feedback leads to better performance
  - Assessment, by itself, creates learning

Observation is infrequent

- Capri et al., 2024
  - Clinical skills never observed by an attending
    - 34% for HX; 25% for PE
- Wang & Vegas, 2022
  - Quarter of students and residents not observed regularly
- Isaacson et al., 2014; Day et al., 1990
  - Been this way for years

Feedback is infrequent even when there is observation

- Shafian et al, 2024
  - Only 26% report receiving regular feedback
    - Less than half reported it was at suitable time, location, clear, and actionable
  - Only 32% reported that faculty had the skills to provide it effectively

Feedback is essential

- Feedback has a significant influence on achievement
  - General education (Hattie, 1999)
    - Meta-analysis of 12 meta-analyses
    - Feedback is among the largest influences on achievement (ES=.79)
  - Medical education
    - Feedback alone effective is effective in 71% of studies (Veloski et al., 2006)
    - Moderate to large effects (Johnson, Weerasuria, Keating, 2020)

Spacing optimizes learning

	Massed Training	**Spaced Training**
Sessions	**Few, Intense**	**Many, Spread Out**
Speed		
Confidence		
Satisfaction		
Retention		
Performance		

Spacing optimizes learning

	Massed Training	Spaced Training
Sessions	**Few, Intense**	**Many, Spread Out**
Speed	Faster	
Confidence	Higher	
Satisfaction	Greater	
Retention		Longer
Performance		Better

Retrieval practice

- Retrieval of information or a performance enhances learning
  - Students read a passage (Roediger & Karpicke, *Psych Science*, 2006)
    - Group 1 took three tests on the passage
    - Group 2 re-read the passage carefully three times
    - On a test one week later, Group 1 did better
  - Students read science text (Karpicke & Blunt, *Science*, 2011)

Current state of workplace assessment

- Measurement
  - Once is not enough
  - One is not enough
  - Reality matters
- Learning
  - Requires observation
  - Provides feedback
  - Creates learning
  - Optimizes learning

The future

"That period of time in which our affairs prosper, our friends are true and our happiness is assured"

A Bierce
The Devil's Dictionary

The greatest challenge for workplace assessment is implementation

Implementation challenges

- Logistics
- Misalignment around purpose
- Lack of a strong educational alliance between faculty and trainees
- Lack of credible bi-directional feedback

Young et al., 2020

Implementation science

- Implementation science offers a way of addressing this challenge
  - Focuses on how evidence-based interventions are adopted, integrated, and sustained in real-world settings
  - Applied to education (Price et al., 2015)
  - Several frameworks that provide a structure to guide research into practice

Implementation science

- Precede/Proceed model is one example that addresses three factors
  - Predisposing
  - Reinforcing
  - Enabling
- Model has been used with success in CME

Some predisposing factors

- Committed examiners
- Longitudinal supervisor-trainee relationships
- Examiners who have the competencies they assess
- Selective use of patients and other healthcare professionals

Some reinforcing factors

- Training
  - Not ‘standardization’
    - Rule out bad reasons for disagreement
    - Teach examiners how to provide feedback
- Monitor and provide feedback
  - Anonymized comparisons among examiners
  - Feedback from trainees
  - Reduce the roles of examiners who are ineffective

Some enabling factors

- Create a system of assessment
  - Clear values
  - Routine bi-directional encounters
  - Support from leadership
- Offer faculty incentives that matter
  - Protected time, financial reward…
  - Recognition perhaps including credentialling…

Summary

- Workplace assessment
  - For good measurement, faculty is the key and more than one is important
  - Strength of workplace assessment is its support and creation of learning
  - Implementation is a challenge that can be addressed from an implementation science perspective