Modified Bloom's Taxonomy for Evaluating Multiple Choice Questions[1,2,3]

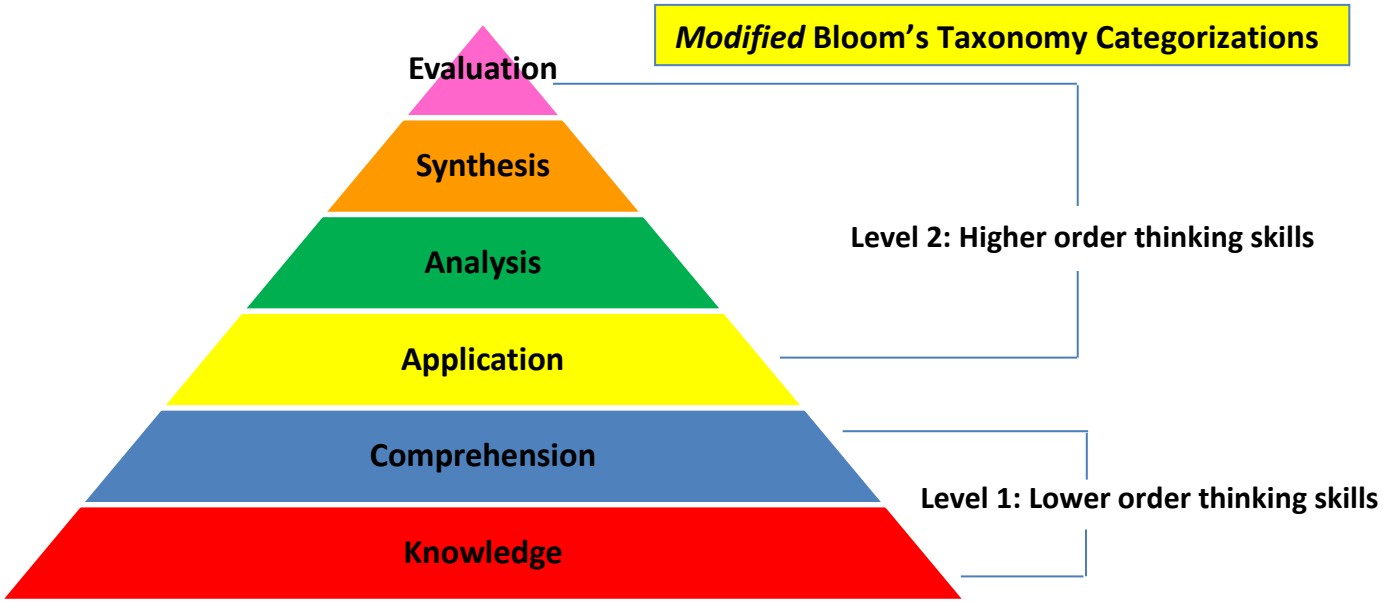

Figure above represents the original Bloom's Taxonomy which is traditionally shown in a pyramid with the lower cognitive levels at the base and higher cognitive levels at the top. Modified Bloom's taxonomy categorizations are outlined to the right of the pyramid.

Modified Bloom's Categorization	Definition	Example MCQ Stems[4]
Level 1: Lower order thinking skills	Questions require the learner to recall or comprehend learned information *without* application of a concept.	Cytogenetic abnormality is best explained by which of the following mechanisms?* Which of the following drugs can cause pain, vomiting, and jaundice?* (*NOTE: Question stems adapted from sample NBME test questions)
Level 2: Higher order thinking skills	Questions require the learner to use prior knowledge to apply new information by incorporating relevant rules, methods, concepts, principles and theories; understand components of a concept; identify relationships between information; and analyze, synthesize and/or evaluate information.	During an experiment, an investigator finds a point mutation (CGG→TGG) in the gene encoding the protein kinase regulatory subunit. The genetic code is shown in the figure. This mutation is most likely to alter the amino acid sequence with which of the following changes? A 12-year-old girl with sickle cell disease has pain in her right arm. An x-ray of the right upper extremity shows bony lesions consistent with osteomyelitis. Which of the following is the most likely causal organism?

Resources

[1] Bloom B. *The Taxonomy of Educational Objectives Handbook 1: Cognitive Domain*. 2nd ed. New York, NY: Addison Wesley; 1984.

[2] Knecht, KT. Assessing cognitive skills of pharmacy students in a biomedical sciences module using a classification of multiple-choice item categories according to Bloom's taxonomy. *Am J Pharm Educ*, 2001; 65(4): 324-334.

[3] Palmer, EJ, & Devitt, PG. Assessment of higher order cognitive skills in undergraduate education: modified essay or multiple choice questions? *BMC Med Educ,* 2007; 7(49).

[4] National Board of Medical Examiners . *USMLE Step 1 Sample Test Questions;* 2014. http://www.usmle.org/step-1/. Accessed June 2, 2014.

Version 1, 6-6-14, prepared by Nikki Zaidi, Ph.D., UMMS Evaluations & Assessment Unit, bibler@med.umich.edu

Type of Question	Definition	Example	Recommendation
Lower Order	Lower order questions require the learner to recall or comprehend learned information *without* application of a concept.	Cytogenetic abnormality is best explained by which of the following mechanisms?	**M1 exams:** < 60% **M2 exams:** < 50%
Higher Order	Higher order questions require the learner to use prior knowledge to apply new information by incorporating relevant rules, methods, concepts, principles and theories; understand components of a concept; identify relationships between information; and analyze, synthesize and/or evaluate information.	During an experiment, an investigator finds a point mutation (CGG→TGG) in the gene encoding the protein kinase regulatory subunit. The genetic code is shown in the figure. This mutation is most likely to alter the amino acid sequence with which of the following changes?	**M1 exams:** minimum of 40% **M2 exams:** minimum of 50%
Case-Based	Case-based questions use a vignette –a clinical scenario or lab scenario—to provide examinees all the information needed to respond to the question.	A 12-year-old girl with sickle cell disease has pain in her right arm. An x-ray of the right upper extremity shows bony lesions consistent with osteomyelitis. Which of the following is the most likely causal organism?	**M1 exams:** minimum of 40% **M2 exams:** minimum of 50%
Integrated	Integrated questions combine content from multiple domains and/or disciplines into a single test item.	A 3629-g (8-lb) male newborn is delivered at term. During the initial examination, urine is found to be leaking from the umbilicus. This patient most likely has an abnormality of which of the following fetal structures?	**M1 exams:** minimum of 10% **M2 exams:** minimum of 20%

General Guidelines for Constructing Written Test Items[1,2]

Guideline	Rationale
Item stems should pose a *single*, clearly formulated problem. Response options should provide *no* additional data.	An examinee should be able to respond to the question without referring to the response options. An ideal format: "[Insert clinical vignette]. What is the most likely diagnosis?"
Avoid using "all of the above" as a response option.	Examinees can easily eliminate this option by identifying just one incorrect answer.
Avoid using "none of the above" as a response option.	This option adds unnecessary ambiguity. It is best practice to replace "none of the above" with an option that is more specific such as "no drug should be given at this time."
Avoid negatively phrased questions such as: "Each of the following is correct EXCEPT" and "Which of the following statements is NOT correct?"	Response options are seldom homogenous; therefore, examinees struggle to assess the level of correctness for each option.
Numeric data should be consistent. a. 20 -**OR-** a. 10%-20% -**NOT-** a. 25 b. 30 b. 30%-40% b. 30-35 c. 50 c. 60%- 70% c. 40-70 d. 100 d. 85%-95% d. 75	Confusion occurs when formats are mixed (i.e. ranges and single data points are used in same set of response options). Examinees will also struggle when options are listed in an illogical or an inconsistent format.
Avoid using ambiguous language such as imprecise phrases and/or vague frequency terms. Imprecise phrases→*is associated with; is useful for;* and *is important* Vague terms→*may, usually,* and *rarely*	Imprecise phrases and vague frequency terms are not consistently defined or interpreted—even by experts. For instance, the term "usually" can mean 70% of the time to some examinees while others will interpret this term to mean higher or lower estimates.
Avoid extreme words such as: *all, always, never, only, nothing,* and *alone*.	Examinees know these words usually signify incorrect response options.
All response options should be approximately the same length.	Examinees know that correct answers are often the longest or the shortest option.
Ensure that all incorrect response options (distractors) follow grammatically and logically from the question stem.	If the question stem ends with the word "an" and only one or two response options begin with a vowel, then the student can easily eliminate options.
Avoid repeating words or phrases that are included in both the item stem and in the correct answer.	Examinees will benefit from cues such as stating "liver" in the item stem and then repeating "hepato" in the correct answer.

Anatomy of a Good Multiple Choice Test Item[1, 2]

Item Stem = **Vignette + Lead-in**

A 21-year-old man has weight loss and severe intermittent bloody diarrhea. Barium enema and colonoscopy show multiple ulcers and inflammatory changes extending from the rectum to the mid-transverse colon. Biopsy specimens taken from multiple sites show acute and chronic inflammation restricted to the mucosa.

The ***vignette*** is a patient scenario (or lab scenario) that provides the examinee **all** the information needed to answer the question.

Which of the following is the **most likely** diagnosis?

The ***lead-in*** should be the final sentence of the stem. It poses the question to be answered.

Response Options = **Key + Distractors**

A. Amebiasis
B. AIDS-associated gastroenteritis
C. Crohn's disease *
D. *Clostridium difficile*-associated colitis
E. *Salmonella* gastroenteritis

* Correct response ("key")

The correct answer is called the "***key***." The key should be most correct to avoid ambiguity.

Distractors are incorrect answers. The distractors should present plausible options and be comparable in length to the correct answer.

Examples of Good Basic Science Questions

I. Several contiguous cells are labeled with a fluorescent dye that cannot cross cell membranes. One cell is experimentally bleached with light that destroys the dye, but soon recovers dye fluorescence. This recovery is **best** explained by the presence of which of the following structures between the bleached cell and its fluorescent neighbors?

A. A basal lamina
B. Desmosomes (maculae adherents)
C. Glycosaminoglycans
D. Gap junctions*****
E. Tight junctions (zonulae occludentes)

This ***lab vignette*** presents a lab experiment and the accompanying ***lead-in*** question asks the examinee to use basic science principles to explain the results.

II. A 65-year- old man has difficulty rising from a seated position and straightening his trunk, but he has no difficulty flexing his leg. Which of the following muscles is **most likely** to have been injured?

A. Gluteus maximus*****
B. Gluteus minimus
C. Hamstrings
D. Iliopsoas
E. Obturator internus

A ***clinical vignette*** is used for this basic science question; however, the examinee should be able to answer the ***lead-in*** question based on an understanding of basic science (i.e. anatomy).

Resources

[1]National Board of Medical Examiners. *Constructing Written Test Questions for the Basic and Clinical Sciences.* 3rd ed.; 2003. http://www.nbme.org/publications/item writing-manual.html. Accessed August 1, 2013

[2]National Board of Medical Examiners. *Writing Multiple Choice Questions: An Introductory Tutorial* [Video file]; 2012. http://www.nbme.org/IWTutorial. Accessed August 1, 2013. **NOTE**: This online tutorial was created by the NBME in collaboration with UMMS Professor N. Cary Engleberg, M.D., Departments of Internal Medicine and Microbiology & Immunology.

Version 1, 8-13-13, prepared by Nikki Zaidi, Ph.D., UMMS Evaluations & Assessment Unit, bibler@med.umich.edu

Principles/Guidelines for Trunk Assessments

Basic Principles for Test Construction and Grading

- All test questions should correlate with clearly stated learning objectives; learning objectives should drive exam content, rather than item quotas based on lecture hours.
- Mandatory curricular sessions should include testable content when appropriate.
- Test questions should integrate information from more than one content domain or area of knowledge. Integrated questions stimulate higher order thinking skills which will train students for complex clinical reasoning.

Blueprint of Tests

- Learning objectives should be used to create assessment items.
- Sequence-based quizzes and exams should include Doctoring and other domains' content when appropriate. For 2015-16 these will be scored separately except in the fused sequences. For 2016-17 there will be integrated scoring of exams.
- Aim: M1 exams–minimum of 40% case-based items, 40% higher-order items, 10% integrated items.
- Aim: M2 exams– minimum of 50% case-based items, 50% higher-order items, 20% integrated items.

Test Item Construction (Quality Control)

- The item stem should pose a clear question that can be answered without the response options.
    - I.e. All items should pass the "cover test"— can examinees pose an answer to the question based solely on the question stem?
- No "all/none of the above" items, including K-type questions, should be used.
    - K-type questions use multiple choices A, B, C, D, and ask the examinee to select from the following combination options 1) A only; 2) both A and C; 3) both B and D; 4) A, B and C, 5) All [or None] of the Above.
- No negatively-worded items should be used *unless* the purpose of the test question is truly to determine what examinees DO NOT know.
- All questions should have meaningful "feedback" provided by the question writer.

Process for Creating Quizzes and Exams

- The Evaluation and Assessment (E & A) Office will be available to consult with Sequence Directors and other Sequence faculty to help implement best practices in item-writing and overall test construction.
- Sequence Directors are responsible for submitting all quizzes and exams to the Evaluation and Assessment Office in one complete packet a minimum of *two weeks prior* to the sequence start date.
- Quizzes/exams will be ready for review at least one week in advance of the assessment sign-on date.
- Final version of all quiz/exams and corresponding answer key should be approved by Sequence Director.

Grading Procedures

- Grading component structure should be clearly defined and implemented as defined in syllabus.
- Poorly performing items should be considered in grading adjustments. Examples of poorly performing questions include:
    - Negative discrimination index—this would suggest that the highest quartile of students performed worse that the lowest quartile of students.
    - Difficulty index ≤ 0.50 (i.e. 50% or less of students answered the question correctly) and discrimination index < 0.30.
    - Valid student queries indicating misalignment between learning objectives and tested content.