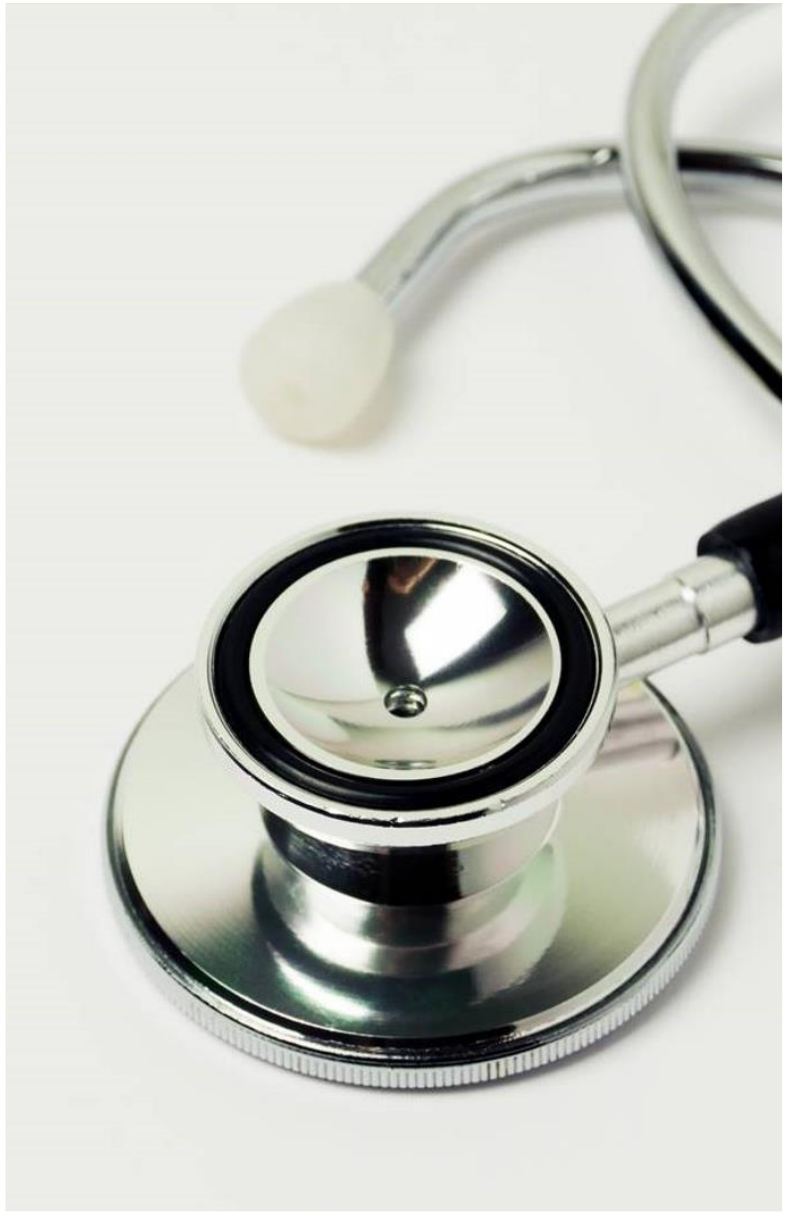
EXPLORING
PHYSICIAN ASSESSMENT AND FITNESS FOR DUTY EVALUATIONS

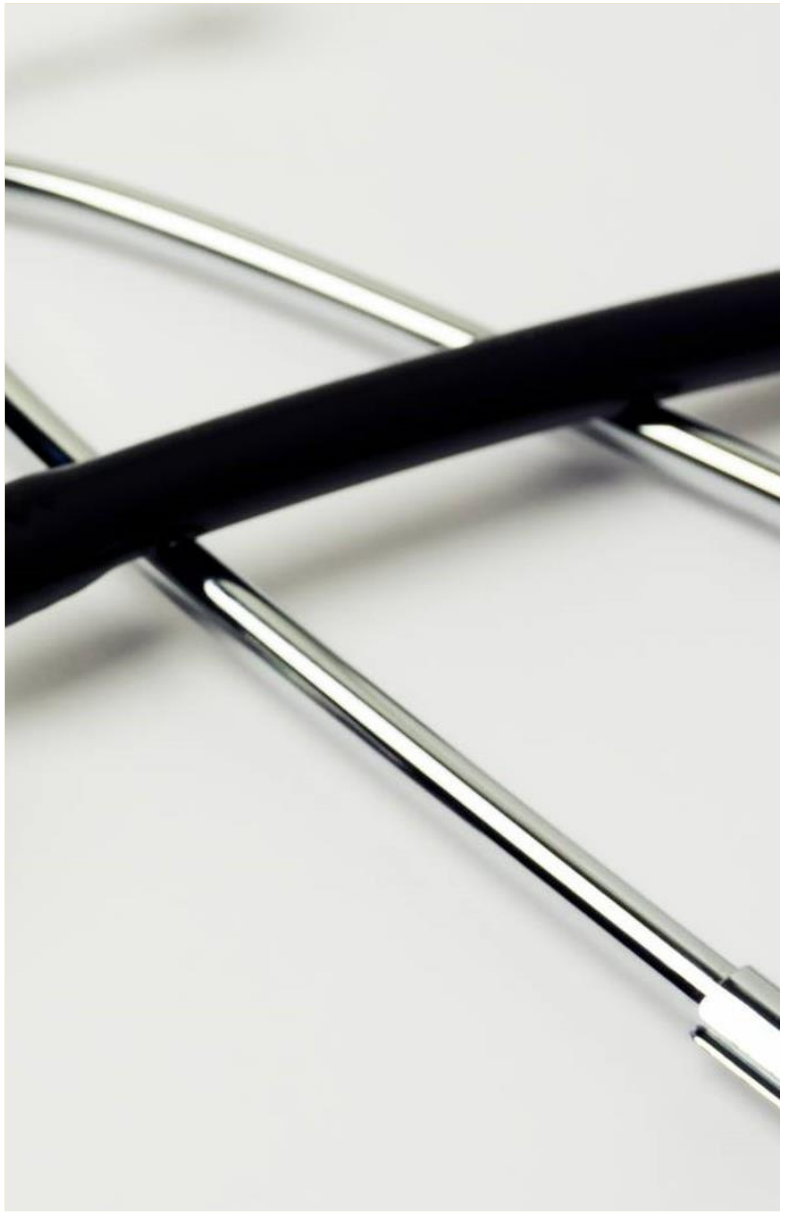
OBJECTIVES

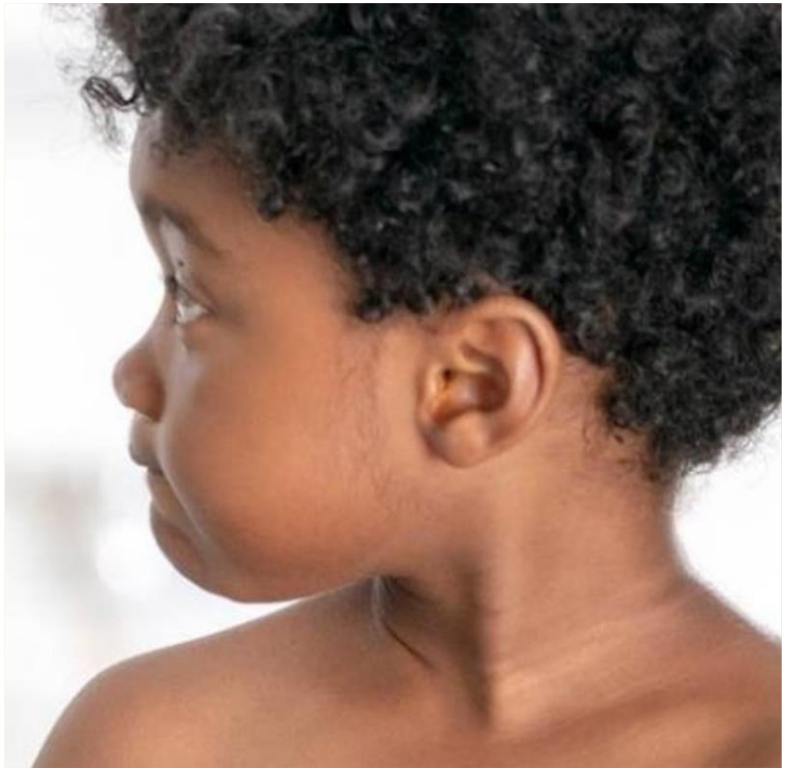
Review the testing components of Physician Assessments.

Review approaches to Fitness for Duty evaluations.

Review and discuss two cases that pertain the above topics.







PART 1: PHYSICIAN ASSESSMENT/CLINICAL PERFORMANCE TESTING

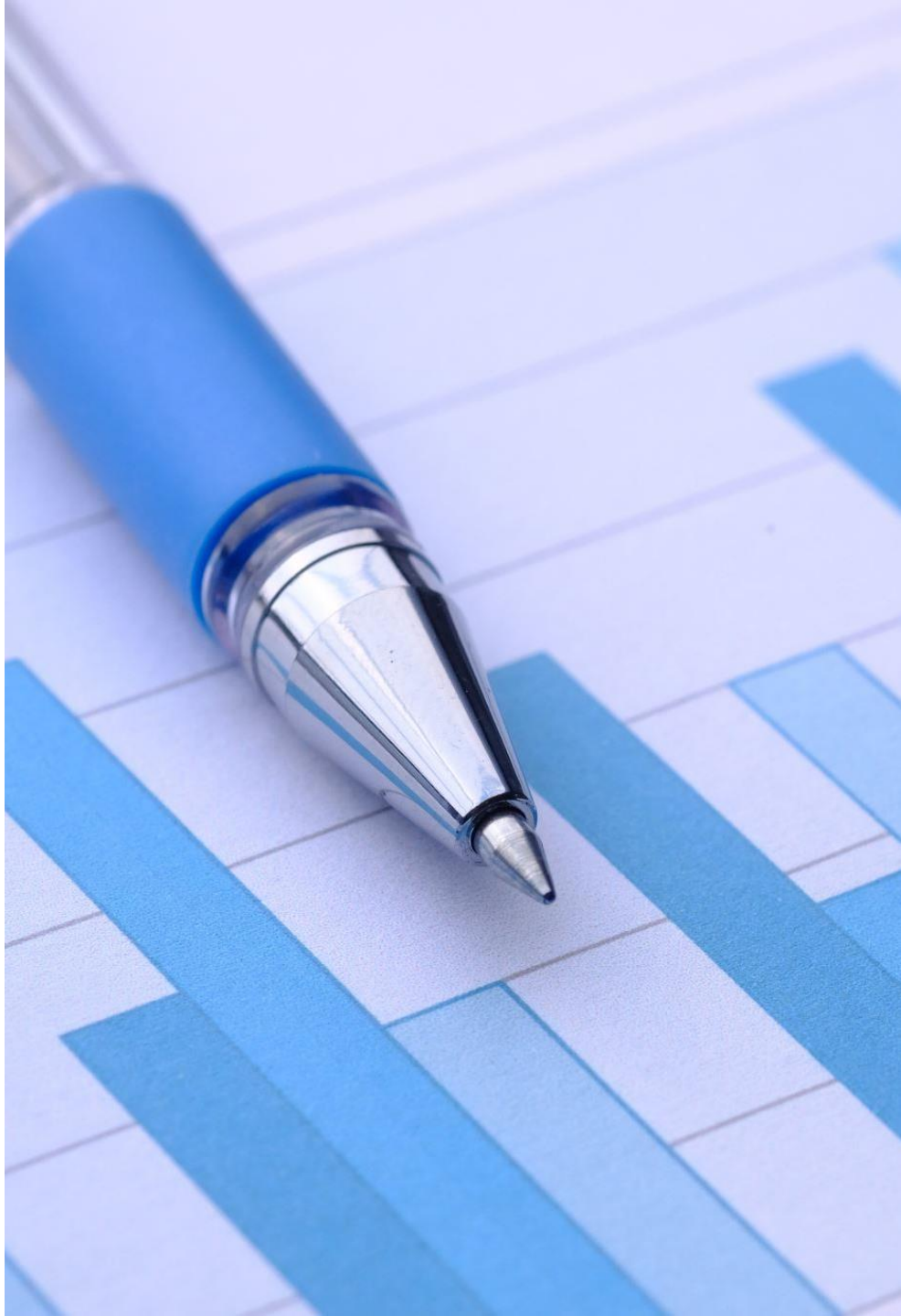
WHAT DO WE WANT TO MEASURE, AND WHY?

- Medical knowledge?
- Clinical skills?
- Intelligence? Cognitive abilities?
- Communication skills?
- Problem-solving skills?

→ **Clinical Performance**

HOW DO WE EVALUATE CLINICAL PERFORMANCE?

- Multiple individual tests
- The tests examine performance from different perspectives
- ACGME Core Competencies serve as a framework for testing (more to come on that…!)
- The data generated is pooled, analyzed, and graded/rated
- All the above make up the assessment (or performance evaluation)



EVALUATION CATEGORIES – PHYSICIAN ASSESSMENT

(I) High stakes

(II) Routine, at intervals

(III) Spot checks

MOST MEDICAL BOARD OR FACILITY-REQUIRED PHYSICIAN ASSESSMENT TESTING IS **HIGH STAKES** TESTING

- Physician's license to practice/career is on the line
- Ability to practice independently and in an unsupervised manner may be threatened
- Practice restrictions can have negative effects on practice
- Reputation is in jeopardy
- Patient safety can be at risk
- A physician may have hidden or undiscovered gaps in competence or health/cognition

QUESTION: SO… WHAT IT THE BEST TEST FOR HIGH STAKES PHYSICIAN COMPETENCY TESTING?

???????

THERE IS NO SINGLE BEST TEST FOR HIGH STAKES PHYSICIAN ASSESSMENT

High stakes testing requires carefully selected testing to get the BEST answers.

No single test can give us everything; multiple and overlapping tests are needed.

QUESTION: SO, HOW MANY TESTS DO I NEED?

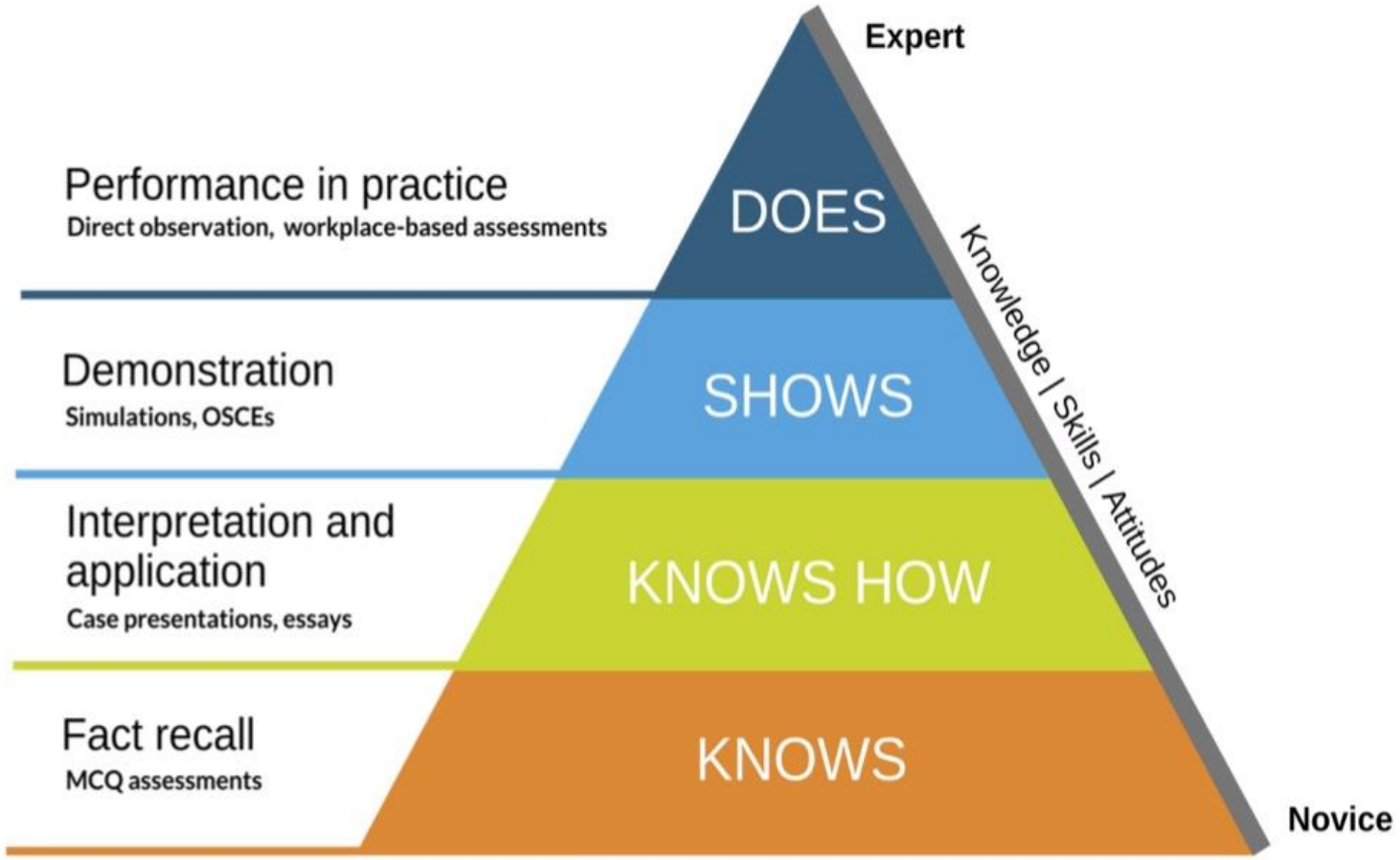
MORE THAN **ONE** or **TWO**

(more to come on this…)

ACCREDITATION COUNCIL FOR GRADUATE MEDICAL EDUCATION (ACGME) CORE COMPETENCIES & EASE OF MEASURING

- Medical Knowledge
- Patient Care
- Interpersonal Communication Skills
- Professionalism
- Systems-Based Practice
- Practice-Based Learning and Improvement

MILLER'S PYRAMID



SELECTION: TESTING MULTIPLE AREAS OF THE PYRAMID

OPERATING SYSTEM

DATABASE

TEST CRITERIA: VALIDITY AND INTERRATER RELIABILITY

Validity: Am I testing what I want to test?

- Using tests that apply to scope of training and/or practice

Interrater Reliability: The reproducibility of measurement between two or more investigators.

- Using trained and calibrated assessors, and probably at least 3 of them

KSTAR INTER-RATER RELIABILITY STUDY

Evaluation of a High Stakes Physician Competency Assessment: Lessons for Assessor Training, Program Accountability, and Continuous Improvement

Elizabeth F Wenghofer [1], Robert S Steele, Richard G Christiansen, Misti H Carter

Journal of Continuing Education in the Health Professions 2021 Apr 1;41(2):111-118.

The 2021 Paul Mazmanian Journal of Continuing Education in the Health Professions Award for Excellence in Research

KSTAR INTER RATER RELIABILITY STUDY

Results

- Intraclass correlation coefficients ranged between **0.756 and 0.876** for all components scored and was highest for overall performance.
- Regression indicated that individual component scores were positively associated with overall performance.
- Levels of variation in component scores were significantly different across quartile ranges with higher variability in poorer performers.

KSTAR INTER RATER RELIABILITY STUDY: DISCUSSION

High-stake assessments (done in a high stakes manner) can be conducted reliably and identify performance gaps of potentially dyscompetent physicians.

Physicians who performed well tended to do so in all aspects evaluated, whereas those who performed poorly demonstrated areas of strength and weakness.

Understanding that dyscompetence rarely means a complete or catastrophic lapse of competence is vital to understanding how educational needs change through a physician's career.

CRITERIA FOR A GOOD TEST

Usability

Acceptability

Adequacy

Purpose

Economy

Meaningfulness of Test Score

Comparability

Availability

CRITERIA FOR A GOOD ASSESSMENT

Validity or Coherence: There is a body of evidence that is coherent ("hangs together") and that supports the use of the results of an assessment for a particular purpose.

Reproducibility or Consistency: The results of the assessment would be the same if repeated under similar circumstances

Equivalence: The same asmt yields equivalent scores or decisions when administered across different institutions or cycles of testing

Feasibility: The asmt is practical, realistic, and sensible, given the circumstances and context

Educational Effect: The asmt motivates those who take it to prepare in fashion that has educational benefit

Catalytic Effect: The asmt provides results and feedback in a fashion that creates, enhances, and supports education; it drives future learning forward

Acceptability: Stakeholders find the asmt process and results to be credible.

Norcini J, Anderson B, Bollela V, et al. Criteria for good assessment: Concensus statement and recommendations from the Ottawa 2010 Conference. Medical Teacher 2011. 33:206-214.

TESTS USED COMMONLY USED IN PHYSICIAN ASSESSMENT

- Multiple Choice Question Tests
- Chart-Derived Testing:: Chart/Reviews/Presentations/Discussions

  -Chart Stimulated Recall

  -Peer Reviewed Chart Presentations
- Oral examinations
- Standardized Patient Encounters and Review/Objective Structure Clinical Exams (OSCEs)
- Cognitive Screening Tests
- Mini-CEX and other tests

MULTIPLE CHOICE QUESTION EXAMS **SPEX®** (SPECIAL PURPOSE EXAM)

- A joint venture of the FSMB and NBME; used by medical boards
- Computerized MCQ examination of medical knowledge pertaining to the **undifferentiated practice of medicine**
- The exam has a suggested cut point: 75% more correct; need to be aware of implications
- Has strong psychometrics
- Gives general feedback (disease categories, age groups, etc.)
- 5-hour test, ~200 questions, $1400
- KNOWS test

MULTIPLE CHOICE QUESTION EXAMS **POST-LICENSURE ASSESSMENT SYSTEM (PLAS)**

- A joint program by FSMB and NBME; they are used by **national assessment programs** to assist them in conducting comprehensive, tailored assessments of physicians' medical knowledge, clinical judgment and patient management skills.
- These exams are **more specific and include specialty tests** such as family medicine, general surgery, neurology; there are a few specific topic tests
- Tests results the given are clustered into content areas, allowing for more depth in detail of feedback
- Most exams are 2-3 hours long, usually have 90-100 questions, test cost is approximately $100
- Most are KNOWS or KNOWS HOW tests

STANDARDIZED PATIENTS ENCOUNTERS

- Actor is trained to assume the role of an ill patient, including history, reported physical finding
- The physician being exam interviews the patient, examines them, explains initial impressions, and explains next steps; takes place in a simulated clinic or simulation center
- Domains tested: communication, medical knowledge, clinical decision-making, mind organization
- Discussion of cases with the examinee afterward enhances data obtained for the assessment
- Very rich source of information, especially in the SHOWS zone of Miller's pyramid, flexible regarding personalization
- Can be hard for examinee to suspend disbelief of the process, expensive to set up and run
- This is a SHOWS test

OBJECTIVE STRUCTURED CLINIC EXAM (OSCE)

- Often used with medical students
- Student rotate through a series of stations – several live patients with a single medical complaint, identification of heart sounds from machine, writing a progress note, etc.
- Scenarios are usually focused, sometimes used as a part of a final exam
- Can be used with residents and practicing physicians, but standardized patients yield more information in the "shows" realm
- This is a KNOWS HOW to SHOWS realm test

ORAL EXAMINATIONS

- A series of clinical scenarios that pertain to the physician's scope of practice used
- Criteria are set for history taking, verbalize clinical findings, clinical rationale and proposed treatments
- Often includes items that must be identified and/or managed appropriately, or the physician fails the scenario
- One or two examiners are usually present for the interview
- Lower cost compared to other testing modalities, relatively easy to create and administer
- Disadvantages: assessor needs to be well-trained to do this, oral examination can be more anxiety-provoking and obtrusive than other testing modalities; if only one assessor doing this, test reliability is not measurable

CHART-DERIVED TESTING: CHART STIMULATED RECALL (CSR) & PEER REVIEWED CHART PRESENTATIONS (PRCP)

- CSR: uses a **physician's own charts**
- PRCP: uses **another physician's redacted chart** (from peer review)
- Examinee and assessor both review charts (independently) prior to assessment
- Examinee presents the chart (single episode vs. longer care episode), presents the chart in a manner consistent with a good handoff of care, identifies pertinent findings, critiques the charts, makes suggestion re: quality improvement and system issues
- CSR demonstrated to have validity and good reliability (3-6 charts, ABEM study)
- PRCP demonstrated to have very good inter rater reliability; less obtrusive than other tests (KSTAR study – 5 charts); this is SHOWS test
- Cost is higher than some tests due to the need for review of multiple charts by assessor(s)

OTHER TESTS: MINI-CEX (MINI-CLINICAL EVALUATION EXERCISE)

- A form of direct observation assessment
- A workplace-based assessment tool that evaluates a trainee's performance during a consultation with a real patient.
- The mini-CEX is a concise and validated method of assessment that evaluates a trainee's skills in a number of areas, including History-taking, Clinical examination, Formulating management plans, Communicating with patients, and Professional and interpersonal skills.
- Relatively low-cost, validity under certain conditions has been established
- Usually one assessor, but there could be more if videotaped; can be time consuming
- This would be a DOES test

OTHER TESTS: COGNITIVE SCREENING

- Done by most assessment programs in the US to identify concern for cognitive impairment or inefficiencies
- MicroCog® (at facility, computer based, has a physician reference group)
- CNS Vial Signs® (at-home, computer based, tests executive functioning)
- Wisconsin Card Sorting Test, Rey-Osterrieth Figure Test, Trail Making Tests A and B (all test executive functions)
- Face to Face Screening with a neuropsychologist
- Other tests are being sought/investigated

ADDITIONAL TESTS

- Electrocardiogram and rhythm strip Interpretation
- Fetal monitoring strip interpretation with scenarios
- Simulated vaginal deliveries (SimMom®, etc.)
- Current pharmacotherapy
- Ethics and professionalism
- Dermatology test (from DermNet, New Zealand)

REVIEW USING MILLER'S PYRAMID

WHAT QUESTIONS DO YOU HAVE?

Thank you!

Rob Steele, MD, FAAFP

Rural and Community Health Institute

Texas A&M Health, Texas A&M University

rsteelefp@gmail.com

steele@tamu.edu

CASE #1: PHYSICIAN ASSESSMENT

Assessing a Physician Wanting to Return to Practice

Please Refer to Cases Handout

PART II:

FITNESS FOR DUTY EVALUATIONS FOR PHYSICIANS

PHYSICIAN HEALTH: "*DOCTOR, DON'T TREAT THYSELF*"

ROSVOLD, E. *DOCTOR, DON'T TREAT THYSELF*. PSNET/AHRQ, SEPT 2004. PSNET.AHRQ.GOV/

- 1 IN 3 PHYSICIANS DO NOT HAVE A PERSONAL PHYSICIAN
- MANY FIND IT DIFFICULT TO ASSUME PATIENT ROLE; 71% FEEL EMBARRASSED THEIR OWN DOCTOR
- OFTEN OPT FOR SELF-DIAGNOSIS AND/OR SELF TREATMENT
- INFORMAL HEALTH CARE OFTEN SOUGHT, INCLUDING CURBSIDE CONSULTS FROM COLLEAGUES
- WHEN SEEN BY A DOCTOR, MANY DESCRIBE THEIR SYMPTOMS IN THE MEDICAL TERMS THAT APPLY ONLY TO THE DIAGNOSIS ALREADY SELF-DESCRIBED

PHYSICIAN HEALTH: SOME CAUSES OF IMPAIRMENT/THREATS TO PATIENT SAFETY

- Alzheimer's dementia
- Parkinson's dementia
- Severe depression/anxiety/other behavioral disorders
- Cerebrovascular disease/stroke
- Dysexecutive syndrome
- Multiple uncontrolled medical comorbidities
- Metabolic disorders
- Burnout!

PHYSICIAN HEALTH: FINDINGS SUGGESTIVE OF COGNITIVE IMPAIRMENT

- Memory loss
- Reduced attention
- Decreased executive functioning
- Erosion of problem-solving abilities
- Increase in unsolicited patient complaints
- Increased medical errors
- Something's not right…

INDICATIONS OF POSSIBLE IMPAIRMENT "SOMETHING JUST ISN'T RIGHT…"

THE TIME BETWEEN NOTICING SOMETHING ISN'T RIGHT AND TAKING ACTION

KSTAR EXPERIENCE:

BETWEEN 6-24 MONTHS

THE NEED TO ADDRESS POSSIBLE IMPAIRMENT OFTEN COMES TO THE ATTENTION OF MEDICAL STAFF LEADERSHIP (FACILITY/HOSPITAL/GROUP)

A FAIR NUMBER COME DIRECTLY TO MEDICAL BOARDS BY COMPLAINT.

FITNESS FOR DUTY EVALUATION

A physician fitness for duty evaluation is a medical assessment that determines if a physician is able to perform their job duties safely.

It determines if a physician can safely perform their job duties without risking injury to themselves or others.

COMPOSITION OF A FITNESS FOR DUTY EVALUATION KSTAR EXAMPLE

- Information from referring facility and physician being evaluated
- Detailed history of events that leading up to the evaluation
- Medical evaluation – comprehensive physical examination including a neurologic exam
- Cognitive screening versus full neuropsychological evaluation
- Occupation therapy – upper extremity evaluation
- Neuropsychiatric evaluation
- Multisource (360) feedback information from coworkers
- Hearing screening

DATA FROM THE EVALUATION

- Reviewed by a Medical Director, clarification sought from examiners when needed
- Pertinent findings are identified and compiled into a final report
- Recommendations are made for any 1) additional testing/treatment, 2) changes in practice (including retirement)
- The final report is given to the referring entity, the physician evaluated; it is also given to a Medical Board or Physician Health Program when indicated

IMPORTANT: DISCUSSING FINDINGS WITH PHYSICIAN EVALUATED

The referring facility is usually contacted first, with findings discussed.

The referring entity meet in person with the physician evaluated to get the report and discuss findings and implications.

The importance of having adequate support available when results are potentially devastating.

The FFD program agrees to speak with and review report with the examinee when requested.

FFD TESTING: SPECIAL NOTES AND CONSIDERATIONS

A team of experienced, independent evaluators are usually used instead a physician's own providers.

When a neuropsychological evaluation is indicated, it is best to use a board-certified clinical neuropsychologist who is not local to nor well know to the physician being evaluated. Contact the provider of the FFD when help is needed in finding a neuropsychologist.

When impairment is found or suspected, consider having the physician contact their State Physician Health Program to help coordinate ongoing work-up and for appropriate protections and advocacy.

DISPOSITIONS AFTER EVALUATION

1

24% had no significant findings; could return to work, no restrictions

2

32% had minor findings that needed follow-up, work continued during expedited workup/treatments

3

27% had findings that required a change in scope of practice and/or supervision

4

17% had to cease practice/retire due to impairment (progressive)

LATE CAREER PHYSICIAN EVALUATION

- Coned down Fitness for Duty Evaluation
- Essentially a screening process for older physicians above a certain age, can be repeated at intervals
- Includes a physical exam, cognitive screening, and review of 5 patient charts
- Can convert to a full FFD if findings are concerning

THANK YOU!

Rob Steele, MD, FAAFP

Rural and Community Health Institute

Texas A&M University

rsteelefp@gmail.com

steele@tamu.edu

CASE #2 FITNESS FOR DUTY EVALUATION

An Anesthesiologist Who Developed Difficulty Managing Complications

Please Refer to Case Handouts