Writing NBME style multiple choice exam questions

Anthony Payne, PhD
Department of Physiology
MCG at AU
Learning Objectives

By the end of this workshop, you will be able to:

1. Explain the role of Learning Objectives in writing appropriate exam questions for medical students.

2. Create multiple choice questions for medical students in the general format of the National Board of Medical Examiners.

3. Recognize and repair common flaws in writing multiple choice questions for medical students.
Outline

• What are we testing?
  • Learning Objectives.

• Utilizing the NBME **Constructing Written Test Questions for the Basic and Clinical Sciences** manual – Chapters 1-4

• Question Formats

• Technical Item Flaws – What not to do.

• Basic Rules for writing one-best-answer MCQs – What to do.

• Workshop – small groups
What Are We Testing?

• When we write an exam, what are we asking the student to do?

  • Repeat memorized facts?
  • Use mathematical equations?
  • Apply a principle in a realistic context?
  • Display knowledge of how to manage a patient with a complex disease and complicating psychosocial factors?

• What do you want the student to be able to DO after you teach them?
Writing Learning Objectives

• LOs are written to inform the learner what the ultimate goal is.
• What do you want the student to DO after learning your material?

• LO Domains:
  • One of 3 domains: Knowledge, Skills, Attitudes
  • Pre-clinical years may focus more on knowledge. Clerkships and Residency ramp up the skills.

• LO Elements:
  • ABCDs of LOs: Audience, Behavior, Condition, Degree

• LO Cognitive Levels:
  • See Bloom’s Taxonomy
Writing Learning Objectives

Bloom’s Taxonomy

- **Understand**
  - Explain ideas or concepts
    - classify, describe, discuss, explain, identify, locate, recognize, report, select, translate
  - Recall facts and basic concepts
    - define, duplicate, list, memorize, repeat, state

- **Apply**
  - Use information in new situations
    - execute, implement, solve, use, demonstrate, interpret, operate, schedule, sketch

- **Analyze**
  - Draw connections among ideas
    - differentiate, organize, relate, compare, contrast, distinguish, examine, experiment, question, test

- **Evaluate**
  - Justify a stand or decision
    - appraise, argue, defend, judge, select, support, value, critique, weigh

- **Create**
  - Produce new or original work
    - design, assemble, construct, conjecture, develop, formulate, author, investigate

- **Remember**
  - Recall facts and basic concepts
    - define, duplicate, list, memorize, repeat, state
Writing Learning Objectives

Bloom’s Ranking of Thinking Skills

<table>
<thead>
<tr>
<th>REMEMBER</th>
<th>UNDERSTAND</th>
<th>APPLY</th>
<th>ANALYZE</th>
<th>EVALUATE / SYNTHESIZE</th>
<th>CREATE</th>
</tr>
</thead>
<tbody>
<tr>
<td>List, Name, Identify, Label, Show, Define, Recognize, Recall, State,</td>
<td>Choose, Summarize, Explain, Describe, Compare, Discuss, Differentiate, Demonstrate, Classify</td>
<td>Solve, Calculate, Illustrate, Use, Interpret, Relate, Manipulate, Apply, Modify</td>
<td>Analyze, Organize, Deduce, Contrast, Compare, Distinguish, Diagnose, Devise a Plan, Solve</td>
<td>Design an Experiment, Hypothesize, Predict, Support, Schematize, Write, Report, Justify</td>
<td>Appraise, Argue, Critique, Defend, evaluate, Judge, Revise, Validate, Test</td>
</tr>
</tbody>
</table>

The words “know”, “understand”, and “learn” are nowhere in these lists.
Writing Learning Objectives

• LOs are written to inform the learner what the ultimate goal is.
• What do you want the student to DO after learning your material?

• Learn the 50 States on a map of the United States. (*knowledge domain*).
• Given a map of the United States (*condition*), the student (*audience*) will be able to correctly label (*behavior*) all 50 States in 20 minutes (*degree*). (*Level: remember*)
Writing Learning Objectives

• **Example:** *Knowledge Domain*
  
  Know how arterial blood gases are altered in the various acid-base disturbances.

• **OR**
  
  Given a patient’s arterial blood gas readings *(condition)*, correctly *(degree)* diagnose the type of acid-base disturbance *(behavior)*.
    
    • *Level: Analyze*

  Given a patient’s arterial blood gas readings with an acid-base disturbance, predict the physiological compensation that will occur.
    
    • *Level: synthesize/evaluate*
Writing Learning Objectives

• **Example:**

  • Perform a “head-to-toe” examination of a standardized patient.

• **OR**

  • In a 15-minute station with a standardized patient, the student will accurately perform a “head-to-toe” physical examination.

  • In a 15-minute station with a standardized patient (*condition*), the student (*audience*) will accurately (*degree*) perform a “head-to-toe” physical examination (*behavior*).

• **Domain:** Skills (and knowledge)

• **Level:** Mechanism (*Mid-Level psycho-motor skills*)
Writing Learning Objectives

• Example:

• Compare and contrast malabsorption-induced diarrhea, osmotic diarrhea, and secretory diarrhea. Name the disorders that can cause each type of diarrhea and the diagnostic tests used to distinguish between them.

• Given the clinical presentation of a patient with diarrheal disease, predict the most likely pathophysiological mechanism (secretory, malabsorption, osmotic) involved and recommend the correct diagnostic tests used to determine the underlying cause(s).

• Compare and contrast diarrheal diseases cause by enterohemorrhagic E. coli (E. coli 0157:H7) and enterotoxigenic E. coli. Discuss the molecular pathogenesis involved.

• In a patient with infectious diarrhea, differentiate whether the diarrhea is caused by enterohemorrhagic E. coli (E. coli 0157:H7) or enterotoxigenic E. coli. Explain the molecular pathogenesis of these two microorganisms.
Writing Learning Objectives

• **Resources**


Writing Learning Objectives

• Workshop time

• Pair up
Required elements for an MCQ
1. Important content. See your LOs.
2. Realistic scenario.
3. Proper question structure.
Question Formats

• True/False vs. One-best-answer

A, C, D, E

• T/F

<table>
<thead>
<tr>
<th>Totally Wrong Options</th>
<th>Totally Correct Options</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>E</td>
<td></td>
</tr>
</tbody>
</table>

• OBA

<table>
<thead>
<tr>
<th>Totally Wrong Options</th>
<th>Totally Correct Options</th>
</tr>
</thead>
<tbody>
<tr>
<td>A, E, C, D</td>
<td>B</td>
</tr>
</tbody>
</table>
Question Format – T/F

• T/F questions disguised as MCQs...

• Which of the following statements about blood pressure is true?

• Which of the following statements about blood pressure is false?

• Stay away from these...

• All of the following are true, EXCEPT:

• All of the following are false, EXCEPT:
Which of the following is an X-linked recessive condition?

A. Hemophilia A (classic hemophilia)
B. Cystic fibrosis
C. Phenylketonuria
D. Tay-Sachs disease

B, C, D  A
True statements about cystic fibrosis (CF) include:
1. The incidence of CF is 1:2000.
2. Children with CF usually die in their teens.
3. Males with CF are sterile.
4. CF is an autosomal recessive disease.

• Flawed question....
• 4- is true.
• 1- in which population, which country?
• 2- in 2016? In the USA?
• 3- not 100% true
Question Format – T/F

In children, ventricular septal defects are associated with

1. systolic murmur
2. pulmonary hypertension
3. tetralogy of Fallot
4. cyanosis

Flawed question....

• How severe is the disease?
• How old is the patient?
• Has the disease been treated?
• Again, all can be true under certain circumstances.
A 32-year-old man has a 4-day history of progressive weakness in his extremities. He has been healthy except for an upper respiratory tract infection 10 days ago. Temperature 37.8°C (100°F), BP 130/80 mmHg, Pulse 94/min, RR 42/min and shallow. He has symmetric weakness of both sides of the face and of the proximal and distal muscles of the extremities. Sensation is intact. No deep tendon reflexes can be elicited; plantar responses are flexor. Which of the following is the most likely diagnosis?

A. Acute disseminated encephalomyelitis
B. Guillain-Barre syndrome
C. Myasthenia gravis
D. Poliomyelitis
E. Poliomyositis
Question Format – One-Best-Answer

**Stem:** A 32-year-old man has a 4-day history of progressive weakness in his extremities. He has been healthy except for an upper respiratory tract infection 10 days ago. Temperature 37.8°C (100°F), BP 130/80 mmHg, Pulse 94/min, RR 42/min and shallow. He has symmetric weakness of both sides of the face and of the proximal and distal muscles of the extremities. Sensation is intact. No deep tendon reflexes can be elicited; plantar responses are flexor.

**Lead-in:** Which of the following is the most likely diagnosis?

**Options:**
A. Acute disseminated encephalomyelitis
B. Guillain-Barre syndrome
C. Myasthenia gravis
D. Poliomyelitis
E. Poliomyositis
• Not all the answer choices are completely wrong.
• Answers exist on a spectrum of most correct to least correct.

A. Acute disseminated encephalomyelitis
B. Guillain-Barre syndrome
C. Myasthenia gravis
D. Poliomyelitis
E. Poliomyositis
Question Format – One-Best-Answer

• Why was that a “good” question?

• Important concept
• Requires *application* of knowledge
• Stem provided enough information, without too much irrelevant info
• Lead-in asked a clear question
• Cannot be answered by the lead-in only
• All distractors are homogenous (same category as correct answer – diagnoses, tests, treatments, prognoses, etc).
• Passes the “hand test”
• Avoids other technical item flaws
The “hand test”

Can you cover the answers with your hand?

Is there enough information in the stem and question to answer it without even seeing the answer choices?
Technical Item Flaws

• Issues related to Testwiseness
  • Can be answered without the stem
  • Grammatical cues
  • Logical cues
  • Absolute terms
  • Long correct answer
  • Word repeats
  • Convergence strategy

• Issues related to Irrelevant Difficulty
  • Options are long, complicated, or double
  • Numeric data not consistent
  • Frequency terms are vague
  • Language in the options is not parallel or in illogical order
  • “Teaching” in the stem
Technical Item Flaws – Testwiseness

• Can be answered without the stem

• A 10-year-old girl was diagnosed with juvenile idiopathic arthritis, and prescribed a 1 month course of prednisone for treatment. After follow-up, it is decided that the patient has responded well and she can begin tapering off the prednisone. Why should patients taper off prednisone instead of stopping immediately?
Technical Item Flaws – Testwiseness

• Grammatical Cues

• A and C do not fit grammatically and can be immediately eliminated as answer choices.

A 60-year-old man is brought to the emergency department by the police, who found him lying unconscious on the sidewalk. After ascertaining that the airway is open, the first step in management should be intravenous administration of

A. examination of cerebrospinal fluid
B. glucose with vitamin $B_1$ (thiamine)
C. CT scan of the head
D. phenytoin
E. diazepam
Technical Item Flaws – Testwiseness

• Logical Cues

• A, B, and C cover all the possibilities

• D and E are likely due to wanting 5 options

Crime is

A. equally distributed among the social classes
B. overrepresented among the poor
C. overrepresented among the middle class and rich
D. primarily an indication of psychosexual maladjustment
E. reaching a plateau of tolerability for the nation
Technical Item Flaws – Testwiseness

• Absolute Terms

• C and D are out

In patients with advanced dementia, Alzheimer’s type, the memory defect

A. can be treated adequately with phosphatidylcholine (lecithin)
B. could be a sequela of early parkinsonism
C. is never seen in patients with neurofibrillary tangles at autopsy
D. is never severe
E. possibly involves the cholinergic system
Technical Item Flaws – Testwiseness

• Long Correct Answer

• C is much longer, includes qualifiers

Secondary gain is

A. synonymous with malingering
B. a frequent problem in obsessive-compulsive disorder
C. a complication of a variety of illnesses and tends to prolong many of them
D. never seen in organic brain damage
Technical Item Flaws – Testwiseness

- Word Repeats
- “Unreal” and “Derealization”

A 58-year-old man with a history of heavy alcohol use and previous psychiatric hospitalization is confused and agitated. He speaks of experiencing the world as unreal. This symptom is called

A. depersonalization
B. derailment
C. derealization
D. focal memory deficit
E. signal anxiety
Technical Item Flaws – Testwiseness

• Convergence Strategy
• The answer with the most in common with the others is likely to be the correct choice

• “Anionic” once – eliminate
• “outside” twice, “inside” three times – eliminate outside
• Charged form 3, uncharged 2
• Therefore, answer is B

Local anesthetics are most effective in the
A. anionic form, acting from inside the nerve membrane
B. cationic form, acting from inside the nerve membrane
C. cationic form, acting from outside the nerve membrane
D. uncharged form, acting from inside the nerve membrane
E. uncharged form, acting from outside the nerve membrane
Technical Item Flaws – Testwiseness

• Convergence Strategy

• Writers often start with the correct answer and create permutations of it that are wrong.

• Numerical answers are usually the middle one, not the highest or lowest.
Technical Item Flaws – Testwiseness

How does the baroreceptor reflex alter heart rate, heart contractility, and arteriole diameter in response to a sudden drop in mean arterial pressure?

<table>
<thead>
<tr>
<th></th>
<th>HR</th>
<th>Contr</th>
<th>Art Diam</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>↑</td>
<td>↑</td>
<td>↑</td>
</tr>
<tr>
<td>B</td>
<td>↑</td>
<td>↑</td>
<td>↓</td>
</tr>
<tr>
<td>C</td>
<td>↑</td>
<td>↓</td>
<td>↑</td>
</tr>
<tr>
<td>D</td>
<td>↑</td>
<td>↓</td>
<td>↓</td>
</tr>
<tr>
<td>E</td>
<td>↓</td>
<td>↑</td>
<td>↑</td>
</tr>
<tr>
<td>F</td>
<td>↓</td>
<td>↑</td>
<td>↓</td>
</tr>
<tr>
<td>G</td>
<td>↓</td>
<td>↓</td>
<td>↑</td>
</tr>
<tr>
<td>H</td>
<td>↓</td>
<td>↓</td>
<td>↓</td>
</tr>
</tbody>
</table>

What if we also include ↔ as a potential variable?
Technical Item Flaws – Irrelevant Difficulty

• “Teaching” in the stem

Drugs such as methotrexate, hydroxychloroquine, or penicillamine are often turned to for managing rheumatoid arthritis that is not controlled adequately with “traditional” NSAIDs (e.g., aspirin, ibuprofen, or indomethacin). Which of the following statements best summarizes how those drugs differ from a typical NSAIDs?

• If the material was taught, it does not need to be reviewed in the stem.
• If it needs to be reviewed in the stem, then it was not taught and should not be tested.
Technical Item Flaws – Irrelevant Difficulty

• Numeric data not consistent

• A, B, and C are expressed as ranges

• D and E are specific percentages.

• All options should either be range or specific

Following a second episode of infection, what is the likelihood that a woman is infertile?

A. Less than 20%
B. 20 to 30%
C. Greater than 50%
D. 90%
E. 75%
Technical Item Flaws – Irrelevant Difficulty

• Vague frequency terms

Severe obesity in early adolescence
   A. usually responds dramatically to dietary regimens
   B. often is related to endocrine disorders
   C. has a 75% chance of clearing spontaneously
   D. shows a poor prognosis
   E. usually responds to pharmacotherapy and intensive psychotherapy
Technical Item Flaws – Irrelevant Difficulty

- Unnecessarily complicated or tricky

Arrange the parents of the following children with Down’s syndrome in order of highest to lowest risk of recurrence. Assume that the maternal age in all cases is 22 years and that a subsequent pregnancy occurs within 5 years. The karyotypes of the daughters are:

I: 46, XX, -14, +T (14q21q) pat
II: 46, XX, -14, +T (14q21q) de novo
III: 46, XX, -14, +T (14q21q) mat
IV: 46, XX, -21, +T (14q21q) pat
V: 47, XX, -21, +T (21q21q) (parents not karyotyped)

A. III, IV, I, V, II
B. IV, III, V, I, II
C. III, I, IV, V, II
D. IV, III, I, V, II
E. III, IV, I, II, V
Summary of What to Do

• Match to a learning objective.
• Test higher order thinking when possible.
• Include as much in the stem as possible. Long stem, short options (see next slide).
• Avoid superfluous, overly complicated, tricky.
• Write grammatically consistent options that are logically compatible with the stem.
• List options in logical or alphabetical order.
• Options should be as similar in length as possible.
• Avoid absolutes (always, never, all). Avoid vague terms (usually, frequently).
• Avoid negative questions (except, not true).
MCQ Structure Instructions

1. Patient’s age and gender
2. Chief complaint/presenting symptoms (including nature of onset)
3. Duration of Symptoms (include changing nature of symptoms)
4. Pertinent history (personal, family, social)
5. Examination findings
6. Pertinent labs
7. Lead in Statement (try not to use NOT or LEAST LIKELY). Statement should be focused so students attempt to answer the question with the answer choices covered up ("The Hand Test").
8. Option set (does not have to be limited to 4-5). Do not use “none of the above” or “all of the above.”
Examples

The compensation for acute respiratory acidosis most likely includes which of the following?

a. Decrease in alveolar ventilation  
   b. Increase in ventilation rate  
   c. Reabsorption of hydrogen by the kidneys  
   d. Renal retention of bicarbonate  
   e. Retention of CO2 by the lungs

A patient’s arterial blood gases reveal: pH 7.27, pO2 54, pCO2 60, HCO3- 28. What will the compensation be for this patient’s current acid-base disorder?

a. Decrease in alveolar ventilation  
   b. Increase in ventilation rate  
   c. Reabsorption of hydrogen by the kidneys  
   d. Renal retention of bicarbonate  
   e. Retention of CO2 by the lungs

More advanced question would include a clinical scenario describing an acute respiratory acidosis with lab values.
A 31-year-old woman presents to the ED after experiencing several episodes of sweating, anxiety, heart palpitations, and headache. Her blood pressure is found to be 180/100 mmHg and remains almost unchanged despite aggressive antihypertensive therapy. A renal ultrasound shows no stenosis or compression of the renal artery, but a retroperitoneal mass is seen. High-resolution CT scan reveals a 6.3 x 5.2 x 3.4 cm mass that appears to be arising from the left adrenal gland. Lab findings are remarkable for elevated urine metanephrines. During excision of the mass, several enlarged lymph nodes and suspicious nodules within the abdominal cavity are also biopsied. Which of the following histologic features is/are most consistent with a malignancy of the excised tumor?

- Necrosis involving >25% of the tumor
- >5 mitotic figures per high-power microscopic field
- >15 mitotic figures per high-power microscopic field
- Tumor cells within a lymph node
- Presence of many large, bizarre nuclei and focal solid areas with loss of typical “zellballen” pattern

61% got this correct. Tough question.
Discrimination Index = 78%

Was this a good question???
A 50-year old man visits his physician for an annual exam. He complains of mild shortness of breath upon exertion. Upon cardiac exam, the physician detects an S3 heart sound and notes a laterally displaced point of maximal impulse. An echocardiogram on this patient is most likely to reveal which of the following?

<table>
<thead>
<tr>
<th>Condition</th>
<th>Score</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Left atrial enlargement</td>
<td>3</td>
<td>2.1%</td>
</tr>
<tr>
<td>Left ventricular dilation</td>
<td>80</td>
<td>55.94%</td>
</tr>
<tr>
<td>Left ventricular hypertrophy</td>
<td>55</td>
<td>38.46%</td>
</tr>
<tr>
<td>Right atrial enlargement</td>
<td>0</td>
<td>0%</td>
</tr>
<tr>
<td>Right ventricular dilation</td>
<td>3</td>
<td>2.1%</td>
</tr>
<tr>
<td>Right ventricular hypertrophy</td>
<td>2</td>
<td>1.4%</td>
</tr>
</tbody>
</table>

Average Grade: 0.56 / 1 (55.94%)
Standard Deviation 49.82%
Point Biserial 0.58
Discrimination Index 76.32%
Data Analysis

After the exam is done, how do we examine the performance of an exam question?

**Question 1**  
Difficulty: 1

A newly married couple with a family history of hemochromatosis presents to a genetic counselor at Augusta University Medical Center. Genetic analysis reveals that the husband is homozygous for the HFE C282Y mutation and the wife is heterozygous for the HJV G320V mutation. What is the most likely risk of this couple having a child who could develop hereditary hemochromatosis?

- 0%, because the two carry different amino acid changes [2 (1.05 %)]
- 0%, because the mutations are on different loci [153 (80.1 %)]  
- 50%, because both mutations cause hemochromatosis [31 (16.23 %)]
- 50%, only if the child is male [2 (1.05 %)]
- 100%, only if the child is male [0 (0 %)]
- 100%, either male or female child [3 (1.57 %)]

**Average Grade:** 0.8 / 1 (80.1 %)  
**Standard Deviation:** 40.05 %  
**Point Biserial:** 0.42  
**Discrimination Index:** 41.18 %

**Question 2**  
Difficulty: 1

A 20-year-old previously healthy man presents to the emergency department with a several weeks' history of fatigue, increased thirst, and increased urination. He reports that he has lost 30 pounds over the past 2 months despite having an increased appetite. He denies any night sweats, easy bruising or bleeding, blood in stools, diarrhea, constipation, nausea, vomiting, or abdominal pain. His vital signs are stable and physical exam is unremarkable. His BMI is 20. His complete blood count and basic metabolic panel are normal with the exception of a serum glucose of 389 mg/dL. What is the most likely etiology of his weight loss?

- New onset diabetes mellitus type 2 [43 (22.51 %)]  
- Acute lymphocytic leukemia [0 (0 %)]
- Inflammatory bowel disease [0 (0 %)]
- Hemochromatosis related diabetes [33 (17.28 %)]  
- New onset diabetes mellitus type 1 [115 (60.21 %)]

**Average Grade:** 0.6 / 1 (60.21 %)  
**Standard Deviation:** 49.08 %  
**Point Biserial:** 0.47  
**Discrimination Index:** 52.94 %
Data Analysis

Determines how good the question was at discerning high-performing students from low-performing students. But How???
Data Analysis – Discrimination Index

What if only 50% of the class gets a question correct?

Discrimination Index (DI) = top 27% - bottom 27%
Example: P=0.50. 90% 35%
DI = 0.90 – 0.35 = 0.55

Tough question, but good discriminator.

What if...
Example: P=0.50 52% 48%
DI = 0.52 – 0.48 = 0.04

Tough question. Poor discriminator. We must examine this closely.
Pair up.

Look at the questions in the handouts on the table.

Work on identifying flaws, if any.
Work on improving the question – clinical scenarios in stems, lead-in questions, answer choices, etc.