Writing Items for the ABR

Heather Hopkins, Senior Content Editor

hhopkins@theabr.org (520) 790-2900, ext 156

Item and Test Validity

- Effective item writing is important to ensure that the materials used to evaluate test-takers are valid assessments of the candidates' knowledge.
- Downing et al. (1995) evaluated the validity of a classroom achievement test in medical education that contained flawed test items (e.g., unfocused item stems, use of *none of the above* and *all of the above*, and negatively worded stems).
- **He found that flawed items failed nearly one-fourth more students than nonflawed items.**

How much time does it take?

- Writing multiple choice questions is difficult and time-consuming, even for those who have been formally trained in their construction.

- Professional item writers plan on at least one hour or more to write one good item.

Adapted from Jannette Collins, MD, MEd, FCCP, American Roentgen Ray Society Education/CME committee, "Writing Multiple Choice Questions for Continuing Medical Education Activities and Self-Assessment Modules"

What is an item?

- **STEM =** Background and situational information, followed by request for answer.
- **OPTIONS =** The **KEY** (correct answer) and 3 or 4 **DISTRACTORS.**

It is fine to have 4 options instead of 5.

It is also acceptable to have 3 options, especially for test items for which there are only 3 plausible choices.

What makes a good stem?

1. Linear delivery of information
2. Complete, clear question
3. Focus on a single concept
4. Positively worded format
5. Clinically relevant
6. Noncontroversial

STEM

1. Linear Delivery of Information

Background + situational info + request for answer

A 32-year-old woman is treated for . . .

+ One year later, she presents with. . . .

+ To diagnose her condition, what examination is most appropriate?

NOTE: Cases should be written in the present tense, when possible. Avoid use of second person (“you”).

STEM

1. Linear Delivery of Information

Background + situational info + request for answer

Nonlinear example:

Which of the following is the most likely diagnosis in a patient with rapid weight gain who presents with fatigue, right upper quadrant discomfort, and abdominal distension six weeks after receiving abdominal irradiation for cancer of the hepatobiliary system?

STEM

1. Linear Delivery of Information

Background + situational info + request for answer

Better example:

A patient with cancer of the hepatobiliary system is treated with abdominal irradiation.

+ Six weeks later, he presents with fatigue, right upper quadrant discomfort, abdominal distension, and rapid weight gain.

+ What is the most likely diagnosis?

STEM

1. Linear Delivery of Information

Better example (without extensive background):

Situational info (includes image, if any) **+ request for answer**

For a patient with an inoperable nonmetastatic Ewing sarcoma of the proximal femur,

+ what is the most appropriate treatment?

STEM

2. Complete, Clear Question

Incomplete examples:

1. Carcinoma:

2. In women, carcinoma:

Do you know what these items are requesting of you?
Could you answer this question without a list of options?
These don't pass the "cover test."

STEM

2. Complete, Clear Question

Better example:

In women, what is the most common site of carcinoma?

This question passes the “cover test.”

Also, remember that stems need to be written as complete sentences, which ask a direct question.

STEM

3. Focus on a Single Concept

Unfocused example:

Which of the following statements about surgical resection of metastatic disease to the lung is true?

A. Lymphangitic spread of breast cancer in a lung can be arrested if the involved lobe is resected. **(resection)**

B. Patients with metastatic lung disease from soft tissue sarcomas and bone sarcomas have been shown to benefit from surgical removal of the metastatic disease.* **(resection)**

C. Nuclear medicine lung scan is the most effective radiographic study to detect, plan, and follow patients who are candidates for resection of pulmonary metastases. **(treatment planning)**

D. For a survival advantage to be seen, the pulmonary metastatic disease must be limited to only one lobe of one lung. **(condition)**

STEM

3. Focus on a Single Concept

Better example:

What **type** of metastatic lung disease is most effectively treated by resection?

A. Invasive carcinoma of the breast **(type)**

B. Soft tissue sarcomas* **(type)**

C. Ovarian cancer **(type)**

D. Brain cancer **(type)**

This is effectively testing knowledge of a single concept.

STEM

4. Positively Worded Format

- Items should be structured to ask for the ***correct*** answer and not a wrong answer.
- Negatively worded items tend to test recall-level information.
- Most negatively worded items fail the cover test.
- The ABR concurs with the *National Board of Medical Examiners* to avoid these types of items.
- Negatively worded items, if used, should be limited to situations that require a negative emphasis, e.g., practices to be avoided.

STEM

4. Positively Worded Format

Negatively worded example:

A 47-year-old woman is diagnosed with stage IB ovarian epithelial carcinoma. Which of the following treatment approaches would be the **LEAST** efficacious?

Remember linear delivery:

Background + situational info + request for answer

NOT

Background + situational info + request for answer, but oh, by the way, give us the reverse of what we've just led you to think we were asking about.

STEM

4. Positively Worded Format

Negatively worded AND unfocused example:

Which of the following statements about desmoid tumors is **FALSE**?

A. Desmoid tumors may arise at any **site** but are most common in the torso and extremities.

B. All patients with positive or close margins after initial surgery should receive **postoperative radiation therapy**.

C. For patients who refuse surgery or have unresectable tumors, radiation therapy is an excellent alternative treatment with **permanent local control achievable in 80%** or more of patients.

D. For optimal results, patients with desmoid tumors **should be treated with** very generous margins to doses of 50 to 60 Gy.

Options are unfocused (and wordy), and some have qualifying clauses.

STEM

4. Positively Worded Format

Better example:

Make 4 positive items out of the one negative item:

- Desmoid tumors most often occur in **what region of the body**?
- A patient with a desmoid tumor has positive surgical margins. What is the **most appropriate treatment**?
- A 30-year-old man has a desmoid tumor that is unresectable. What is the **likelihood** that radiation therapy will offer permanent local control?
- For optimal results, a patient with a desmoid tumor **should be treated with** very generous margins to doses in what range?

STEM

5. Clinically Relevant

- Even a well-constructed item is of little value if it does not relate to use in practice.
- Items should be clinically, medically, and scientifically relevant.
- A good clinical item asks the candidate to choose the most likely explanation, diagnosis, treatment, etc., for the condition(s) specified in the stem.

STEM

6. Noncontroversial

Avoid topics that are:

- Trendy, but not yet proven
- Not yet accepted in the mainstream
- Dependent on or specific to the work or studies of a particular person/group/ institution

What makes a good key?

1. Clearly the right answer
2. Properly completes the stem
3. Similar to distractors in length and structure
4. Not controversial or dependent on regional or institutional practice patterns
5. Supported by medical research

For each item, please provide at least one credible reference that supports the key.

What makes a good distractor?

1. Focused on same concept as the key
2. Properly completes stem
3. Similar to the key in length and structure
4. Plausible to some degree
5. No tricky language or clues

DISTRACTORS

1. Focused on Same Concept as the Key

- All options—including distractors—should be conceptually related to each other.

- Stem, key, and distractors need to test knowledge of a specific area.

DISTRACTORS

1. Focused on Same Concept as the Key

Example: Mixed options

Which of the following statements about prostate cancer is true?

A. Ninety-five percent of carcinomas occur in the central zone.

B. The probability of seminal vesicle involvement is 3/2 PSA + 9 GS-3 x 10.

C. Bone scans are positive in approximately 10% of patients with PSA < 10 ng/mL.

D. Despite screening PSA, there has been an increased incidence of lymph node metastases at diagnosis.

E. In men with PSA levels between 4.1 and 10.0 ng/mL, lower-percent free PSA levels increase the likelihood of a cancer diagnosis.*

DISTRACTORS

1. Focused on Same Concept as the Key

Better example:

In men with **prostate-specific antigen (PSA) levels** between 4.1 and 10.0 ng/mL, what effect do lower-percent free PSA levels have?

A. They increase the likelihood of cancer diagnosis.

B. They decrease the likelihood of cancer diagnosis.*

C. They increase the likelihood of cancer survival.

D. They decrease the likelihood of cancer survival.

- What percentage of prostate cancers occur in the central zone?
- In prostate cancer, how is the probability of seminal vesicle involvement calculated?
- A patient with prostate cancer has a PSA of 9 ng/mL. What is the likelihood that a bone scan will be positive?

DISTRACTORS

1. Focused on Same Concept as the Key

Example: Mixed options

Patients with superior vena cava obstruction syndrome are commonly noted to have which of the following?

A. Response to radiation correlated with daily fraction size

B. **Syndrome progression** leading to death

C. Nonsmall cell lung cancer **primary tumors**

D. Response to therapy independent of tumor histology

E. **Syndrome improvement** independent of collateral venous circulation

Choose **characteristic** or treatment response, but stick to one.

DISTRACTORS

1. Focused on Same Concept as the Key

Better example:

Create two items that are more focused.

1. Which of the following is a **characteristic** of superior vena cava obstruction syndrome?

2. Patients with superior vena cava obstruction syndrome are commonly noted to have what response to treatment?

DISTRACTORS

2. Properly Complete the Stem

Example:

A 74-year-old man is admitted for evaluation of a T4N2b squamous cancer of the hypopharynx. He has a sudden cardiac arrest. The most appropriate first step in management is to:

A. cardiopulmonary resuscitation (CPR).*

B. administer IV liquids.

C. initiating defibrillation.

D. perform a CT scan.

DISTRACTORS

2. Properly Complete the Stem

Better example:

A 74-year-old man is admitted for evaluation of a T4N2b squamous cancer of the hypopharynx. He has a sudden cardiac arrest. What is the most appropriate first step in management?

A. Cardiopulmonary resuscitation (CPR)*

B. IV intubation

C. Defibrillation

D. CT scan

Note: If you make the stem a question (not a sentence to be completed), proper completion is rarely a problem.

DISTRACTORS

2. Properly Complete the Stem

Example:

What is the most common symptom associated with the lesion shown in the image?

A. Pain
B. Fatigue
C. Nausea
D. Asymptomatic*

Note: Each option must be the correct part of speech to make sense as an answer to the question asked.

DISTRACTORS

2. Properly Complete the Stem

Better example:

Patients with the lesion shown in the image most commonly present with which of the following?

A. Pain
B. Fatigue
C. Nausea
D. No symptoms*

DISTRACTORS

3. Similar in Length to the Key

Example:

A conventional supraglottic laryngectomy is contraindicated when a patient has which of the following findings?

A. Extensions to one arytenoid

B. Pre-epiglottic space invasion

C. False vocal cord involvement

D. If there is a partial fixation of one vocal cord on direct laryngoscopy*

E. Pyriform sinus involvement

The longer and more specific option draws attention, and it's usually to the right answer.

In this case, the key does not follow the format of the other options (or complete the stem very effectively), which also draws attention to it.

DISTRACTORS

4. Plausible to Some Degree

- To the unprepared or underprepared examinee, there should be enough feasibility in the option for it to be considered.

- Ideally, distractors should be written to represent the compelling, rational, logical *wrong* answers that examinees would come up with if no choices were provided.

DISTRACTORS

5. No Tricky Language or Clues

Tricky language may cause candidates to select incorrect answers.

- Negative words (such as *not*, *never)*
- Multiple parts to one option (e.g., elements 1 and 2, or 2 and 3, or 3 and 4)
- Jargon, slang, older terminology
- Abbreviations, acronyms
- Ambiguity

DISTRACTORS

5. No Tricky Language or Clues

Example: Multiple and overlapping parts

This type of carcinoma is most likely to metastasize to which of the following organs?

A. Lungs
B. Lungs and heart*
C. Lungs, heart, and brain
D. Kidneys

The candidate who selects option A or C is partially correct but gets no credit, so the question does not accurately discriminate. Also, the repetition of lungs and heart allows a knowledgeable test-taker to count the number of times each “part” appears and guess that option B is the correct answer.

DISTRACTORS

5. No Tricky Language or Clues

Better example:

This type of carcinoma is most likely to metastasize to the lungs and what other organ?

A. Heart*
B. Brain
C. Pancreas
D. Kidneys

Now there is clearly one correct answer and no “tricks” or “clues.”

DISTRACTORS

5. No Tricky Language or Clues

Example: Ambiguous options

Non-Hodgkin lymphoma is diagnosed most often in what age group?

A. Young adults
B. Elderly
C. Infants
D. Mature adults
E. Generation Z

These descriptors may not be understood by every candidate and most will not be interpreted in exactly the same way.

DISTRACTORS

5. No Tricky Language or Clues

Better example:

Non-Hodgkin lymphoma is diagnosed most often in what age group?

A. 0 to 5 years

B. 10 to 15 years

C. 20 to 25 years

D. 40 to 50 years

E. > 60 years*

Describing the age groups this way guarantees that every candidate understands exactly what group of people each option represents. Note also that the options are arranged in numerical order, which further increases rapid comprehension.

DISTRACTORS

5. No Tricky Language or Clues

Clues may help unknowledgeable but test-wise candidates select the correct response.

- Vague terms (*might*, *may*, *could*, *can*, *should*)
- Absolute terms (*always*, *never*)
- *All of the above*/*None of the above* options (not used by the ABR)
- Correct response that is grammatically different, longer, or more specific than the other options
- Mutually exclusive pairs unless the pattern spans all options

DISTRACTORS

5. No Tricky Language or Clues

Example: Pairs

A 72-year-old man with lung cancer presents with fatigue, confusion, dyspnea, and imbalance. A whole-body PET scan is performed. Based on the images, what is the most likely diagnosis?

A. Metastasis to the bone

B. Metastasis to the brain*

C. Left-sided congestive heart failure

D. Pleural effusion

Pairs can lead the examinee to think that one of them must be the answer, or that neither is the answer. Either way, this format narrows the chances of guessing right to 50/50. We want to test knowledge, not test-taking savvy. On the other hand. . .

DISTRACTORS

5. No Tricky Language or Clues

Better example: Pairs

A 72-year-old man with lung cancer presents with fatigue, confusion, dyspnea, and imbalance. A whole-body PET scan is performed. Based on the images, what is the most likely diagnosis?

A. Metastasis to the bone
B. Metastasis to the brain*
C. Left-sided congestive heart failure
D. Right-sided congestive heart failure

Two sets of pairs negate the clue. This format will still discriminate effectively.

Summary

- Keep it focused.
- Make it linear.
- Use complete sentences in the stem.
- Ask for the *correct* answer.
- Use clear and accurate phrasing.
- Spell out all abbreviations.

And don’t forget . . .

Other musts:

- Indicate the key (correct answer).
- Classify (code) the item.

For more information . . .

Writer's Checklist

Writer's Checklist

Before you submit your items, check each item for the following:

I did:

- ☐ Spell out ALL acronyms/abbreviations, even those commonly used in the field.
- ☐ Use accurate phrasing and avoid jargon or "shorthand" so that colleagues from different institutions and those for whom English is a second language completely and accurately understand the question.
- ☐ Ask for the correct, not the "wrong," answer (positively-worded stems).
- ☐ Use complete sentence(s) in the stem.
- ☐ Include complete information and only relevant details.
- ☐ Cover the options to see if I can answer the question without looking at the option list for cues (the cover test).
- ☐ Include 4 or 5 total options based on the instructions given.
- ☐ Check that options make sense, are presented in similar format (same tense, verb usage, etc.), are independent not overlapping, and are given in a meaningful order.
- ☐ Keep options grammatically consistent with the stem.
- ☐ Indicate the correct answer (key).
- ☐ Provide graphics (if any) in the required format.

I did **NOT**:

- ☐ Use "negatively-worded" stems (i.e., *Which of the following is NOT*, *All of the following are true EXCEPT*, *Which of the following is FALSE*).
- ☐ Use options that are *All of the above*, *None of the above*, and/or multiple answers (i.e., *Both A and C* or *A, D, and E*).
- ☐ Use absolute terms such as *always*, *never*, *all*, or *none*.
- ☐ Use imprecise terms such as *seldom*, *rarely*, *occasionally*, *sometimes*, *few*, or *many*.
- ☐ Use clues such as *may*, *could*, or *can*.

Quick References

Stem and Options Factsheets

Editorial Factsheet: Cover Test and Flawed Stems

This factsheet is a quick reference that defines the cover test, also known as the *cover-the-options* test, and shows examples of the following types of flawed stems: unfocused and negatively-worded. Often flaws are related, overlap, and appear in combinations. Note that "nonflawed" items may need further revision.

Quick links (Press Control and click on the words to be taken to that section)

The Cover Test	Unfocused, Negatively-worded, Fails the Cover Test	Notes
Unfocused Stem		Sources
Negatively-worded Stem	Linearity	For more information

The Cover Test

Can the question be answered with the responses covered? Can a knowledgeable candidate answer the question before reading the responses? Well-constructed items could be administered as write-in questions.

To check the effectiveness of a stem, cover the option list and see if there is enough information in the stem to understand what is being asked and to formulate the correct answer. The stem should present all the information necessary for the candidate to determine the answer without having to look for clues in the option list.

Anaplastic oligodendroglioma is generally:

Isochromosome 1p is present in a location where anaplastic oligodendroglioma exists. How will the neoplasm react to chemotherapy?

Unfocused Stem

In women, carcinoma:

In women, what is the most common site of carcinoma?

Editorial Factsheet: Flawed Options

This factsheet is a quick reference that shows examples of the following types of flawed options: mixed options, double and multiple options, dissimilar lengths, mutually exclusive options, and clues that include absolute and vague terms. Flaws are related, often overlap, and often appear in combinations. Note that "nonflawed" items may need further revision.

Quick links (Press Control and click on the words to be taken to that section)

Mixed options	Dissimilar length	Sources
Double options	Mutually exclusive pairs	For more information
Multiple-part options	Clues	

Mixed options

Which statement about surgical resection of metastatic disease to the lung is true?

A. Lymphangitic spread of breast cancer in a lung can be arrested if the involved lobe is resected. (resection)
B. Patients with metastatic lung disease from soft tissue sarcomas and bone sarcomas have been shown to benefit from surgical removal of the metastatic disease.* (resection)
C. Nuclear medicine lung scan is the most effective radiographic study to detect, plan, and follow patients who are candidates for resection of pulmonary metastases. (treatment planning)
D. For a survival advantage to be seen, the pulmonary metastatic disease must be limited to only one lobe of one lung. (condition)

Resection is an effective treatment of lung metastasis caused by what type of disease?

A. Invasive carcinoma of the breast (type)
B. Soft tissue and bone sarcomas* (type)
C. Ovarian cancer (type)
D. Brain cancer (type)

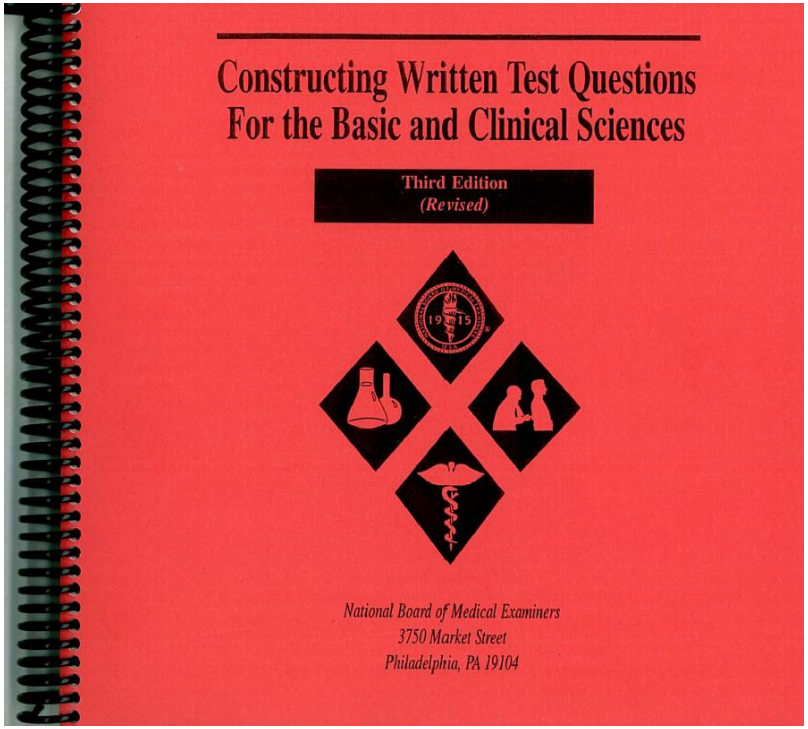
Double options

A lesion at 1-cm depth in tissue is to be treated with 6-MeV electron beam with bolus. A dose of 1.5 Gy to 80% is prescribed. If output is 1cGy/MU, SSD is 104 cm and cone factor is 0.8, what should the thickness of the bolus be, and how many MU should be delivered?

A. 0 cm, 153 MU
B. 0 cm, 170 MU
C. 1 cm, 209 MU
D. 1 cm, 270 MU
E. 2 cm, 302 MU

Redesign the item so there is only one question to be answered. This often means two or more items could be created. Spell out abbreviations.

More Information



www.nbme.org/publications

Thank you for your time and effort on behalf of the American Board of Radiology!