

OVERVIEW OF ITEM FORMATS

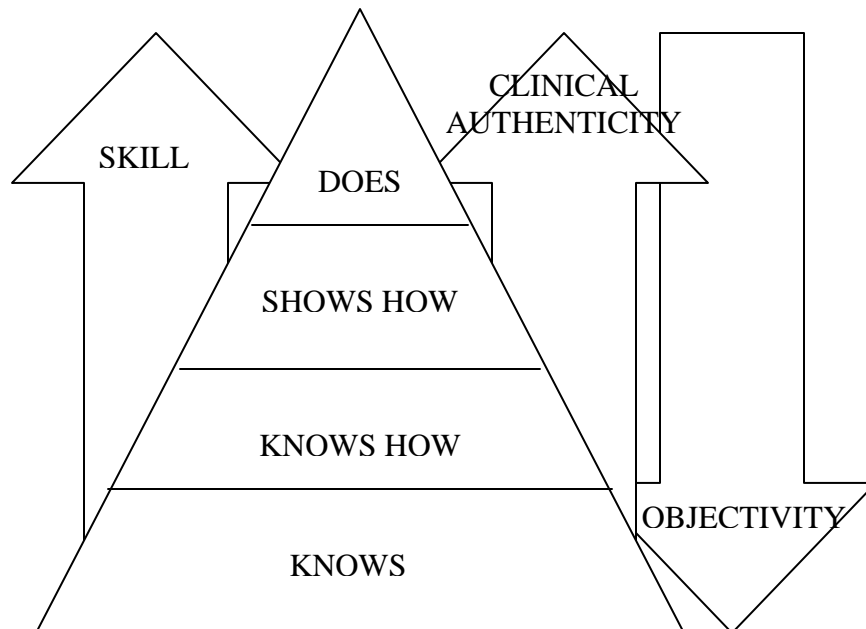
1. Item format according to Miller's Pyramid

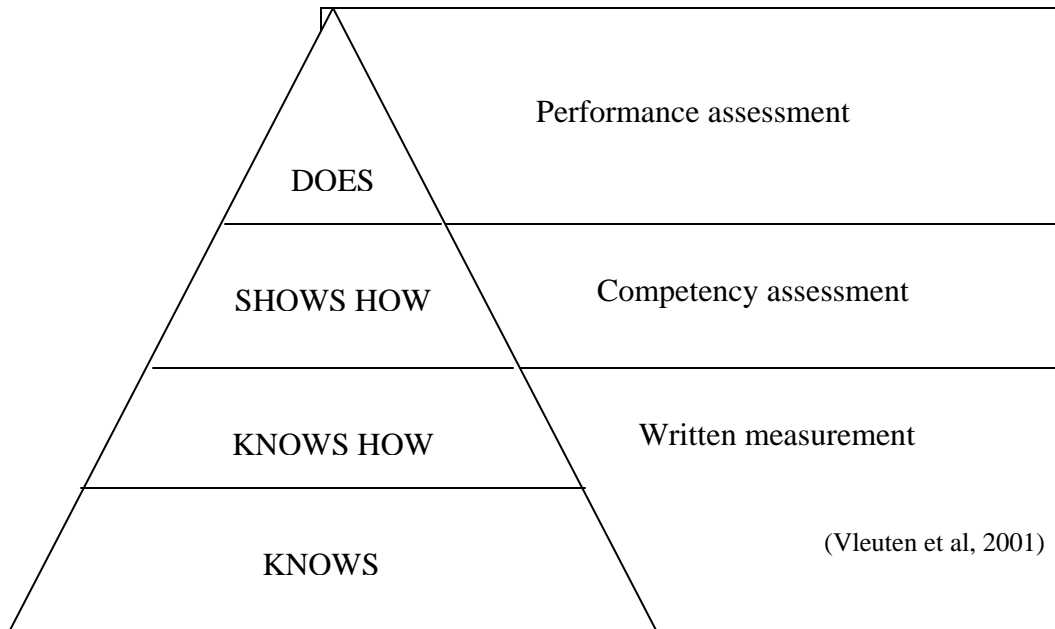
- Pen – and – paper methods
 - Free response
 - Selected (fixed) response
- Competence assessment
 - OSCE
- Performance assessment
 - Observation, etc.

2. Principles

- Response format vs. stimulus format
- Authenticity (realistic)

3. Miller's Pyramid (1990)





4. Stimulus and response formats

- Stimulus format
 - A hypertensive 66-years-old obese lady complains of epigastric pain – CONTEXT RICH
 - Differential diagnoses of epigastric pain include – CONTEXT FREE
 - Stimulus format is the more important determinant of the thought process invoked
 - Don't be fooled by the name, look at the stimulus and look at what you need to test
- Response formats
 - Stimulus being equal, choose the most objective response formats.
- Some consideration
 - Do not fall into 'either/or' thinking; frequently better to use both

- Select the item type that provides the most direct measure of the intended learning outcome
- An effective evaluation system contains several methods that reflects competence in identified areas of performance
- Final judgement on student progress should be made on the basis of multiple assessments, obtained on different occasions, using a variety of methods based on stated goals (Norman and Shannon, 1995)
- Every formats has their advantages and disadvantages

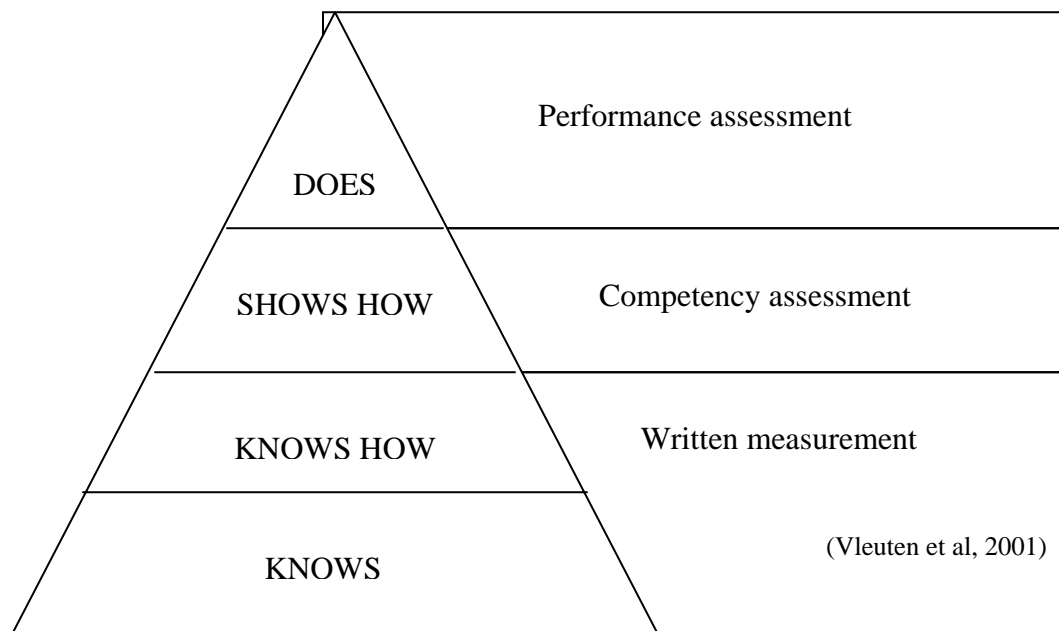
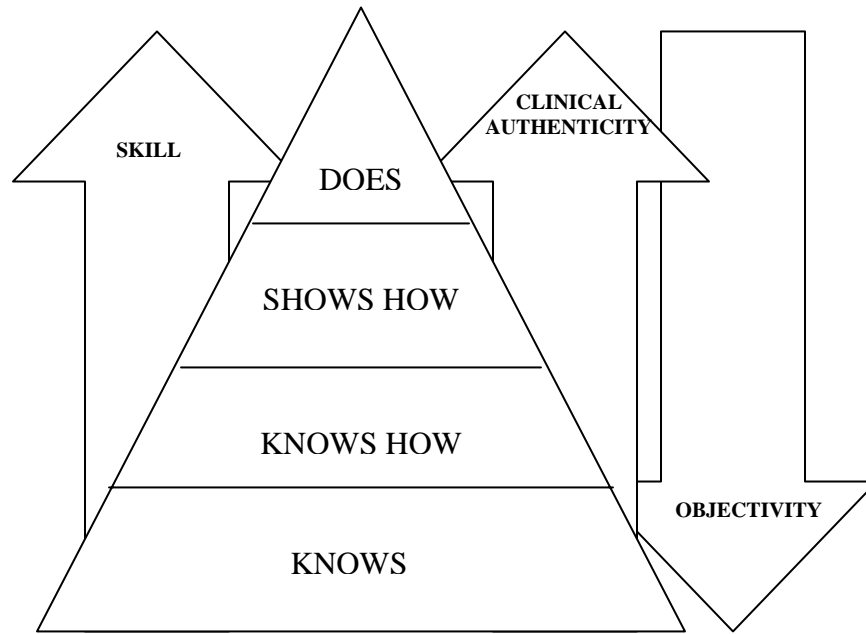
5. Authenticity

- Improved authenticity = improved validity
- Trends
 - Include assessment day-to-day practice
 - Integrate competencies
 - Authenticity at all level of the pyramid → support multiple methods
 - Don't ignore general professional competencies
 - E.g. teamwork, professional behavior, reflection
 - Complex competencies → increased reliance on professional judgement → how to make vigorous (validity and reliability)

ASSESSMENT OF 'KNOWS' AND 'KNOWS HOW'

1. Revision of basic concepts:

- Miller's Pyramid (1990)



- Response format & stimulus format
-

		Open-ended / Free response	Select / multiple choice
Context - Rich	} STIMULUS FORMAT	} RESPONSE FORMAT	
Context - Free			

2. Measuring 'thinking skill'

- Look at cognitive process required
 - E.g. recall/memory, interpretation/comprehension, problem-solving/reasoning.
 - Difficult to determine:
 - Depends on the background of candidate as well as question content
 - The same answer is arrived at using different processes.
- Look at task of examinee
 - Simpler, more objective
 - Requires rote memory for isolated facts → recall item
 - Requires to reach a conclusion/make a prediction/select a course of action → application of knowledge item
 - E.g. to assess knowledge of the days of the weeks

- Context - free
 - List the days of the week
- Slightly context – rich
 - Work days in Kelantan include
- Context – rich
 - A ruler in Manaland decided to have an 8-day week and inserted a new day, Fuadday, as the first day of the weekend. In Manaland people start work on Tuesday and now work 6 full days. The day after Fuadday is?
- Promoting application of knowledge
 - Using novel situation
 - In context → context rich
- As far as possible all items should be assess application of knowledge

3. Paper based methods

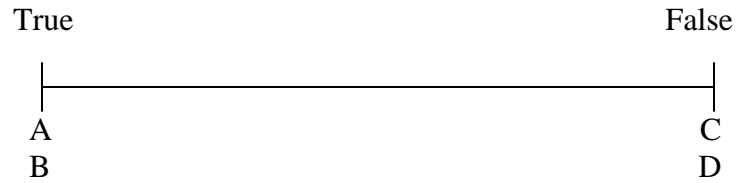
- Select type
 - Type X (Multiple true-false)
 - Type A (Single best response)
 - Type R (extended matching)
 - The progress test
- Supply type
 - Essay
 - Modified Essay Question
- Mixture
 - Key Feature Problems/ Clinical Reasoning Exercise

4. General characteristics of select type (MCQ):

- Strengths
 - Effective:

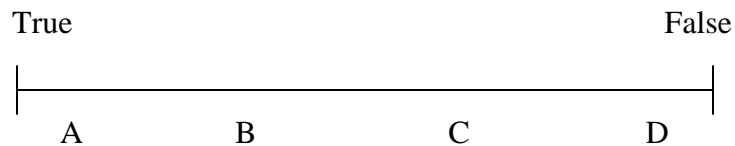
- contributes to high validity and reliability per hour of testing time
- Objective:
 - contribute to reliability
- Cost-effective:
 - Administration and scoring
- Higher-order thinking skills measurable if properly formatted
- Weakness
 - Resource-intensive:
 - Requires skill and time to write, especially context-rich questions
 - Often used to test recall of facts; trivialisation
 - Promotes surface learning in students if used to test primarily recall
 - Promotes trivialisation and avoidance of important areas and cognitive skills among item-writers if faced with difficulty in writing
 - Not able to assess summarizing and writing skills, critical thinking processes
- Types available
 - Type X (MTF)
 - Strengths
 - Concise; most effective among types
 - Useful for evaluation of factual information
 - Weakness
 - Difficult to construct flawlessly; often leads item-writers to ask for pure recall
 - A student who knows something is false; does he know what the true answer is?
 - A rather complicated scoring procedure

- Abandoned by NBME and use discouraged by most authorities
- Visualization



○ Type A (Single Best Response)

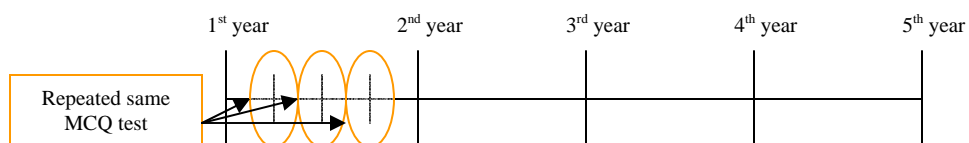
- Strengths
 - Highest reliability among the types
 - More suited to context-rich questions
 - Often easier to construct than type X
- Weakness
 - Often used to test fact
- Visualization



○ Type R (Extended Matching)

○ The Progress Test

- A strategy, not a specific method
- Overcome negative steering effect
- Regular MCQ examinations for all students
- Aimed at graduation level
- Upward progression expected
- Visualization



4. Type X: uses and advantages

- Efficient: 3 type X vs. 2 Type A in same time (Ebel & Frisbie, 1991)
- Simple and direct measure of verbal knowledge
- Most useful where there are 2 alternatives
 - Right/left, more/less
- Distinguishing
 - Fact/option, cause/effect, relevant/irrelevant, valid/invalid
- Identification of facts and definitions

5. Type X: requirement

- Stem must be clear and unambiguous
- Responses/ alternative/ branches/ options must be absolutely true or false. No shades of grey.

6. Type X: limitations (Linn & Gronlund, 1995)

- Difficult to construct unambiguous statement that measure complex learning outcomes (McAlear, 2001)
- Most field has exceptions – unsuitable type X
- Not especially useful beyond knowledge

7. Status of type X MCQ

- Distinction between true or false not clear
- To avoid ambiguity, writers are often pushed towards assessing recall of isolated facts
- There is move away from this type (McAlear 2001, Anderson, 2004)
- The NBME recommends not to use type X MCQs

8. Essay Question

- Many variations
- Traditional essay
 - Able to see 'how a mind work'
 - Problem in content specificity
 - Major problems in reliability → due to subjectivity
 - Overcoming

- Specify content (model answer)
- Multiple trained scorers
- Both difficult to achieve

9. The Modified Essay Question (MEQ)

- Consist of a patient situation given successively
- Students have to answer questions related to the situation
- Presented as a booklet/ a series of sequential questions
- Appealing in a problem-based learning
- Features:
 - Students cannot go back to the previous questions/ answers
 - Time guide provided
 - Assesses problem-solving and application of knowledge
 - In-depth knowledge can be assessed but loses in breadth of knowledge
- Disadvantages
 - Time-consuming to grade (hand-marked)
 - Content specificity
 - Greater number of invigilators needed to ensure compliance

10. Key-feature problem/ clinical reasoning exercise

- Features
 - Short, multiple scenarios
 - Variable response formats
- Promising research results for use to measure 'clinical thinking' of candidates

11. Extended Matching question (type R MCQ)

- For example:
Theme: Peripheral nerve supply to upper limb
Options:
 - a) Axillary nerve
 - b) Radial nerve

- c) Ulnar nerve
- d) Median nerve
- e) Musculocutaneous nerve
- f) Circumflex nerve
- g) Suprascapular nerve
- h) Brachial plexus nerve
- i) Posterior interosseous nerve
- j) Spinal nerve C 8

Lead-in (instruction):

For each of the given condition below, choose the most appropriate lesion to the nerve from the list above

(stems)

- Following an injury to the upper arm, the patient is unable to extend the elbow against resistance and has wrist drop. (d)
 - A patient complains of paraesthesia of the thumb, index and middle finger of the right hand. On the examination the thenar eminence is wasted. (e)
 - A patient complains of progressive inability to extend his ring and little fingers on the left hand. On examination, these fingers are held in fixed flexion. (c)
- Why choose EMQ?
 - Testing applied knowledge in contrast to recall knowledge
 - Better than MCQ because minimal cueing effect: recognitions, elimination, hints
 - Many options and options applicable to all vignettes or stems

EMQ versus multiple true false MCQ

EMQ	MTF MCQ
Applied knowledge	Inclined on recall knowledge
Testing matching and one best answer discriminatory ability	Discriminatory ability contaminated by guessing
Guessing has low correct probability	Guessing has 50-50 probability
No negative score penalty	Require negative score penalty
Can be scored by scanner	Can be scored by scanner
Wide and deep coverage based on question theme	Usually wide coverage
Integration well achieved	Integration limited achieved
New and hence difficult	Established and hence easy
Under one theme, 3 or more questions are developed	Variable topics

EMQ pros and cons

	Essay	Short answer	MCQ	EMQ
Application of knowledge	Excellent	Good	Poor	Good, can be improved with justification
Assessment tool for higher order skill	Excellent	Good	Poor	Poor to good if justification is required
Coverage of topic	Poor	Good	Excellent	Excellent
Reliability of score	Poor	Good	Excellent	Excellent
Ease of scoring	Poor	Moderate	Excellent	Excellent
Preparation time	Minimal	Moderate	Maximal	Moderate
Total cost	Large	Moderate	Low	Low
Cheating (sneak-a-peak)	Most difficult	Difficult	Easy	Easy unless justification required

- How to write EMQ?
 - Determine the theme
 - Write the options

- Write the lead-in
- Write the stems as vignettes
- Determine the theme
 - Diagnosis
 - Investigations
 - Causation
 - Management steps
 - Complications
 - Drugs
 - Clinical features
 - Prevention
 - Pathogenesis
 - Structure and function
 - Pathophysiology
 - Clinical correlation
 - Pharmacodynamics
 - Pathogens
- Write the options
 - Short
 - No verbs
 - Each options clear entity
 - 10 options required
- Write the lead-in
 - Unambiguous
 - Relate to stem and options
- Write the stem
 - Short, authentic description
 - Varying prototypically
 - Only one best answer
 - Use real life situation as a basis

- Good distribution of presentation types
- How popular is EMQ?
 - Sheffield University MBBS program
 - PLAB 1 – professional and linguistic Assessment Board
 - USMLE – US Medical Licensing Examination
 - PPSP Family Medicine MMed Program