Assessment of learning:
More than just a test

Office of Medical Education
Creighton University School of Medicine

Last Updated July 2015
Session objectives

After completing this session the learner should be able to:

• Differentiate between assessment and evaluation
• Describe common assessment types
• Create effective multiple choice questions
• Describe assessment as a process for continuous quality improvement
What is assessment?

What is evaluation?

Is there really a difference?
Definitions

The accrediting body for medical schools in the U.S. and Canada is the Liaison Committee on Medical Education (LCME). The LCME uses the following definitions:

**Assess** refers to assessment of medical student performance

**Evaluate** is reserved for faculty, resident, course and clerkship/ clerkship rotation, and program evaluation

*Source: LCME Functions and Structure of a Medical School, 2011*
Assessment Types

• Formative
• Summative
Formative Assessment

• To assess learning progress
• Often provided without a grade
• Used throughout a course or clerkship
• Intended to provide opportunities for students to ask questions and improve knowledge and skills

EXAMPLE: In-class quiz using audience response system clickers or cell phone polling
Summative Assessment

• To assess learning at the end of a curricular experience (e.g., course, clerkship)
Summative Assessment

Traditionally

Exams
- Multiple Choice Tests
- Essay Tests
- Oral Exams

Now

Performance Tests
- Labs
- Simulations
- OSCEs

Plus papers, portfolios, journals
The Assessment Process

• Assessment is more than giving an exam
• It is a process
• Like any process, there are opportunities for continuous quality improvement
Assessment Process

- Assessment Blueprint
- Item Development
- Administration And Scoring
- Item Evaluation
Create a blueprint for your assessment

Identify the learning outcomes and levels of expertise *before* you design the assessment instrument.
But first...

Ask yourself, are you assessing

Recall of memorized information?
Comprehension of material?
Application of new material?
Analysis?
Synthesis?
Evaluation?
Planning and designing an exam

- Start before the course/conference begins
- Create questions that match the objectives
- Ask colleagues to review questions
- Remember to weight more important topics
- Focus on concepts, not trivia
- Focus on the application of knowledge, not recall
Multiple Choice Questions

Strengths

• Efficient for large groups
• Can use to assess a range of thinking skills
• More responses reduces the chance of guessing correctly
• Grading is usually quick and straightforward

Weaknesses

• Difficult to write good questions
• It takes time to write and validate questions
Anatomy of the MCQ

• Stem: contains the text
• Options: all answer choices
• Key: the correct answer
• Distractor: the incorrect answers
MCQ Format used by the NBME and CUSOM

Category One: Single-best-answer formats

A (4 or more options, items or sets)
B (4 or 5 option matching, sets of 2-5 items)
R (extended matching, sets of 2-20 items)

http://www.nbme.org/about/itemwriting.asp
A 65-year-old man comes to the physician for a follow-up examination after the results of a bronchoscopy showed squamous cell carcinoma. When the physician tells the patient the diagnosis, the patient becomes tearful and responds, “No, you’re wrong! This must be a mistake. This can’t happen to me. Let’s do more tests.” This patient is most likely at which of the following stages of grief?

(A) Anger
(B) Bargaining
(C) Denial
(D) Depression

(Case and Swanson, 2004)
Rules for One-Best-Answer Items

• Each item should focus on an important concept
• Each item should assess application of knowledge, not recall of isolated fact
• The stem should pose a clear question
• All distractors should be homogeneous
• Avoid technical item flaws the cue “testwise” students or pose irrelevant difficulty

(Case and Swanson, 2004)
Best Evidence for Writing MCQs

- Invest in the stem; make sure the item can be answered without looking at the options.
- Include language in the stem instead of repeating it in each option.
- Each item should be clear and plausible.
- Avoid using different grammar (e.g., verb tense) in stem and distractor.
- Avoid negatively phrased items, definitive (e.g., always), and vague (e.g., many) language.
- Avoid trickery and irrelevance.
More MCQ Tips

• Randomly distribute the correct answer position throughout the test
• Present the options in a logical order (e.g., alpha, chronological)
• Allow 1-2 minutes per question, depending on the reading time
• Instructor should take exam first, then double or triple the time estimate
One Best Answer: Extended Matching

This format is used occasionally for some exams

4 Components of a good EMQ:

1. A theme;
2. An option list;
3. A lead-in statement;
4. At least two item stems.

(Case and Swanson, 2004)
Retired MCQ Formats

*No longer used by the NBME or Creighton

Category Two: True/False

• Negative A questions, e.g., “Each of the following is correct except...”
• True/False formats - select all true options
  • C (A/B/Both/Neither items)
  • X (simple true/false)
  • K (complex true/false where one or more options may be true)
For more MCQ info

- Consult the guide “Test Question Formats for M1 and M2 Quizzes and Exams at Creighton University School of Medicine” that is posted in this module folder

- Use the online NBME tutorial
  http://download.usmle.org/IWTutorial/intro.htm

- Review the NBME item writing guide available here:
  http://www.nbme.org/about/itemwriting.asp

- Contact the Office of Medical Education
Testing Process

Item Evaluation → Test Blueprint → Test Administration And Scoring → Item Development
The next step: Evaluating the test

- Examine reliability (consistency) and validity (did you assess or measure what you intended)
- Examine response patterns
- Examine item quality
Item Difficulty

• Proportion of examinees selecting each response
• Determine item difficulty (0.00-1.00)
• Easy = 85% answer correctly
• Difficult = 35% answer correctly
• Item analysis will show you this statistic for each item

Consult your Component Director or the Office of Medical Education for exam statistics
Item Discrimination

• Point Biserial Correlation indicates correlation (+1 to -1) between whether a student selected a particular alternative and the student’s total score on the test
• Large, positive RPBI indicates students with higher scores answered correctly
• Low, positive (<.20) suggest ability not related to success
• Negative indicates that low-achieving students performed better than high-achieving students
Finally, take a step back and reflect on the exam overall

A difficult question (p=.35) does not mean it is a good question!

If RPBI = .18 (i.e., low positive) for the same item, you should review distractor statistics or delete item entirely.
Now that the exam is over

• Take time to review exam items
• Flag items that did not perform well
• Talk to the course or component director to learn why items were flawed
• If an item was too difficult, revisit your lecture materials and update for the next iteration of your lecture
Close the loop

- Item Evaluation
- Test Blueprint
- Test Administration And Scoring
- Item Development
References and Resources


For further reading


Gronlund N. Writing Instructional Objectives for Teaching and Assessment. 7th ed. Upper Saddle River, New Jersey. Merrill Prentice Hall, 2004


http://scoring.msu.edu/writitem.html#intro