

**Constructing Written Test Questions for the Basic and Clinical Science
Third Edition-National Board of Medical Examiners**

Complete Issue Available at <http://www.nbme.org/publications/item-writing-manual-download.html>

Purposes of Testing

- Communicate to students what material is important
- Motivate students to study
- Identify areas of deficiency in need of remediation or further learning
- Determine final grades or make promotion decisions
- Identify areas where the course/curriculum is weak

What Should Be Tested?

- Exam content should match course/clerkship objectives
- Important topics should be weighted more heavily than less important topics
- The testing time devoted to each topic should reflect the relative importance of the topic
- The sample of items should be representative of the instructional goals

Multiple Choice-Item Formats and Structure

Multiple-choice questions (MCQ's) can be divided into two families of items: those that require the examinee to indicate all responses that are appropriate (true/false) and those that require the examinee to indicate a single response (one best answer).

True/false-item formats require that examinees select all options that are true

- A / B/ Both/ Neither items
- Complex true/false items
- Simple true/false items
- Simulations such as Patient Management Problems (PMPs)

One-best-answer item formats require that examinees select the single best response

- 4 or more options, single items or sets
- 4 or 5 option matching items in sets of 2-5 items
- Extended-Matching items in sets of 2-20 items

One-best-answer with single items questions make explicit the number of options to be selected. These items are the most widely used multiple-choice-item format. They consist of a stem (e.g., a clinical case presentation) and a lead-in question, followed by a series of choices, typically one correct answer and four distractors.

Basic Rules for One-Best-Answer Items

Subject each question to the five “tests” implied by the following rules. If a question passes all five, it is probably well-phrased and focused on an appropriate topic.

1. Each item should focus on an important concept, typically a common or potentially catastrophic clinical problem.

Don't waste testing time with questions assessing knowledge of trivial facts. Focus on problems that would be encountered in real life. Avoid trivial, “tricky,” or overly complex questions.

2. Each item should assess application of knowledge, not recall of an isolated fact.

The item stems may be relatively long; the options should be short. Clinical vignettes provide a good basis for a question. For the basic sciences, patient vignettes may be very brief; “laboratory vignettes” are also appropriate. For the clinical sciences, each should begin with the presenting problem of a patient, followed by the history (including duration of signs and symptoms), physical findings, results of diagnostic studies, initial treatment, subsequent findings, etc. Vignettes may include only a subset of this information, but the information should be provided in this specified order.

3. The stem of the item must pose a clear question, and it should be possible to arrive at an answer with the options covered.

To determine if the question is focused, cover up the options and see if the question is clear and if the examinees can pose an answer based only on the stem. Rewrite the stem and/or options if they could not.

4. All distractors (i.e., incorrect options) should be homogeneous.

They should fall into the same category as the correct answer (e.g., all diagnoses, tests, treatments, prognoses, disposition alternatives). Rewrite any dissimilar distractors. Avoid using “double options” (e.g., do W and X; do Y because of Z) unless the correct answer and all distractors are double options. Rewrite double options to focus on a single point. All distractors should be plausible, grammatically consistent, logically compatible, and of the same (relative) length as the correct answer. Order the options in logical order (eg, numeric), or in alphabetical order.

5. Avoid technical item flaws that provide special benefit to testwise examinees or that pose irrelevant difficulty.

Do **NOT** write any questions of the form “Which of the following statements is correct?” or “Each of the following statements is correct EXCEPT.” These questions are unfocused and have heterogeneous options. Avoid using absolutes such as *always*, *never*, and *all* in the options; also avoid using vague terms such as *usually* and *frequently*.

Summary of Technical Item Flaws

Issues Related to Testwiseness

- **Grammatical cues** - one or more distractors don’t follow grammatically from the stem.

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 B1 (thiamine)
- C. CT scan of the head
- D. phenytoin
- E. diazepam

- **Logical cues** - a subset of the options is collectively exhaustive.

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

- **Absolute terms** - terms such as “always” or “never” are in some options.

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

- **Long correct answer** - correct answer is longer, more specific, or more complete than other options.

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

- **Word repeats** - a word or phrase is included in the stem and in the correct answer. Sometimes, a word is repeated only in a metaphorical sense, e.g., a stem mentioning bone pain, with the correct answer beginning with the prefix “osteo-”.

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

- **Convergence strategy** - the correct answer includes the most elements in common with the other options.

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

Issues Related to Irrelevant Difficulty

- **Options are long, complicated, or double**

Peer review committees in HMOs may move to take action against a physician’s credentials to care for participants of the HMO. There is an associated requirement to assure that the physician receives due process in the course of these activities. Due process must include which of the following?

- A. Notice, an impartial forum, council, a chance to hear and confront evidence against him/her.
- B. Proper notice, a tribunal empowered to make the decision, a chance to confront witnesses against him/her, and a chance to present evidence in defense.

- C. Reasonable and timely notice, impartial panel empowered to make a decision, a chance to hear evidence against himself/herself and to confront witnesses, and the ability to present evidence in defense.

- **Numeric data are not stated consistently.**

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%

- **Terms in the options are vague (e.g., “rarely,” “usually”).**

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

- Language in the options is not parallel.

- Options are in a nonlogical order.

In a vaccine trial, 200 2-year-old boys were given a vaccine against a certain disease and then monitored for five years for occurrence of the disease. Of this group, 85% never contracted the disease. Which of the following statements concerning these results is correct?

- A. No conclusion can be drawn, since no follow-up was made of nonvaccinated children.
- B. The number of cases (ie, 30 cases over five years) is too small for statistically meaningful conclusions.
- C. No conclusions can be drawn because the trial involved only boys.
- D. Vaccine efficacy (%) is calculated as $85 - 15/100$.

- “None of the above” is used as an option. It will often be possible to fix such items by replacing “none of the above” by an option that means roughly the same thing but is more specific. For example, in an item asking an examinee to specify the most appropriate pharmacotherapy, replacing “none of the above” by “no drug should be given at this time” will eliminate the ambiguity of “none of the above.”

Which city is closest to New York City?

- A. Boston
- B. Chicago
- C. Dallas
- D. Los Angeles
- E. None of the above

- Stems are tricky or unnecessarily complicated
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
- The answer to an item is "hinged" to the answer of a related item