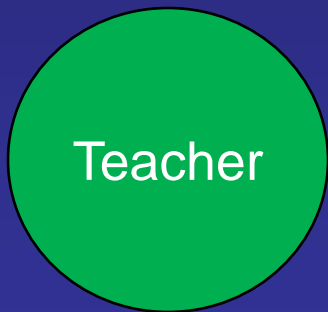


Principles of Assessment: Can This Learner Take Care of Your Mother?

James M Shumway, PhD
Associate Dean for Medical Education
Professor of Medicine and Pediatrics
West Virginia University
School of Medicine

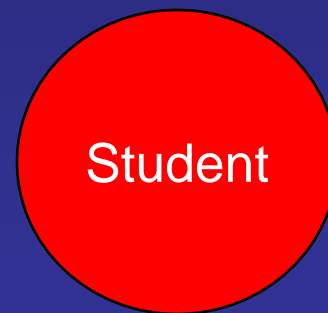
(jshumway@hsc.wvu.edu)



Curriculum



Teacher



Test:

1. When did the Pilgrims land at Plymouth Rock?



1620.



AS YOU CAN SEE, I'VE MEMORIZED THIS UTTERLY USELESS FACT LONG ENOUGH TO PASS A TEST QUESTION. I NOW INTEND TO FORGET IT FOREVER. YOU'VE TAUGHT ME NOTHING EXCEPT HOW TO CYNICALLY MANIPULATE THE SYSTEM. CONGRATULATIONS.



THEY SAY THE SATISFACTION OF TEACHING MAKES UP FOR THE LOUSY PAY.

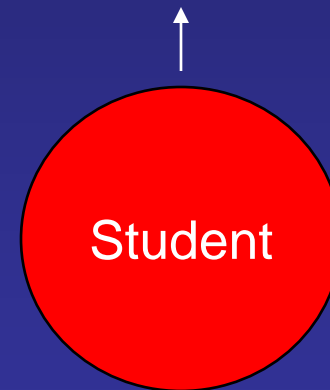
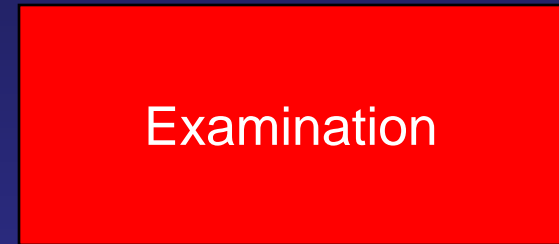
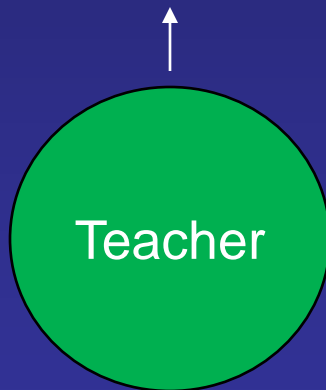
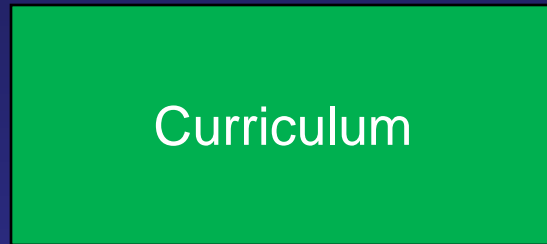


Examination

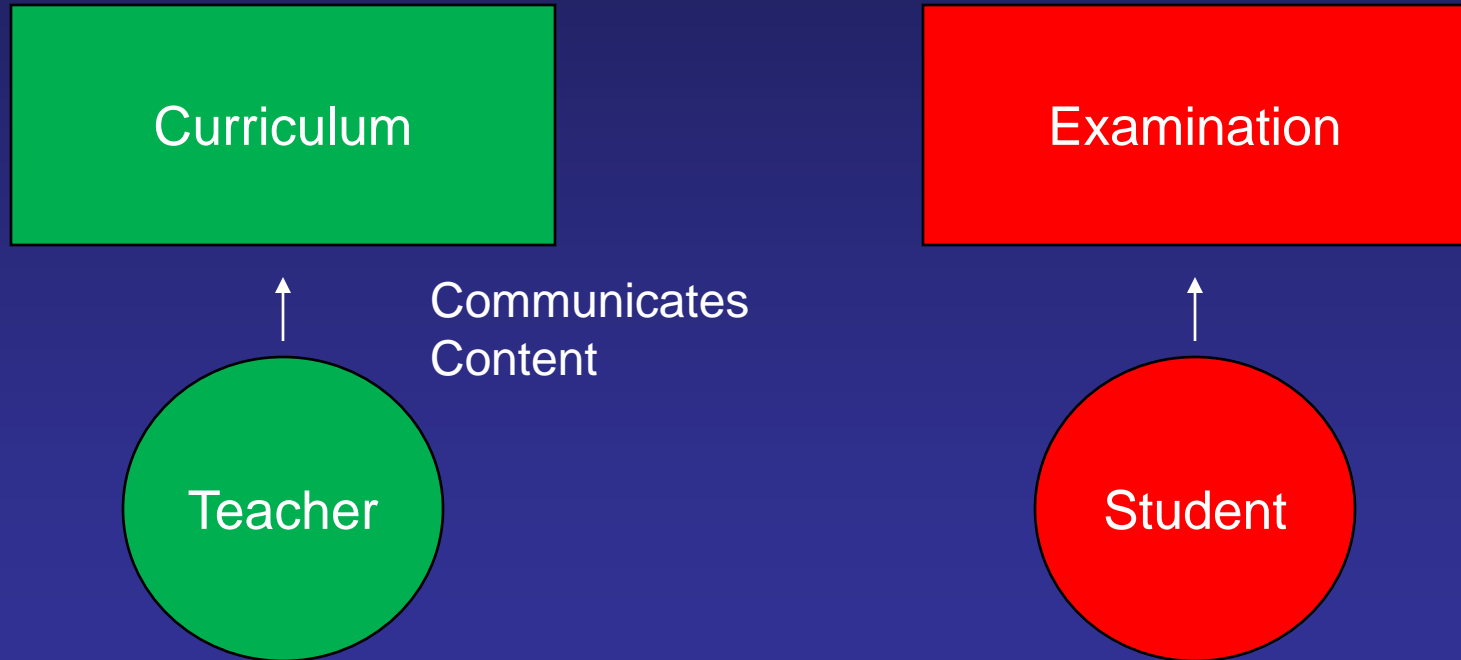
Student



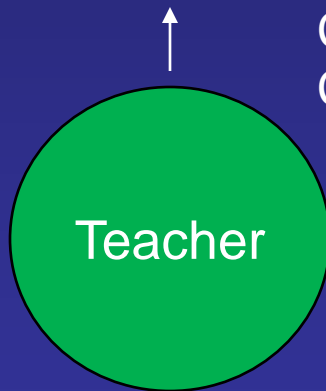
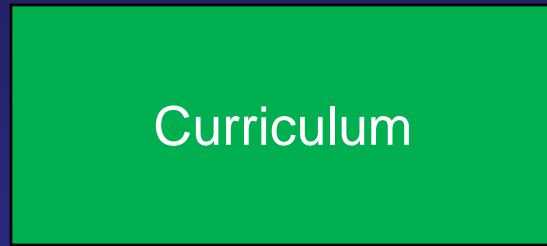
A Teacher's View vs. A Student's View



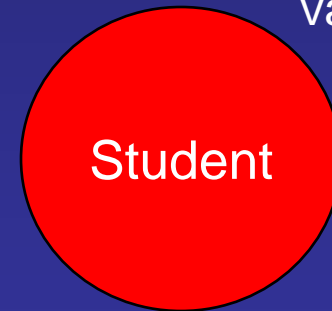
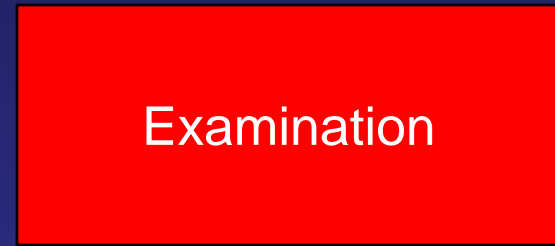
A Teacher's View vs. A Student's View



A Teacher's View vs. A Student's View



↑
Communicates
Content



↑
Communicates
Values



"... I can't go bowling tonight, Freddie, I'm cramming for an IQ test tomorrow..."

Source: Reprinted by permission: Tribune Media Services.

Presentation Objectives:

As a result of attending this presentation you should be able to:

- ... explain that assessment drives learning
- ... define context specificity and generalizability
- ... explain the difference between norm-referenced and criterion-referenced assessment (standard setting)
- ... identify the key concepts of formative and summative assessment, reliability, and validity
- ... describe assessment authenticity and the move to more performance-based assessment practices including the methods in current use
- ... describe the current issues surrounding assessment practices and the need for further development and study

Purposes of Assessment

- To determine whether course or clerkship objectives are met.
- To certify or judge competence of students.
- To evaluate the teaching program.
- To predict future performance of students.
- To guarantee to the public that we are producing safe practitioners (accreditation).

Key Concepts

- Formative and Summative Assessment
- Test Blueprinting
- Reliability and Validity
- Cost
- Impact on Learning
- Context Specificity and Generalizability
- Standard Setting (Norm- vs. Criterion-Referenced assessment)

Key Concepts

- **Formative and Summative Assessment**
- Test Blueprinting
- Reliability and Validity
- Cost
- Impact on Learning
- Context Specificity and Generalizability
- Standard Setting (Norm- vs. Criterion-Referenced assessment)

Formative and Summative Assessment

- Formative
 - Process focused
 - Feedback oriented
 - Low stakes
 - Is corrective so student can improve before educational experience is completed
- Summative
 - Outcome focused
 - Determines achievement of student
 - High stakes
 - Requires strict quality control

Key Concepts

- Formative and Summative Assessment
- **Test Blueprinting**
- Reliability and Validity
- Cost
- Impact on Learning
- Context Specificity and Generalizability
- Standard Setting (Norm- vs. Criterion-Referenced assessment)

Test Blueprinting

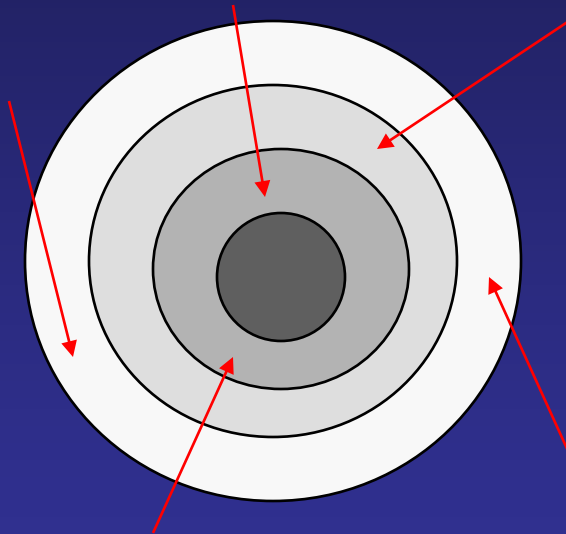
- Sampling of content reflective of objectives and material
- Creation of a table of content
- Identification of assessment methods in cells of table
- Assignment of faculty to develop test questions for each cell
- Assignment of weight to questions

Key Concepts

- Formative and Summative Assessment
- Test Blueprinting
- **Reliability and Validity**
- Cost
- Impact on Learning
- Context Specificity and Generalizability
- Standard Setting (Norm- vs. Criterion-Referenced assessment)

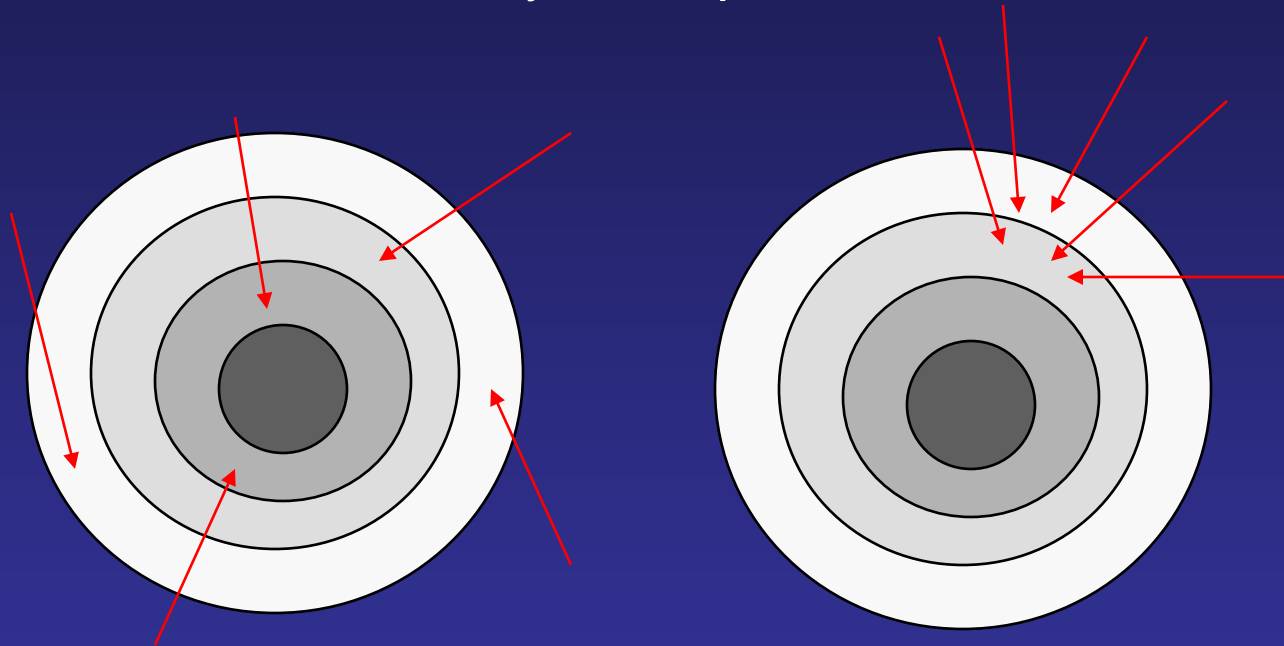
Reliability

- To assess the consistency of responses



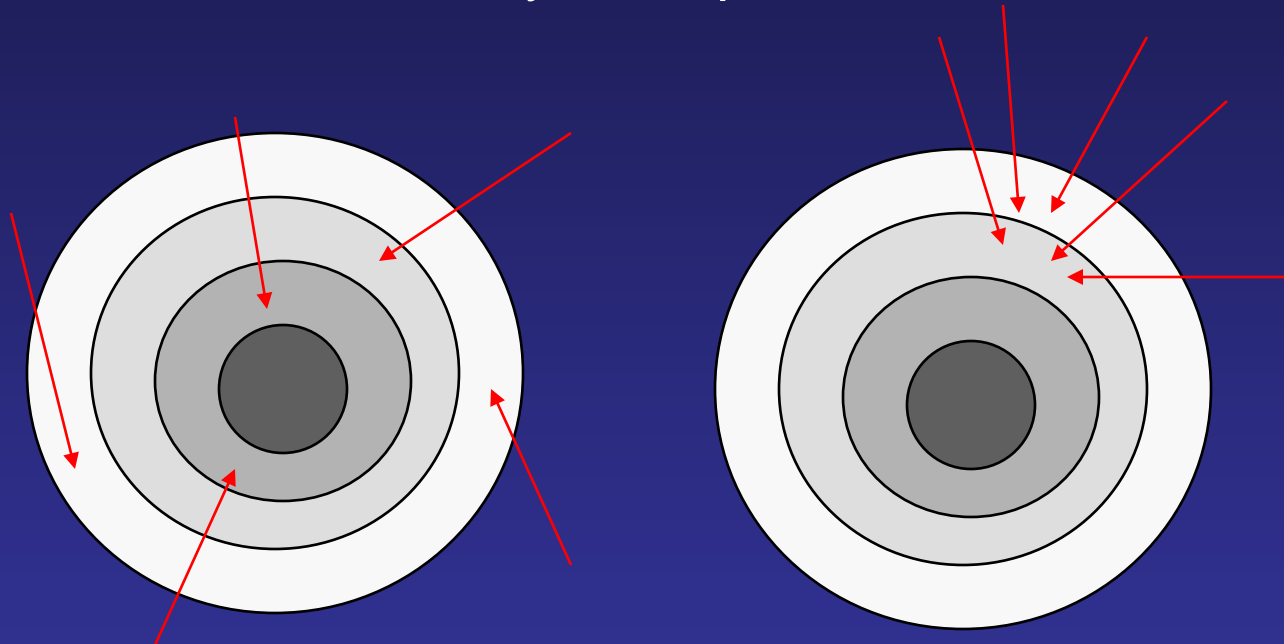
Reliability

- To assess the consistency of responses



Reliability

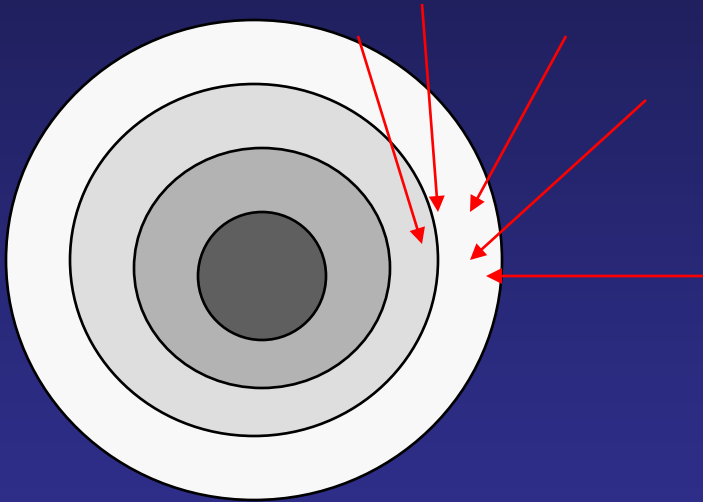
- To assess the consistency of responses



- Internal consistency (correlating items)
- Consistency over time (test-retest)
- Rater reliability

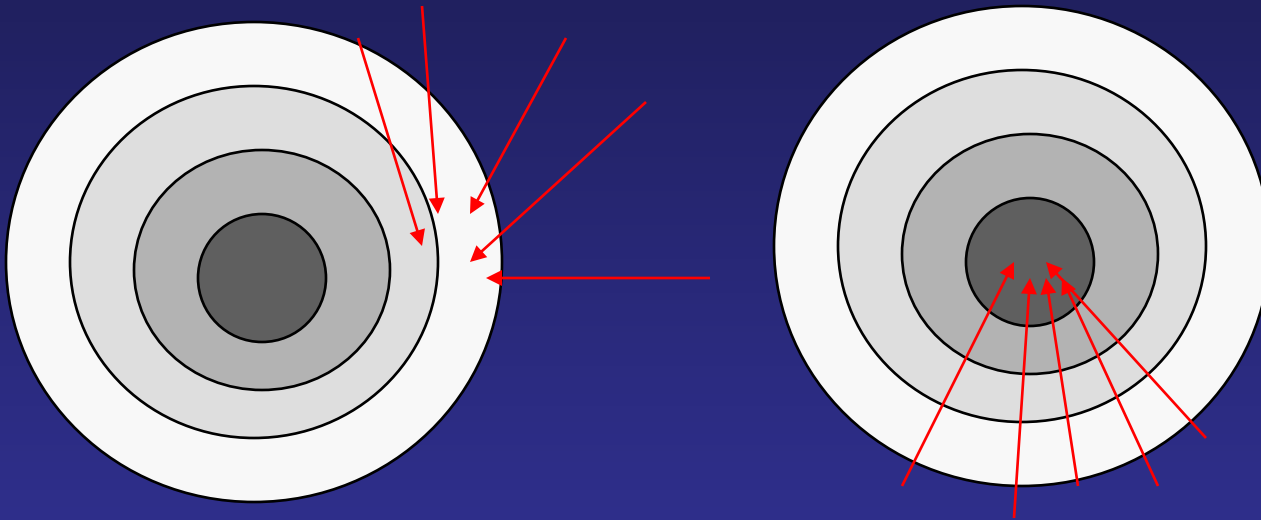
Validity

- Truthfulness (measures accurately what you want it to measure).



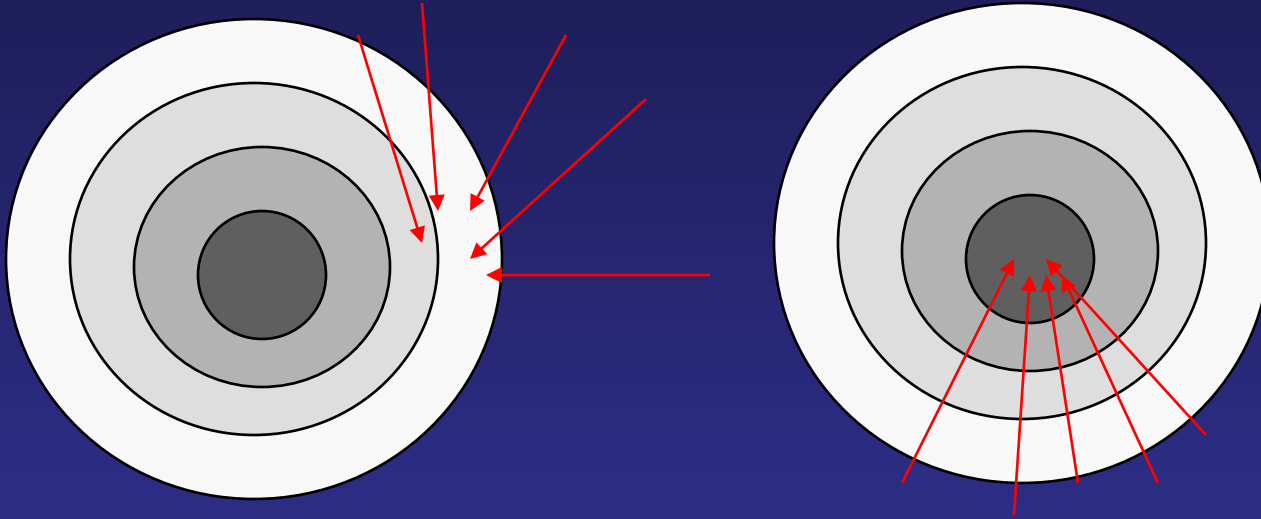
Validity

- Truthfulness (measures accurately what you want it to measure.)



Validity

- Truthfulness (measures accurately what you want it to measure.)



- Compare results with other measures' results.
- Compare results with results from diverse groups to see if expected differences match.

Key Concepts

- Formative and Summative Assessment
- Test Blueprinting
- Reliability and Validity
- **Cost**
- Impact on Learning
- Context Specificity and Generalizability
- Standard Setting (Norm- vs. Criterion-Referenced assessment)

Cost

- Good assessment is costly:
 - Time to develop, administer, and grade
 - Faculty development
 - The cost of assessment often requires compromise in practice.
 - Investing in assessment is investing in teaching and learning.

Key Concepts

- Formative and Summative Assessment
- Test Blueprinting
- Reliability and Validity
- Cost
- **Impact on Learning**
- Context Specificity and Generalizability
- Standard Setting (Norm- vs. Criterion-Referenced assessment)

Impact on learning

- Assessment drives learning through its content, format, information given, and programming (frequency, timing)
 - Exams are the hidden curriculum. Exams define academic success for the learner.
 - Exams can reinforce (un)desirable learning behavior.
 - The effects of assessment are often difficult to assess.
 - Any assessment action will result in an educational reaction.

Key Concepts

- Formative and Summative Assessment
- Test Blueprinting
- Reliability and Validity
- Cost
- Impact on Learning
- **Context Specificity and Generalizability**
- Standard Setting (Norm- vs. Criterion-Referenced assessment)

Context Specificity and Generalizability

Context Specificity (*case specificity*):

- There is no generic problem solving approach.
- Performance is problem specific.
- Therefore, need multiple sampling strategies, cases, raters, and items to have confidence in students' performance.

Generalizability:

- Applicability of test results to predict performance beyond the test.

Key Concepts

- Formative and Summative Assessment
- Test Blueprinting
- Reliability and Validity
- Cost
- Impact on Learning
- Context Specificity and Generalizability
- **Standard Setting** (Norm- vs. Criterion-Referenced Assessment)

Standard Setting

- A standard is a score that serves as a boundary to separate those who pass or fail.
- Understanding the differences between
 - Norm-referenced assessment
 - Criterion-referenced assessment

Norm-Referenced:

- Individual vs. Others.
- Standard is relative.
- Standard is not known in advance. Influenced by others' abilities.
- Purpose is to discriminate among individuals.
- Distribution is assumed to be normal.
- Feedback relative to performance is dependent on individual's abilities.

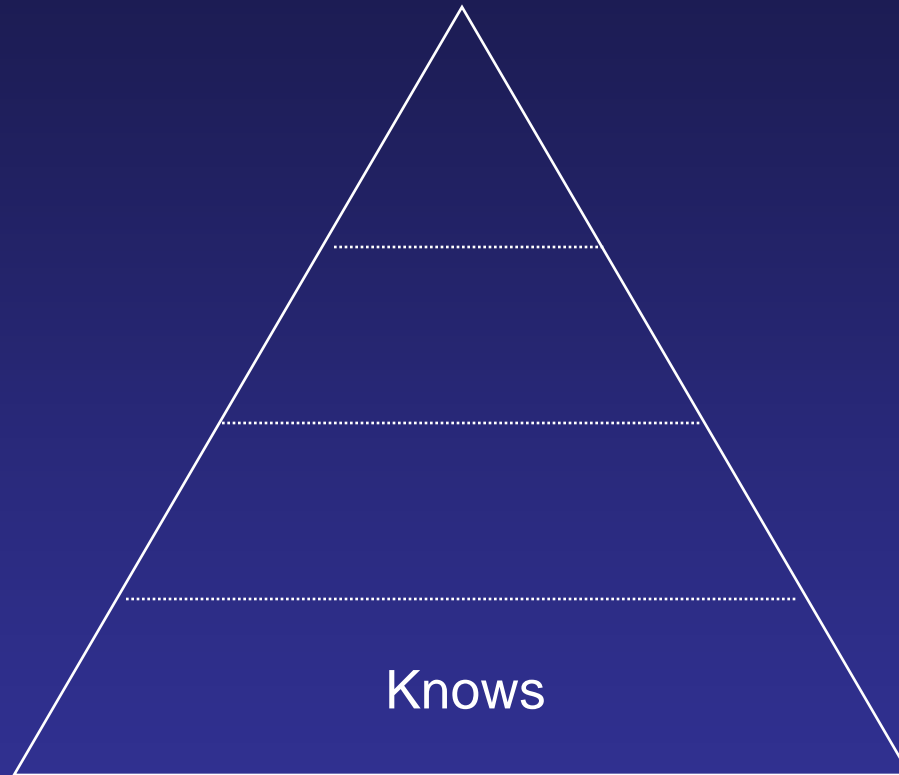
Criterion-Referenced:

- Individual vs. fixed standard.
- Standard is absolute.
- Standard is known in advance.
- Purpose is to assess competence.
- Pass-fail decision is independent of others' performance.
- Feedback relative to performance is clear.

An Assessment Model

Miller Learning

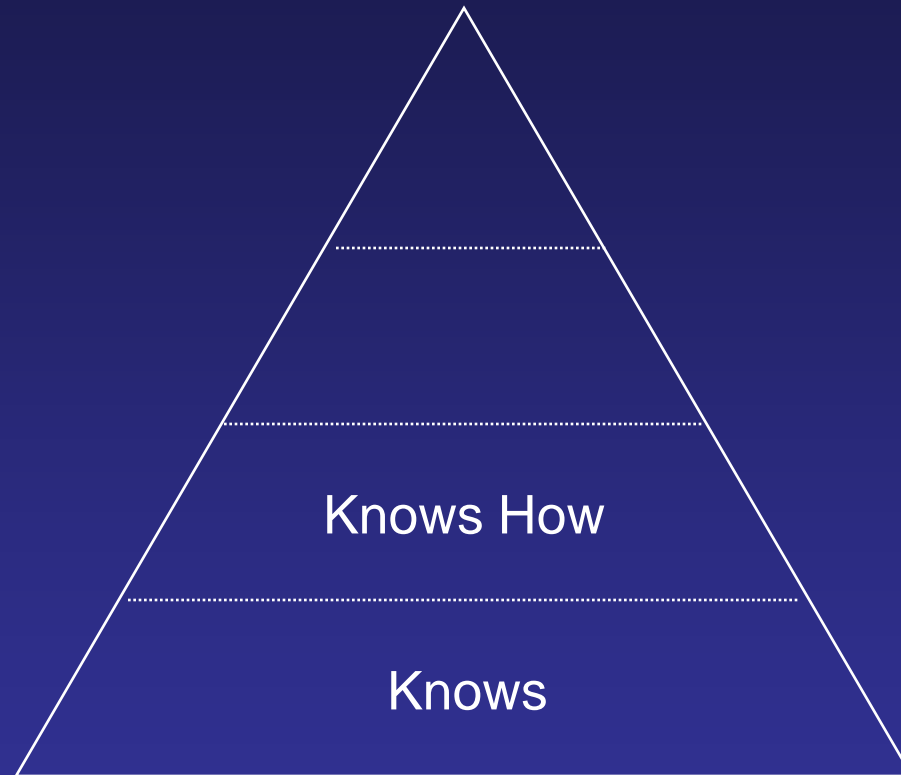
Pyramid



Miller GE, 1990

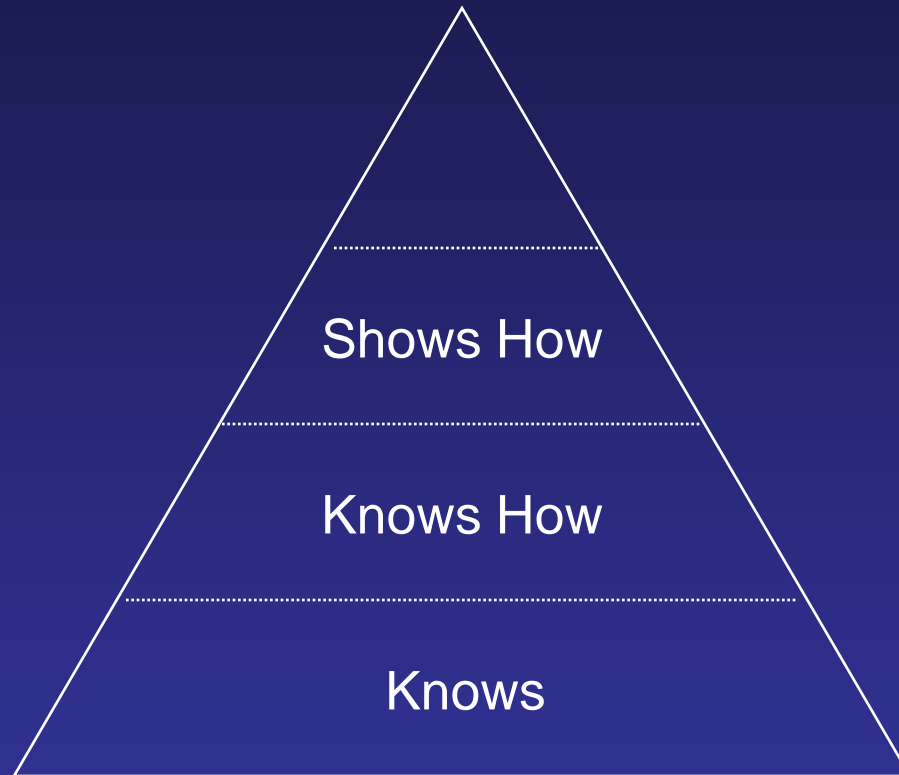
Miller Learning

Pyramid



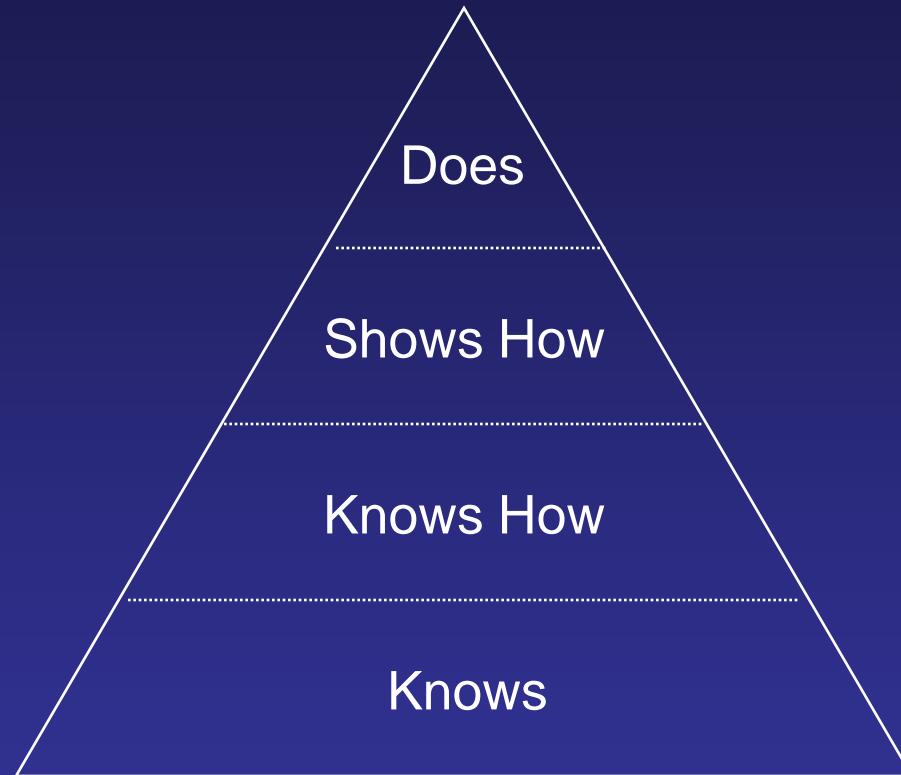
Miller Learning

Pyramid



Miller Learning

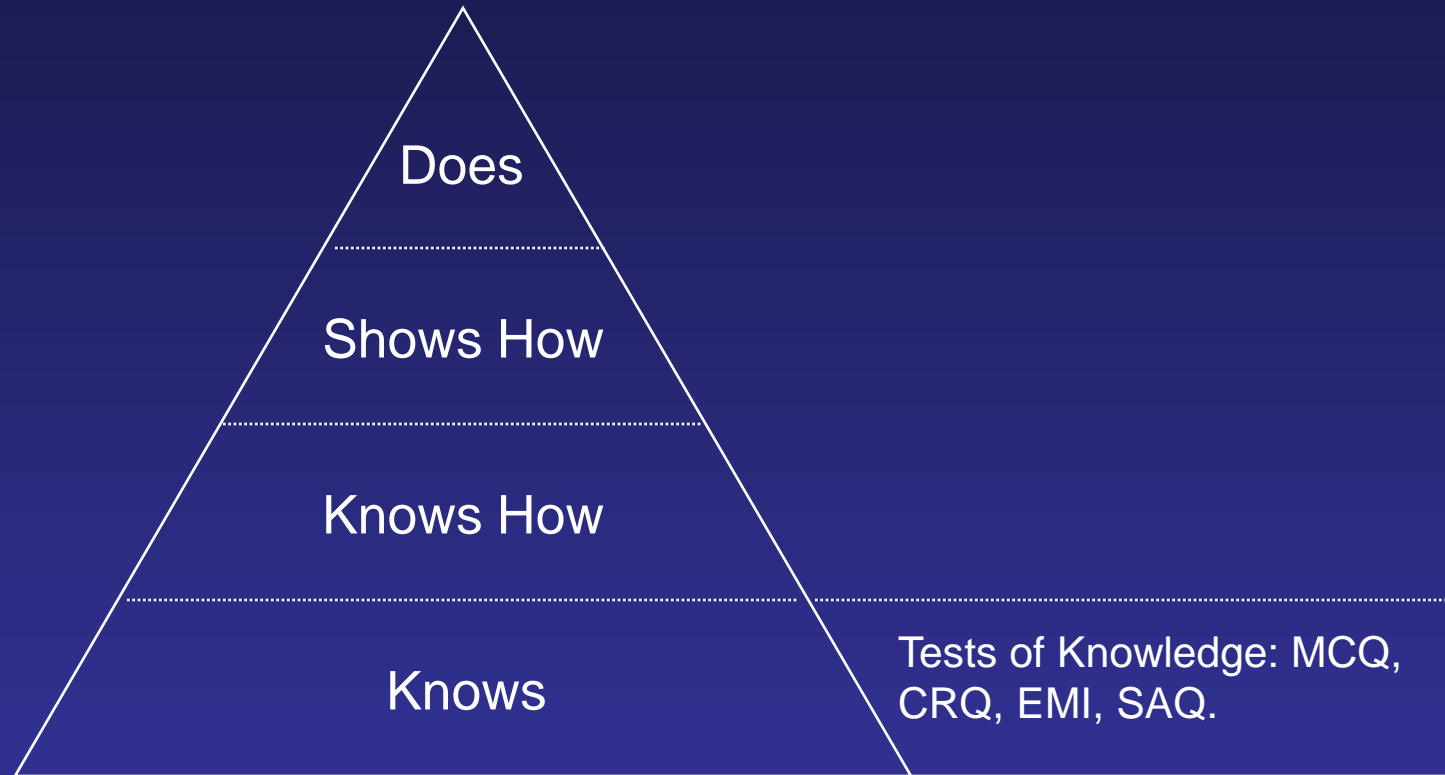
Pyramid



Categories of Assessment Instruments

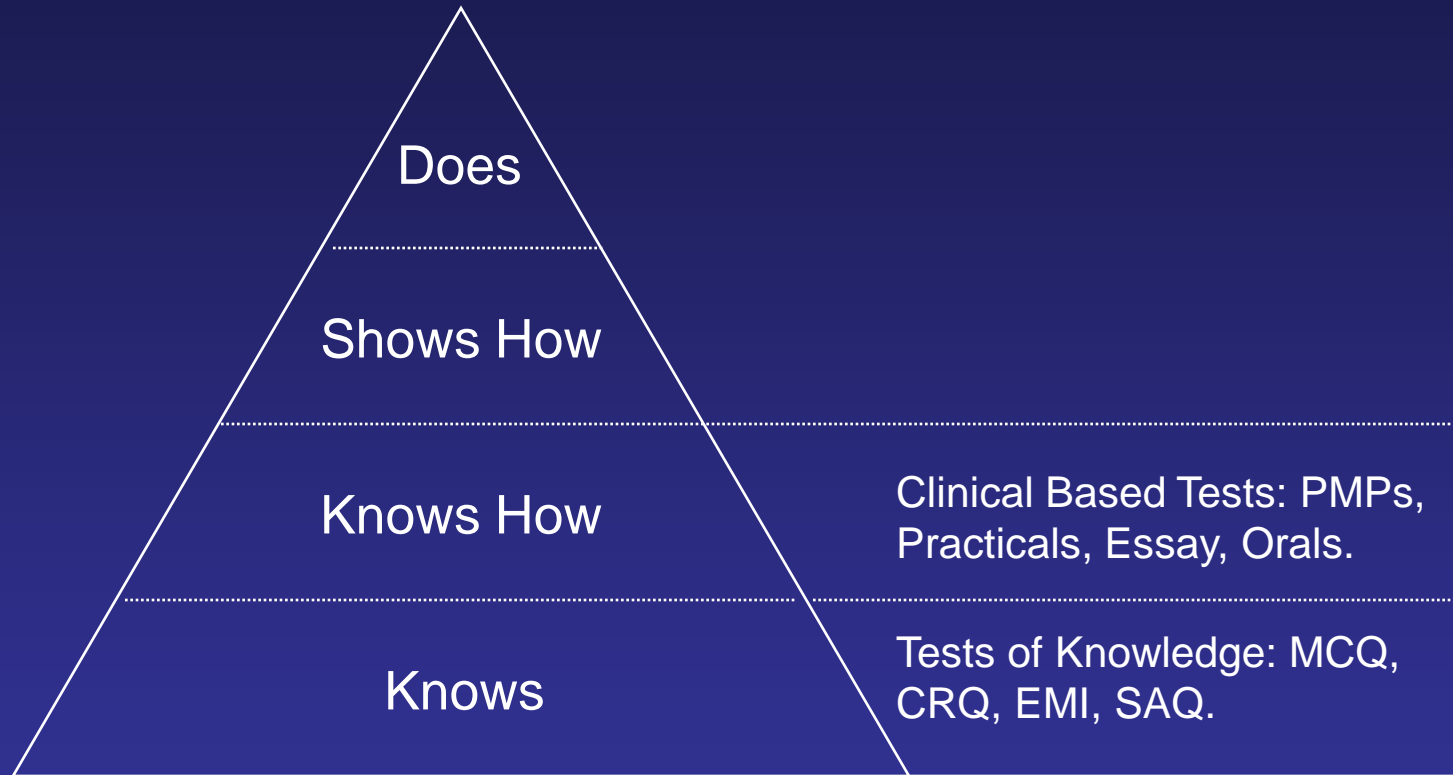
- Written Assessments
 - MCQs, CRQs, EMIs, PMPs
 - Essays, Short Answer
 - Practicals
 - Diagnostic Thinking Inventory
 - Progress Tests
- Portfolio Assessments
 - Logbooks &
 - Portfolios
- Multiple Station Exams
 - OSCE
- Simulations
 - Standardized (Simulated) patients (SPs)
 - Simulations & Models
 - Role playing
- Observation
 - Checklists
 - Rating Scales
 - Peer- & Self- Assessment

Miller-Van der Vleuten Learning-Assessment Pyramid

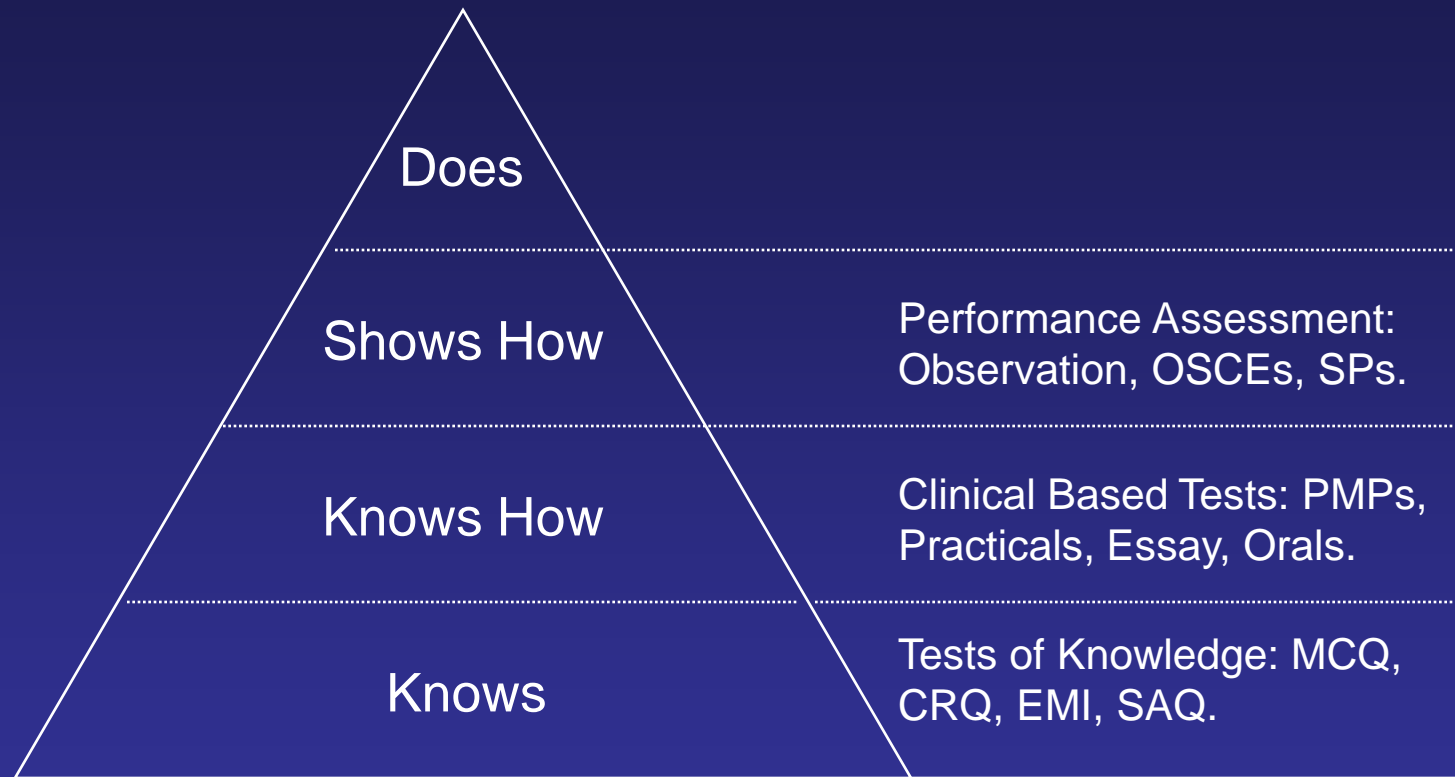


Miller GE., 1990
Van der Vleuten C., 2000

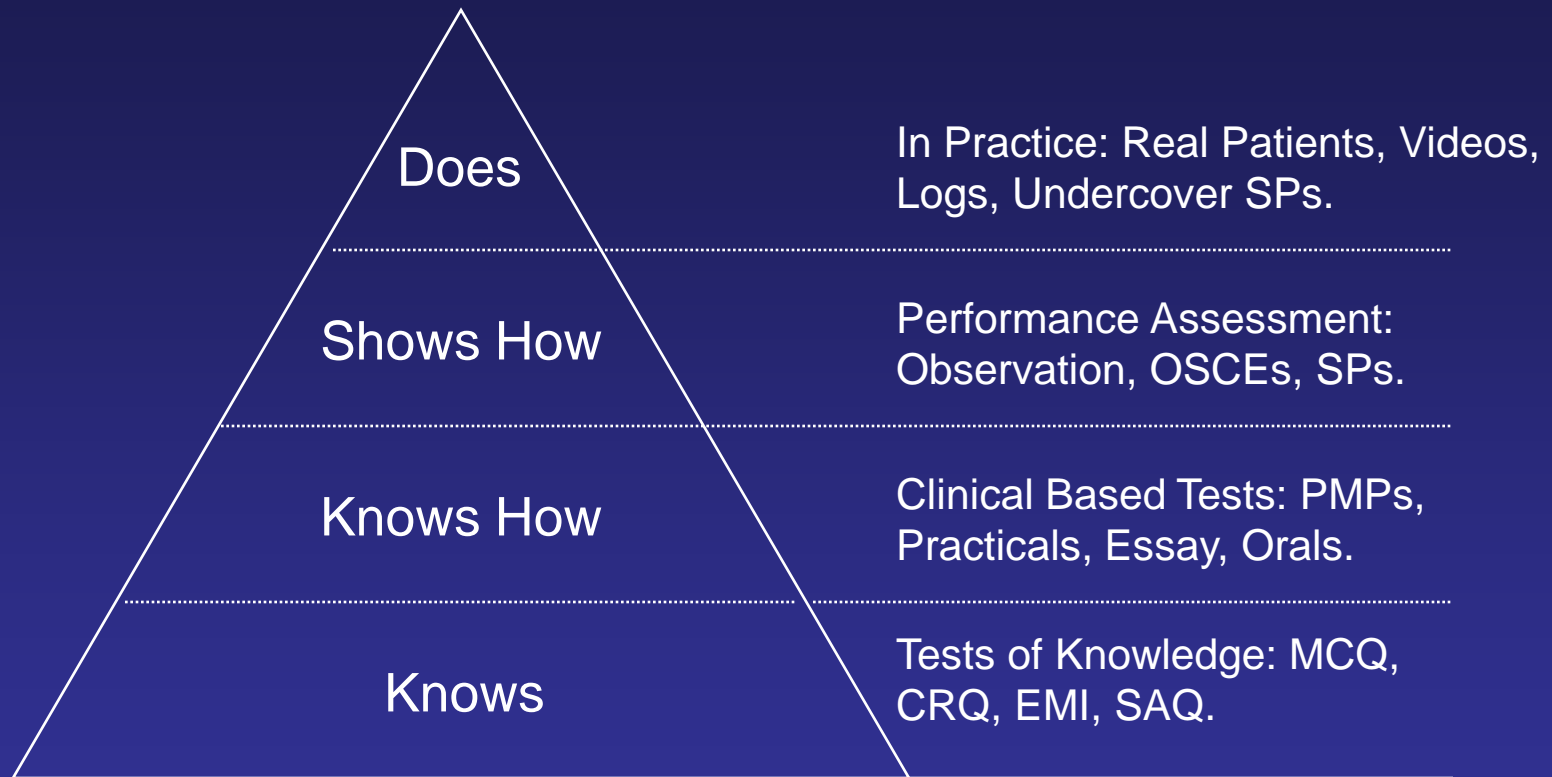
Miller-Van der Vleuten Learning-Assessment Pyramid



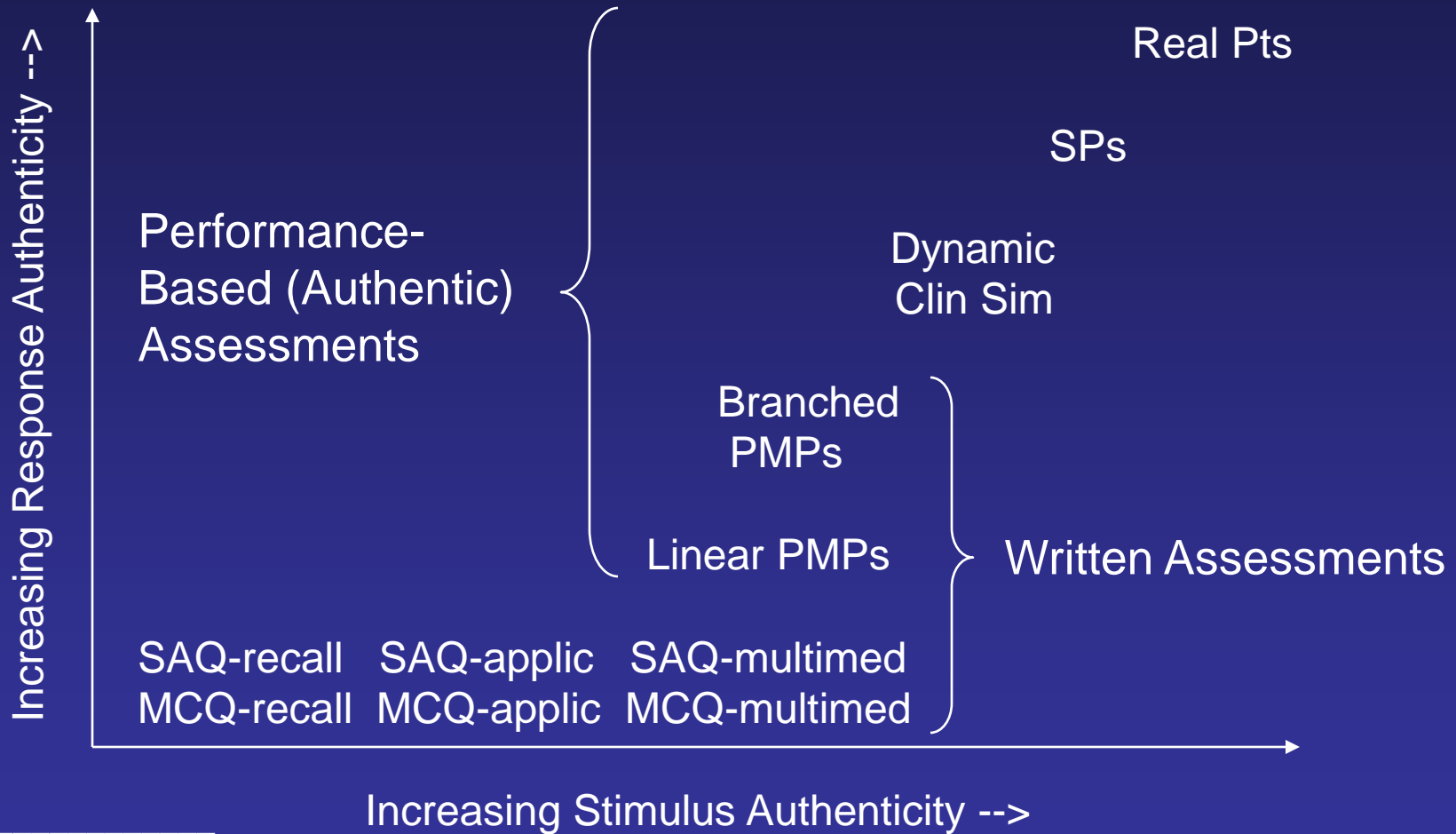
Miller-Van der Vleuten Learning-Assessment Pyramid



Miller-Van der Vleuten Learning-Assessment Pyramid



Swanson's Assessment Authenticity Graph



Harden's Assessment by Authenticity Table

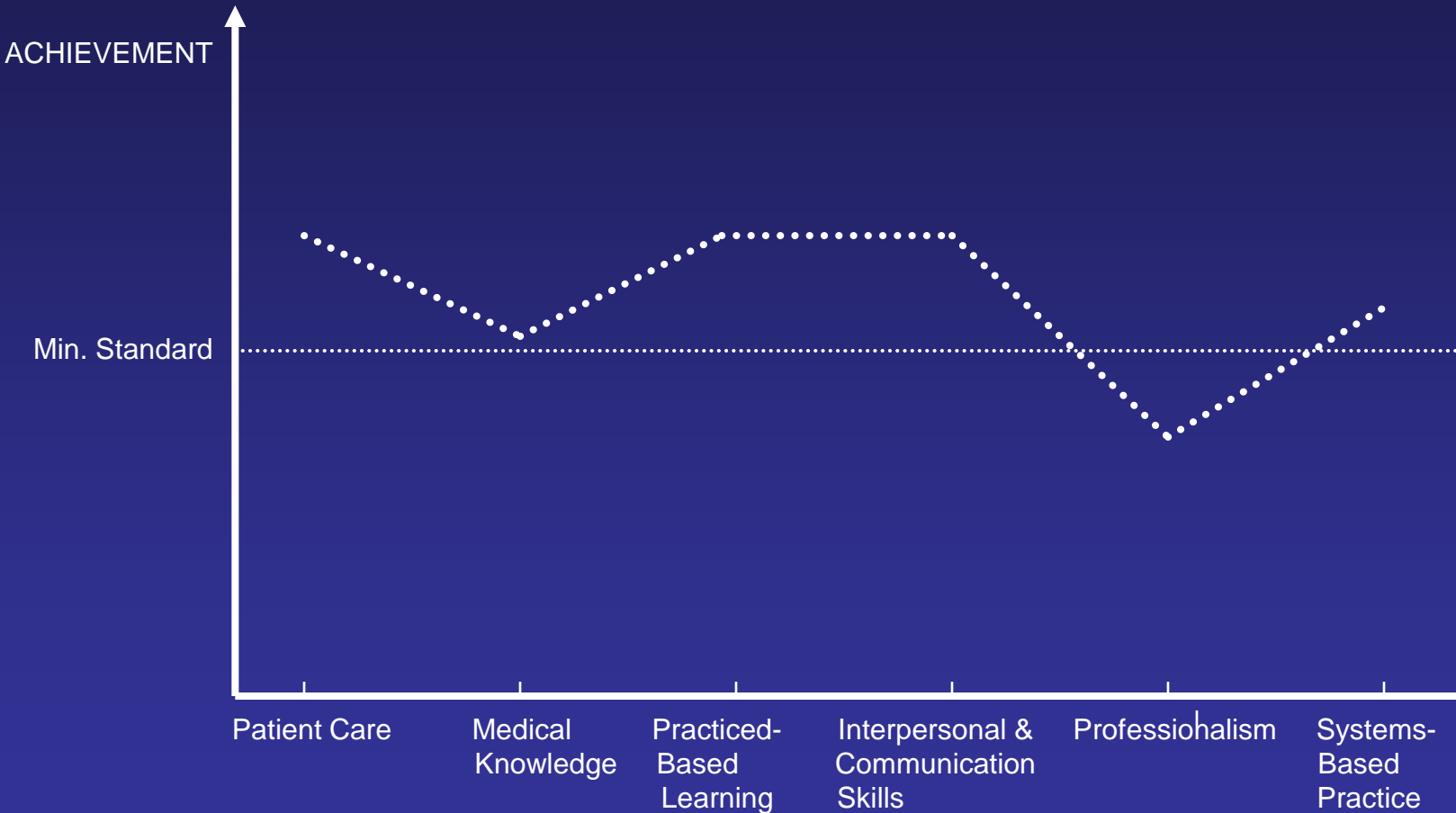
Direct	Orals. Problem-solving exercises.	OSCEs. SPs.	Checklists. Rating scales.
Indirect	MCQs. Essays. EMIs. Progress tests.	Simulations. Models. Role Playing.	Portfolios. Logs.
	Mediated	Simulated	In Practice

Authenticity →

Matching Assessment Approaches to Competency

- Patient Care
 - OSCEs, SPs, Simulations, Logs, Practicals, Observation
- Medical Knowledge
 - MCQ, Essay, Progress Tests
- Practice-Based Learning & Improvement
 - Peer- & Self-Assessment, Portfolios
- Interpersonal & Communication Skills
 - Peer- & Self-Assessment, Portfolios, Observation, OSCEs, 360 Evaluations
- Professionalism
 - 360 Evaluations, OSCEs, Observation
- Systems-Based Practice
 - Observation, Portfolios

Example of an Assessment Profile for a Student



e.g.: ACGME Competencies

A Few Specific Assessment Methods

MCQ example

A 60-year-old man comes to the physician because of a 6-month history of gradual progression of fatigue and a 4.5-kg (10-lb) weight gain. He has had only occasional shortness of breath during this period. He has smoked 2 packs of cigarettes daily for 35 years. Physical examination shows ankle edema. Crackles are heard at the lung bases. Which of the following is most likely to be decreased in an arterial blood sample?

- (A) Bicarbonate concentration
- (B) Erythropoietin activity
- (C) Hematocrit
- (D) PO₂**

MCQ

- Reliability (↑)
- Validity (↑ objective)
- Cost (↑ if done properly)
- Impact on learning
 - Students learn to pass the test.
 - Students study for test rather than learning material.
 - Faculty instruction designed around what needs to be learned to pass the test.

New on the Scene: Portfolios

An Example

One school of nursing began using student portfolios in response to the National League for Nursing's requirement for schools to develop outcome assessment programs to document students' abilities in critical thinking, communication, and therapeutic nursing interventions. This school started small and focused on "critical thinking".

Students were instructed by the faculty to begin their portfolios in the first semester professional nursing course and complete them in the capstone course. Portfolios were to include (a) copies of the students' Personal Professional Beliefs Papers (PPBP) from these 2 courses and (b) one completed class assignment from each semester that they felt best represented their writing and critical thinking abilities. With each chosen document, students were to include a self-reflection piece that summarized the strengths and weaknesses of the completed assignment and how that work represented their growth as a nursing professional.

Portfolios

- Reliability
 - Due to ability of rater to agree on standards and criteria.
 - Triangulation of evidence from a number of sources.
- Validity
 - Extent to which they accurately document experiences that are indicators of mastery of desired learning outcomes.
- Cost
 - Requires staff time and resources (especially external review)
 - Additional \$\$ when developing “electronic” portfolios.
- Impact on learning
 - Because document what the learner has done and asks them to reflect on their strengths, weaknesses, and accomplishments.

Clinical and Practical Assessments: OSCEs and SPs

- Assesses performance, concerned with what student “can do” rather than what “knows”.
- Samples performance
- Evaluates in simulated “high fidelity” situations without putting learner or patient at risk.

Clinical and Practical Assessments: OSCEs and SPs

- Reliability
 - ↑ number of stations, ↑ reliability (e.g., 15-20)
 - ↑ reliability when checklists and ratings scales are used by SP raters.
- Validity
 - ↑ number of stations, modest validity increase.
- Cost = ↑, time and resources
 - Maintaining security of exam over long run.
 - Startup and continuing costs of SP program.
- Impact on Learning
 - Positive experience for students.
 - Formative evaluation as the student is participating in it.
 - Or, students may prepare for exam by compartmentalizing skills and not understand their connection or flow in clinical care.

Peer- and Self-Assessment

- Strengths
 - Focuses on an important question: Who should do the evaluating?
 - Particularly used in assessing attitudes and communication skills.
 - Checklist and rating scales can be used.
 - Peers have advantage of observing under real clinical conditions.
 - Peers can be more discriminating because of increased exposure of observation.
 - Shown to correlate highly with faculty evaluations.
- Weaknesses
 - May carry less weight than faculty ratings, therefore used in conjunction with.
 - Difficulty in training of peer raters (same with training of faculty raters).
 - Under used in health professions education.

Peer- and Self-Assessment

- Reliability
 - Adequate with trained raters.
- Validity
 - Adequate with well designed focus on observable behaviors.
 - Use of behaviorally-anchored rating scales.
- Cost
 - Those normally associated with training of raters and design of rating scales (time)
- Impact on learning
 - Power to transform students, esp. with self-appraisal.
 - But, if done poorly, could cause mistrust, suspicion, and peer rivalries.

Trends in Assessment

- Accreditation bodies increasingly impact assessment practices.
- Use of more “performance-based” exams.
- Increased use of “external” standardized tests for student promotion decisions.
- Movement of providing “less” feedback to students on what details of items they missed.
- Increasing public calls for “accountability” often demonstrated by increased use of assessment in our education programs.

Implications for Further Development and Study

- Develop faculty and staff.
- Pay attention to standard setting:
 - Understand what “marks” mean.
 - Have a quality control process in place.
 - Listen to student input.
 - Use evidence-based assessment practices.
 - Use “test blueprints” to adequately sample the knowledge, skill, attitude, and behavior (KSAB) domains.
 - Set standards for assessing in the clinical setting.
- Include the use of qualitative measures.

An Assessment Manifesto

- Assessment should help students learn and play a positive role in learning.
- Assessment should accommodate individual differences among students.
- Assessment purposes should be clearly understood by all involved – students, faculty, staff, administration.
- Assessment needs to be reliable and valid.
- Assessment needs to provide formative information for the learner in addition to summative information for the institution or program. In other words, it should help the student develop.
- Assessment should provide opportunities for faculty, staff and students to reflect on their learning and what they are doing.
- Assessment should be designed as part of the curriculum, not attached to it as an additional appendage.
- The amount of assessment needs to be appropriate for the given learning outcomes and expectations.
- The criteria for assessment and standard setting needs to be understood by all involved. Students, especially, need to understand what is required of them and how they will be assessed at each stage of their learning.
- Multiple assessment strategies should be used to “triangulate” on the behavior being observed.

References (1 of 2)

- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999) *Standards for Educational and Psychological Testing*. American Educational Research Association, Washington, D.C.
- Amin Z & Khoo HE (2003) Overview of Assessment and Evaluation. In *Basics in Medical Education*, 251-260 (Oxford University Press, Oxford, UK)
- Brown SR, Smith B. (1996) *500 Tips on Assessment*. London: Kogan Page.
- Melnick D. (2001) Presentation at Marshall U. February 12.
- Miller GE. (1990) The assessment of clinical skills / competence / performance. *Academic Medicine*, 65 (Supplement): p. S63-S67
- Newble D (1998) Assessment. In: Jolly B & Rees L (eds.) *Medical Education in the Millennium*, 131-142 (Oxford University Press, Oxford, UK)

References (2 of 2)

- Robertson JF, Elster S, Kruse G. (2004) Teaching Evaluation: Portfolio Outcome Assessment – Lessons Learned Nurse Educator, March/April, Vol 29, No. 2, PP 52-53.
- Shumway J.M. and Harden R.M. (2003) AMEE Guide No. 25. The assessment of learning outcomes for the competent and reflective physician. *Medical Teacher*, Vol 25, No. 6, pp. 569-584.
- Suskie LA. (1996) *Questionnaire Survey Research: What works*. Association for Institutional Research, Tallahassee, Fl.
- Swanson D. (2000) Performance-Based Testing. *Ottawa Conf. Presentation*. South Africa.
- Van der Vleuten C. (2000) The Assessment of Professional Competence: Developments, Research and Practical Implications. *Ottawa Conf. Presentation*. South Africa.