Item Writers' Guide

Last updated August 6, 2010
Questions or comments? E-mail editors@theabr.org

HIGHLIGHTS IN THIS EDITION

1. **TWO TO FOUR DISTRACTORS**
   Items can have 2 to 4 distractors (3 to 5 total options).

2. **AVOIDING THE USE OF THE "ALL ARE TRUE EXCEPT" CONSTRUCTION**
   Why it should be avoided.

3. **REWORKING NEGATIVELY-WORDED ITEMS**
   Items conceived in a negative format can be changed to positive ones.

4. **SAMPLES OF PROBLEMATIC ITEMS AND HOW TO FIX THEM**

Note
Most of the items in this guide were deleted from the item pool and have been reworked for instructional use. Therefore, they may contain outdated material. Also, items are included for the discussion of structure, and they may not be relevant to your subspecialty.

Foreword

Thank you for agreeing to write items for the American Board of Radiology (ABR) examinations. We appreciate the gift of your time.

This guide focuses only on multiple-choice (single best answer) items. This booklet is designed for easy reference as you write. Points of item writing are explained in detail in the body of the manual. As you get comfortable with the process, you may find that all you need is a checklist to guide you from start to finish. A list of "Item-writing Basics" is below, or you may use the table of contents, which is designed to be copied for that purpose.

Please feel free to offer feedback on this manual to information@theabr.org.

Item-writing Basics

General

- It is permissible to include tables, graphs, or diagrams. As appropriate, items for which the answers depend on interpretation of diagnostic images can be used.
- State the needed information as concisely as possible and at a comfortable language level.
- In addition to being targeted to the understanding of a resident, all items should be clinically relevant, noncontroversial, and up-to-date.
- Focus the stem and options; avoid clues and tricks.
- Make sure the items pass the cover test (see p. 5).
- Cite your references and code (classify) your items.

Stems

- Focus on one concept. To focus an item, think of what the question is really asking the candidate to do (e.g., recall information, apply knowledge, solve a problem).
- The stem should present all the information necessary for the candidate to formulate an answer without having to look for clues in the option list (the cover test).
- The stem should be presented in a linear fashion. The information should be chronological, if relevant, and the question to be answered should appear immediately before the list of options.

Options

- Use plausible options to avoid clues. Also, avoid gradations or degrees of correctness; there is one best answer among the choices, and to the prepared candidate, it is clearly the right answer.
- Keep options similar in structure and concept. Test one bit of specific information with each item.
- Options should properly complete the stem.
- An item may have four (4) OR five (5) answer options. Current pyschometric standards also allow as few as three (3) answer options.

Table of Contents/Checklist

Conflict of Interest
Specific questions and answer options written for the ABR cannot be used for any other activity. This includes classroom discussion, mock exams, lectures, seminars, or any other outlet that might compromise the security of our exams.

Copyrighted Images
Any image that is copyrighted—that came from a book, article, or other copyrighted source—cannot be used by the ABR unless we have written permission from the publisher/copyright holder (not the author) to use it. If you want to include such an image, it is your responsibility to get clearance in writing.

Item Format.

This guide focuses exclusively on multiple-choice items. Each item focuses on one concept to be tested. There is ***one best answer*** among the choices, and to the prepared candidate it is clearly the right answer. Our objectives are to:

- test knowledge of the subject matter presented;
- test one piece of information per item;
- present items clearly and concisely so the candidate can answer as many as possible during the time allotted.

Graphics format.

We are moving toward using more image-based items on our exams. You may include tables, diagrams, scan images, and screen captures as part of your items. Please follow the instructions you received or e-mail any graphics relevant to your items to items@theabr.org. We can convert other file types, but the .png file format is preferred. **DO NOT** send graphics as imbedded objects in text files or templates. This greatly degrades our ability to manipulate the image. Send graphics as separate files, clearly labeled with your name and the appropriate item's keyword so that we can place them with the correct items. Additional instructions may be requested from the Information Technology department via e-mail: agudenkauf@theabr.org.

Start early.

One part of successful writing is the review. To give yourself as much objectivity as possible, you should allow at least several days between writing and assessing the items before submission for consideration and review. This allows you to clear your mind of the process, "forget" what you wrote, and look at the items with a fresh perspective. Waiting until the end of the cycle to do your writing robs you of this important perspective.

Consider your audience.

It is important to consider who will be the recipients of your work. If what we present is not relevant to their experience or to their expected field of work, the item does not serve its purpose. The candidates taking these exams are being evaluated for their ability to practice safely and effectively. Items should not only be relevant, but ***important*** to practice. For Initial Certification examinations, items should deal with subjects that are taught in most residency programs. For Maintenance of Certification exams, items should be practice oriented.

Decide on your topic.

Is it relevant?
One thing that makes even a well-written item unusable for testing is if it asks for information that has little to do with the practice of radiology or to case management in general. Remember that we are testing the candidate's ability to practice radiology safely and effectively. All items should be pertinent.

Is it something the candidate must know to be successful in practice?
There are basic concepts that candidates must clearly understand. There are also perception, interpretation, and deduction skills that should be developed in training. We test all of these things.

Is it based on controversial opinions, studies, or material?
Items that relate to specific studies, regional practices, or the methods of particular doctors should not be used. Likewise, items should not focus on trendy subjects (those that may change in a short period of time) or on practices and theories that are open to debate.

Simple or complex?
There are two basic types of questions used on radiologic physics exams, each valuable for testing different skills and levels of understanding. A "simple" question would have either zero calculational steps, or at most, a relatively minor arithmetic component (the focus being more on the knowledge of relevant facts). A "complex" question requires some advanced reasoning and two or more calculational steps.

Cite your references now.
We must document this information for a variety of reasons. Cite the references while they are in front of you so you won't have to try to find them again later.

Choose and cite your reference(s).

Decide what sources you will be using. We need at least one complete reference for every item. References need to be credible sources that are nonspecialized, not controversial, and widely studied, replicated, and verified. Among other things, these are our defense in case an item is challenged. Therefore, provide ***at least one credible reference*** for every item. Be sure the reference supports the key.

Decide what *single* task you want the candidate to perform.

Focus on one piece of information.

An effective item points toward a single option; therefore, the stem should be focused in one direction, making the path to the answer clear. The item should ask the candidate to complete *one* task, answer *one* question.

(This should NOT be confused with the difference between writing simple and complex questions. You should be pointing the candidate toward a single answer. Whether that takes multiple steps is another matter. See "*Simple or complex?"* above.)

Recall information.
Some information should be memorized and easy for a radiologist to bring to mind. The following is an example of a recall item:

> What is the most common cause of superior vena cava syndrome?
>
> A. Tuberculosis
> B. Malignancy
> C. Central venous catheters
> D. Fibrosing mediastinitis
>
> (Key is B)

Apply knowledge.
The candidate must be able to employ knowledge in specific situations. This type of item builds on recall skills but requires more of the candidate, such as abstracting meaning from data, recognizing implications of clinical findings, identifying abnormalities on radiographs, or comparing possible treatment approaches. The following is an example of an interpretive item:

Note
For new items, the current preferred format is to use complete sentences in the stem or to use colons rather than blanks. The ABR is moving to a system in which all test questions can be used on any exam, therefore, all items should follow the current preferred format.

If needed, the calculator available during exams will be the standard Microsoft calculator.

A guidewire-induced spasm of the popliteal artery occurs during angioplasty of the superficial femoral artery. The treatment of choice is intra-arterial administration of which of the following?

A. Alcohol
B. Lidocaine
C. Papaverine
D. Tolazoline
E. Nitroglycerin (Key is E)

Solve a problem (not necessarily through mathematical calculation).
Here we are testing whether the candidate can make judgments and decisions to deduce answers to complex clinical questions or to select a strategy. The following are examples of problem-solving items:

A brain image is obtained using a spin-echo MRI pulse sequence with TR = 000 ms and TE = 100 ms. What is the anatomy with the highest signal intensity?

A. Gray matter
B. White matter
C. Lens of the eye
D. Cerebrospinal fluid
E. Intracranial fat (Key is D)

Immediatetly after balloon angioplasty of an atherosclerotic stenosis of the superficial femoral arter, angiography shows acute closure of the vessel. What is the most likely cause?
A. Spasm
B. Thrombosis
C. Dissection
D. Rupture
E. Recoil (Key is C)

Other tasks the candidate may perform.
Some of the other ways an item may be used to evaluate a candidate's interpretive and problem-solving skills are:

- Interpreting diagnostic images, data, graphs, and tables
- Using pathophysiology to predict findings from certain conditions
- Explaining why something has occurred
- Ordering and interpreting diagnostic studies
- Designing an overall treatment program (including surgery, radiation therapy, and chemotherapy) or quality management program
- Selecting or optimizing a radiation therapy treatment plan
- Calculating a tumor dose
- Determining a prognosis

Some items that employ a multiple-answer format do exist in the database, so you may recall seeing them. However, we are moving away from that format. Most of these items have been revised or eliminated.

Design the item so there is only one question to be answered.

For scoring reasons, it is recommended that items ask for one answer only. The following would present a problem:

> 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, source-to-surface distance (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
>
> (Key is C)

However, the material would be usable if the item were split into two items (they would most likely not appear in the same exam):

> 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, source-to-surface distance (SSD) is 104 cm, cone factor is 0.8, and bolus thickness is 1 cm, how many MU should be delivered?
>
> A. 153 MU
> B. 170 MU
> C. 209 MU
> D. 270 MU
> E. 302 MU (Key is C)

and

> 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, source-to-surface distance (SSD) is 104 cm, cone factor is 0.8, and the MU to be delivered is 209, what should the thickness of the bolus be?
>
> A. 0.5 cm
> B. 0 cm
> C. 1 cm
> D. 1.5 cm
> E. 2 cm (Key is C)

Decide what depth of knowledge you want to test.

Consider the following illustration:

> What city is the capital of Arizona?
>
> A. Albuquerque
> B. Dallas
> C. Phoenix
> D. Sacramento

One does not have to have much knowledge about western cities to choose the right answer. However, if the item is written this way ...

What city is the capital of Arizona?

A. Tucson
B. Flagstaff
C. Phoenix
D. Yuma

... a person must know much more about Arizona cities to make the correct choice. Think about where the stem is going to lead the candidate in his or her "personal database." What level of knowledge do you want to test?

Here is an example of what to avoid. (*The ABR does not use "None of the above" constructions*, but to illustrate a point, we'll include it.)

For a patient suspected of having a hepatocellular carcinoma, which of the following tumor markers is useful?

A. Placental alkaline phosphatase
B. Beta human chorionic gonadotropic hormone (βhCG)
C. Tumor necrosis factor
D. None of the above (Key is D)

This item shows how we can actually miss testing the knowledge the writer intended. A candidate could technically get this item right, but not know that the useful tumor marker for hepatocellular carcinoma is α-fetoprotein.

Write the stem.

The #1 flaw we find in submitted items: Lack of focus.

The stem is the part of the item that asks for a response. The most effective ones are questions, though in some instances it may better serve your topic to present a statement to be completed (not to be confused with "fill in the blank," which is not recommended). ***The stem should present all the information necessary for the candidate to figure out the answer without having to look for clues in the option list.***

The Cover Test
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 an answer.

Suppose a stem is written this way:

In pair production:

or

Which of the following statements about pair production is true?

Avoid the "Which of the following is true" construction! Although used in the past, this type of stem is unfocused and usually leads to the inclusion of mixed options. Rewrite the item so that it asks a single, focused question that passes the cover test. Often more than one focused item can be generated from an item with this construction.

What is being asked about pair production? Any number of ideas could run through the candidate's mind. The intent is not clear. These stems fail the cover test. (This type of stem also leads to writing mixed options.)

This delivery is better:

In pair production, what is the role of positrons?

Now the candidate understands what we want to know. If he/she knows the answer, it's just a matter of finding it in the options.

Be aware of sentence structure and linearity.
The most effective way to deliver information is in a linear fashion—one that allows the mind to follow a path, arriving smoothly at the point at which an answer must be given. ***Focusing the stem in this way will also help you focus the options.***

The formula most suited to linear thought is:

Background info + Situational info (or equation) + Request for solution

An example of weak, nonlinear structure would be:

> What percentage of patients will survive at least 42 months if they are part of a group that exhibits lifetimes that are normally distributed with a mean of 36 months and a standard deviation of 3.0 months?

In this instance, the candidate would already be trying to formulate an answer (or wondering how it can be done) before crucial information is given.

A better and linear way in which to present the information would be:

> A group of patients exhibits lifetimes that are normally distributed with a mean of 36 months and a standard deviation of 3.0 months. What percentage of these patients will survive at least 42 months?

Use common and precise language.
We are not testing vocabulary. Sometimes when people write, they tend to use specialized words when common words would be just as effective. More common words are actually more effective because they ensure that candidates are not laboring to understand the item, and that we are in fact testing knowledge of the subject matter—not language skills.

Preferred structure is "a patient with _____" rather than "a _____ patient." For example, "a patient with diabetes" rather than "a diabetic patient."

Likewise, language needs to be medically, scientifically, and technically precise and accurate. This may mean rephrasing from "this is how we say it" to increase clarity, reduce ambiguity, and to use nomenclature consistently. This is equally important for the medical field in general and for the examinations leading to board certification.

Be clear and concise.
One objective of an exam is to present as many items as possible. The more answers a candidate gives, the more precisely proficiency can be assessed. Stems should be clear and concise.

Wordiness is not appropriate. Long, involved explanations or histories should be avoided. State the needed information as concisely as possible and at a comfortable language level. The following stem goes overboard:

> A 57-year-old man with typical anatomy has a squamous cell carcinoma involving the distal 9 cm of the esophagus that extends to the cardia with slight extension into the paraesophageal fat. There is no regional or distal metastasis. The treatment plan is for a 50-Gy tumor dose in 25 fractions using concurrent chemotherapy (5-FU and cisplatin). The first 36-Gy tumor dose was given with anterior-posterior portals. The spinal cord has received 110% of the prescribed dose. The patient now comes for a 14-Gy boost in 7 fractions through reduced fields. Which of the following field arrangements is most likely to give adequate coverage of the tumor volume with an acceptable risk of toxicity? (Assume a spinal cord tolerance of 45 Gy.)

Problems with "All of the following are true EXCEPT:" construction
In a research study, Downing evaluated the validity of a classroom achievement test in medical education that contained unfocused and/or negatively-worded item stems. He found that these items failed nearly one fourth more students than nonflawed items. It can also lead to awkward completion of the stem (e.g., "All of the following statements about… are true EXCEPT:" then the full statements have to be added as part of this sentence.) Please avoid this construction.
Some argue that this construction simulates the process of elimination inherent in radiology (e.g., differential diagnoses). However, this construction for an item stem is unfocused, fails the cover test, and asks for the *wrong* answer. A different item type may be better for these types of content; discuss options with your coordinator.

Further arguments against negative construction
Negative construction (e.g., "Which of the following is FALSE?") violates at least five principles of item writing:

1. Keep vocabulary simple and straightforward.
2. Minimize the amount of reading time (rereading is often necessary).
3. Make the stem's direction very clear.
4. Ask for the correct answer, not a "wrong" answer.
5. Avoid tricky language.

Negative stems are also typically unfocused, which leads to mixed options.

"[T]his guideline comes from a consensus of experts in the field of testing who feel that the use of negative words in the stem [has] ***negative effects on students and their responses*** to such items." Haladyna, *Developing and Validating Multiple-choice Test Items*, 2nd Edition, Erlbaum, 1999. (Bold and italics added.)

A better way to present this would be:

> A 57-year-old patient has squamous cell carcinoma of the distal esophagus. The spinal cord has already received 36 Gy through anterior-posterior portals. Planned treatment now is 14 Gy in 7 fractions. Which of the following field arrangements is most likely to give adequate coverage with minimal risk of spinal cord injury? (Assume a spinal cord tolerance of 45 Gy.)

Stay focused.
Although it might be appropriate to include some extraneous information in a stem to test whether the candidate can glean the pertinent points, it is important that the stem maintain a high degree of focus. A stem that rambles, delivers information disconnectedly, or includes too much information not directly related to the question being asked can be confusing and can draw the candidate away from the task at hand. Remember that the more items the candidate answers, the better we can truly assess knowledge. We don't want to waste the candidate's time with extraneous reading.

Example:

> A 60-year-old female patient presented with rectal bleeding and was found to have an adenocarcinoma of the midsigmoid colon. It was completely resected by low anterior resection, and five of five lymph nodes were negative. No adjuvant therapy was given. Two years later she presents with anorexia and a ten-pound weight loss and is found to have liver metastasis. What is the most appropriate treatment now?

This stem could be rewritten more simply as:

> Two years after resection of an adenocarcinoma of the midsigmoid colon, a patient develops liver metastases. Which of the following is the most appropriate treatment option?

Avoid negatives.
We are trying to ascertain what the candidate knows, not how easily a test-taker can be confused. It is best to avoid items that require reverse thinking. This complicates the presentation of information. What a negative stem sets up is:

Background info + Situational info (equation) + Request for solution, but—oh, by the way—give us the opposite of what we've led you to think about

Avoid writing items that include:

- The statement that does **NOT** reflect…:
- Which of the following is **FALSE**?
- Which of these is **NOT** an indicator…?
- All of the following are true **EXCEPT**: (see sidebar)

This kind of item is easy to write because you just have to take 3 or 4 facts from the references and make up one bogus one. But asking for a false "answer" is more a test of the candidate's ability to think conversely than knowledge of the subject. Often, even experts will read this type of stem and mentally proceed to the CORRECT answer. Test-takers tend to choose the first correct answer in the list and often don't remember that they were supposed to be seeking the INCORRECT option from the list.

Remember: Ask for the *correct* answer, not the *wrong* answer.

Reworking Negatively-Worded Items

What if I think of a good item, but it comes to me in negative format?
There is hope! We have been going through old items, trying to make them psychometrically sound. We have found a LOT of negative items. The good news is that in many cases each option in a negative item can serve as the focus of one or more new positive stem.

Start with the distractors, which in this case would be the wrong answers, even though they give correct information. (Do you see why negative wording causes problems?) Each distractor deals with a specific piece of information. Work with that information and make it a stem and key. Then fill in the distractors.

Also use the right answer (key)—the negative response, the correct option, but the wrong answer. (It is confusing, isn't it?) What is the information you are testing with that key? What is it that you wanted to make sure the candidate knew was wrong? Flip it around and ask the question directly.

If your options in the negative item were homologous (similar in concept and structure), you may want to use them as parts of one positive item; however, it is still OK to write multiple, similar items. The rapidly changing field of radiology means we constantly need to add new items to our exams. By converting a negatively-written item, we can create up to five items on the same topic, and we can use different items on multiple exams to test basically the same knowledge. This helps guard the security of the exams.

Example
Original negative item:

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. Patients who have 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.

(Key is B)

The stem is both negative and unfocused. It does not pass the cover test. This led the writer to include mixed options—all options should be the same category and similar in concept, structure, and length. The following are examples of items that could be constructed using the options of this item as bases for new stems.

Where do desmoid tumors often occur?

A. Extremities
B. Abdomen
C. Head and neck
D. Digestive system

(Key = A)

A 30-year-old man has a desmoid tumor that is unresectable. What is the likelihood that radiation therapy will offer permanent local control?

A. 50%
B. 60%
C. 70%
D. 80%

(Is D key? References didn't agree)

For optimal results, how should a patient with a desmoid tumor be treated?

A. Generous margins to 30 Gy
B. Generous margins to 60 Gy
C. Conservative margins to 30 Gy
D. Conservative margins to 60 Gy

(Key is B)

Item from the old key (negative) option:

A patient with a desmoid tumor has positive surgical margins. What is the recommended treatment?

A. Repeat surgery for wider excision
B. Chemotherapy
C. Postoperative irradiation

(Key = ?)

See more examples starting on page 16.

Write the correct answer.

It should offer the most plausible response.
To the prepared candidate, there should be no doubt that the answer is the only option that could possibly be selected from the list of choices. The prepared candidate will readily choose it.

It should grammatically and properly complete the stem.
Whether the answer completes a sentence, or answers a question with a phrase or a full sentence, it should be in proper syntax.

It should be similar in length to the distractors.
Nothing gives away the right answer faster than one choice that is well-thought-out and carefully worded, and the others are choppy, too short, too long, etc. Candidates can also be drawn to distractors that are different from the others. To ensure that options are chosen because they are the best answers, and not because of some clue, try to make them as similar as possible.

Vary placement of the key in the options list.
Sometimes we have to rearrange the options; this might be necessary if (for instance) we have too many items on an exam for which the answers are "A." For some items, changing the order of options isn't a problem. For others, such as numeric lists (percentages, dosages, etc.) it can be problematic. Those have to be listed in numeric order.

For these types of item, if you have 5 options, it helps if you avoid putting the key in the middle of the list. Statistically, the unsure candidate will pick the middle option. (Remember, we want to test what candidates really know, not their guessing skills.) If one of these items has to be "rearranged," it really means coming up with other options, and we don't know what's feasible (e.g., if you've given of list of percentages of 10, 20, 30, 40, and 50, is 60 still in the range of possibility, or should a number be added at the lower end of the scale?).

Write the distractors.

Distractors are the other possible answers in the option list. Distractors are perhaps the hardest part of the item to write because they must, on some level, seem like reasonable options.

Make them plausible.
To the unprepared or underprepared candidate, distractors have to seem like they could be the right answers.

Provide 2 to 4 distractors (3 to 5 total options).
It is better to have fewer plausible distractors than a list of options with more distractors (see sidebar). A distractor that adds nothing (or doesn't fit the list) can actually negatively skew the statistics for that item.

Research has shown that 2 distractors (3 total options) does adequately discriminate. Three total options can be used, especially when a fourth plausible option is difficult or impossible to provide (e.g., when the question asks the candidate to determine whether a dose or agent increases, decreases, or remains the same under specific circumstances).

Avoid mixed (nonhomologous) options.
Be sure distractors are rooted in the same material as the correct answer. The following item has several problems, one of which illustrates our point here:

Which of the following statements about esophageal cancer is **FALSE**?

A. The esophagus has a well-developed submucosal lymphatic plexus that predisposes to lymphatic metastasis. [anatomy and characteristic of that anatomy]
B. According to current AJCC staging, celiac lymph node involvement is not classified as a regional nodal disease. [characteristic of the disease]
C. Barrett esophagus is associated with an increased risk of adenocarcinoma of the esophagus. [risk factor]
D. Tylosis is not associated with an increased risk of squamous cell cancer of the esophagus. [negative—so now we have a double negative— + risk factor]

Can you figure out the key?

Note
There are some exceptions to the mixed options rule. In the real-life clinical situation, management sometimes requires making a choice between various interventions or further diagnostic tests. In such situations, the options should reflect the various possibilities that might realistically be considered by the candidate. The options could include further diagnostic tests or treatment intervention.

The stem is both negative and unfocused. Added to that is the problem of options that are not similar in content. These are called "mixed options." Mixed options are a common problem in items that have unfocused stems and begin with statements such as "Which of the following statements about a condition is true (or false)." If you find yourself writing mixed options, go back and look at the structure of your stem.

Supposedly, the "key" for this item is option D.

Make sure distractors are similar in length to the correct answer.
As was discussed for writing the key, you shouldn't have one choice that is well-thought-out and carefully worded, while the others are choppy, too short, too long, more specific, use qualifying phrases, etc. Try to make all options as similar as possible.

Why is Lipowitz metal (Cerrobend) is preferred over lead for custom blocking?

A. It is much easier to machine than lead.
B. It has a much lower melting point than lead.
C. It is composed primarily of cadmium, a less toxic substance than lead.
D. Custom blocks made of Lipowitz metal (Cerrobend) have a sharper penumbra than custom blocks made of lead.
E. Its linear attenuation coefficient is higher than that of lead.

(Key is B)

Option D is longer and worded differently than the others. In this case it is not the key, but its differences could draw candidates to think it is right. If it is changed to read...

D. Custom Cerrobend blocks have sharper penumbra than custom lead blocks.

... it will more closely match the other options. Therefore, if it is chosen, it will be because the candidate thinks it is the right answer, not because of a trick. (It could also be deleted from the list of options.)

Make sure the distractors complete the stem properly.
An option that does not fit the sentence structure of the stem sticks out. Make sure the correct answer and the distractors all "blend in"—that is, they complete the stem correctly and grammatically.

Tricks
Language or construction that may cause candidates to select an incorrect option

Don't use tricks.
Negatives
We are not trying to ascertain whether candidates can be tricked with language or vague cues. This is also a reason to avoid negative stems: choosing an option in the negative is more a test of reverse thinking than of knowledge. The following item, in addition to having an unfocused stem and mixed options, is further weakened by tricky language:

Which of the following statements most accurately describes Turner syndrome?

A. A webbed neck is not found frequently.
B. The ovaries usually have normal function.
C. The chromosomal pattern is typically 45,x.
D. The cause for short stature is clearly understood.
E. Patients rarely have shield-like chests. (Key is C)

Remember
We are not trying to trick the candidate. On the contrary, we are trying to make it possible for each person to answer as many items as possible in the allotted time, thereby affording the greatest chance of success.

Option A contains a negative term (*not*), which a candidate could easily miss. Options A, B, and E are also examples of reverse truths, two of which are pretty obvious. A reverse truth is a true statement that is made false simply by changing or inserting a word or two. Option D is improbable, using the word "clearly." For this item, it is not hard to choose C.

Jargon, slang, etc.
Other tricks that can be confusing to the candidate include use of jargon, slang, eponyms, and abbreviations.

> A 74-year-old man was admitted for evaluation of a T4N2b squamous cancer of the hypopharynx. He arrested and was coded. Which of the following would be the most appropriate next step?
>
> A. Bolusing with "roids"
> B. Plugging into an IV
> C. Tubing
> D. Bagging
> E. Scoping
>
> (No key is given)

Acronyms
With few exceptions, ABR policy is to spell out terms on first use in an item, with the acronym following in parentheses. The acronym can stand alone in the remainder of the item. This is to avoid confusion of terms and to make information absolutely clear.

The issue of fairness arises in this situation because not all candidates may be familiar with these particular terms. Standard medical terminology should be used in all examination items for the purpose of clarity and fairness. This also applies to acronyms (see sidebar).

Don't give clues.
Some candidates have mastered the art of test-taking, including how to figure out what choices are not correct by looking for clues. Some giveaways are:

- Vague terms*: might*, *may*, *can*—clues to the key because they indicate that almost anything is within the realm of possibility
- Absolute terms: *always*, *never*—clues to distractors because there are no exceptions

The following item illustrates both points:

> Which of the following statements about Ewing sarcoma is correct?
>
> A. It always involves the diaphysis.
> B. It involves the metaphysis more commonly than the diaphysis.
> C. It may involve the epiphysis.
> D. Extraosseous Ewing sarcoma has never been reported.
> E. It almost always involves long tubular bones.
>
> (Key is C)

Clues
Language or construction that may help unknowledgable but test-wise candidates to select the correct option

Notice that the stem is unfocused and that the options are mixed. In addition, for the observant candidate, Options A and D can be eliminated from consideration in this item because the terms "always," and "never" mean there are no exceptions. On the other hand, Option C is almost certain to be the correct response because the term "may" includes all possibilities.

Can this item be saved? Let's focus the stem, remove the clues, and make the options homologous:

Which of the following is usually involved in Ewing sarcoma?

A. Diaphysis
B. Metaphysis
C. Epiphysis

(Key is C)

Now the stem asks a clearer question. (Although "usually" is vague, moving it to the stem makes it equally applied to all options.) The answers are straightforward. There is no ambiguity and there are no clues.

The savvy test-taker understands the thought processes of the item writer and will pick up clues as to what answers are right and wrong. We try to make sure we don't leave a path to follow.

Double/multiple options

Double/multiple options are responses that contain two or more pieces of information. As explained under "Design the item so that there is only one question to be answered," these can present scoring issues. Then there is the added problem that many test-takers understand the thought process of the writer and can figure out the correct answer based on the frequency with which each part of the option occurs in the list. For example, if the elements of the options appear with this frequency....

A. 1 & 2
B. 1 & 3
C. 1 & 4
D. 2 & 5

... the saavy examinee will note that 1 appears 3 times, and 2 appears twice. The outliers (4 and 5) are least likely to be right. The most likely answer, then, is the one that contains both 1 and 2 (A).

For an item with three elements per option it gets more complicated, but this structure can still help the examinee narrow the field. ***It is better to stick with one-element answers whenever possible.***

All of the above

This type is typically not a good measurement tool. If candidates recognize two correct choices, they know the rest of the list probably follows suit and may not even read the other choices. These may seem easy to write—just pull a few facts from the references—unfortunately, candidates know the technique.
We do not use this construction.

None of the Above

There is an argument that "none of the above" responses can be effective. Sometimes there is no definitive answer to the question, and the item is trying to determine that the candidate realizes this. A "none of the above" item may actually entice the candidate to study all of the options more thoroughly. However, we feel that the possible negative effects are more relevant to our situation. Therefore, the ABR does not use any items in this format.

None of the above

Although this type of option may have other motives (see sidebar), some writers use it to escape having to supply a good answer. Many candidates know that statistically, when it is used, "none of the above" is the correct response more than 50% of the time. This may entice the underprepared candidate to choose incorrectly.
For this reason, the ABR does not use this item type.

Collectively-exhaustive/mutually-exclusive options
Paired options only work if ***all*** in the list effectively cancel each other—one cannot be correct if another is—therefore rendering the rest of the options invalid. It is only a valid format if ***all*** of the options are involved in the matrix.

If there is one pair of mutually-exclusive options...

> Lacunar skull is most frequently due to which of the following?
>
> A. Chiari I malformation.
> B. Chari II malformation.
> C. aqueductal stenosis.
> D. neurofibromatosis.
>
> (Key is B)

... the saavy test-taker will pick out the pair and figure that one of them is correct (and it usually is). Now the statistics on this item are skewed because we've really given the candidate a 50/50 chance of success. On the other hand, if there are two sets of mutually-exclusive options...

> Which of the following venous access devices would be most appropriate for a patient requiring a long-term, continuous infusion of medication for pulmonary hypertension?
>
> A. Implanted port in the chest wall with catheter inserted via the internal jugular vein
> B. Implanted port in the upper arm with catheter inserted via the basilic vein
> C. Hickman catheter (cuffed) tunneled from chest wall to insertion site in the internal jugular vein
> D. Quinton catheter (large bore, 2-lumen, cuffed) tunneled from chest wall to insertion site in the internal jugular vein
>
> (Key is C)

...we have negated the advantage and returned the odds to 25%.

Note: The second example on page 9 also shows good use of two mutually-exclusive pairs.

Key the answer.

Just a reminder that we need to know what option is correct. The correct answer is called the "key." Please make sure to indicate what option you intend to be the correct response. (And remember to vary the position of the key in the list of options.)

We need to have:
- **a key**
- **a classification**
- **two references**

for each item.

Classify the item.

Classification of all items is required. Each item should be classified based on the key (correct answer). *We cannot use an unclassified item.* On the template, please indicate the proper classification for each item.

Review the item.

As previously mentioned, it is best to write items well before the deadline, then lay them aside for a while. When writers try to revise their own work right away, they tend to miss areas that need correction or clarification—for example, the writer often "sees" the intended content, whether it is actually there or not. Returning to them later decreases this risk.

When you do review the items, consider the following points:

Cover the options and see if you can answer the question.
The stem should give sufficient information so that the candidate can formulate an answer without having to look at the option list for cues (the "cover test").

Look at stem structure.
Be sure you are delivering the information in the most direct way possible. Be sure the presentation is focused, concise, and linear.

Writing your items well before the deadline gives you the opportunity to look at them with "fresh eyes" before submission, thus catching mistakes you might otherwise miss.

Check the key.
Will the prepared candidate recognize that it is clearly the right answer?
It is the same category and similar in concept, structure, and length as the distractors?
Does it properly complete the stem?
Make sure it is clearly marked.

Consider the validity of the distractors.
Do any of them sound ridiculous?
Are they the same category and similar in concept, structure, and length to the right answer?
Do they complete the stem appropriately?
Do they avoid tricks, clues, and convoluted presentation?

Check your reference citation.
Each item must have at least *two* complete references.

Keep turning! Still ahead:
- Sample item rework
- Ideas to get you started

Problematic Items and How to Fix Them.

Negatively-worded, unfocused stem; mixed options. Example 1

Weak:

Which of the following statements regarding soft tissue sarcoma is **FALSE**?

A. With the use of aggressive chemotherapy, the lung is the most common location for distant metastases.
B. The overall incidence of lymph node metastases in soft tissue exceeds 20%.
C. Overall 5-year survival for soft tissue sarcomas is 40%.
D. Over 80% of local recurrences will occur in the first 2 years of followup.
E. Over 80% of distant metastases will occur within 5 years of followup.

(Key is B)

Why:

Not only is the stem negative, but there is very little focus. Try the cover test; cover the options with a piece of paper and try to formulate an answer. We know it is asking about soft tissue sarcoma, but what about it? Symptoms? Treatment? Prognosis? It could be anything. This stem structure also allows for the creation of options that are not the same category and not similar in concept, structure, and length. Option A is about a therapy-related complication. Option B reports the instance of one type of metastasis. Options C, D, and E report prognostic statistics.

Better:

Positive, focused items can be gleaned from the options (stems only here). Although these items are similar, we can use multiple items on the same subject for different exam administrations.

1. A patient with soft tissue sarcoma is treated with aggressive chemotherapy. Metastasis is most likely to develop in what organ?

2. What is the likelihood that a patient with soft tissue sarcoma will develop lymph node metastases?

3. For patients with soft tissue sarcoma, what is the overall 5-year survival rate?

4. In patients with soft tissue sarcoma who are treated with chemotherapy, local recurrence is most likely to occur within what period of time?

5. In patients with soft tissue sarcoma who are treated with chemotherapy, distant metastases are most likely to occur within what period of time?

Negatively-worded, unfocused stem; mixed options. Example 2

Weak:

Which of the following statements regarding bone metastases is **FALSE**?

A. After surgical stabilization of a metastasis that compromises the strength of the cortex of supporting bone, postoperative irradiation is associated with a significant increase in functional status.
B. In patients with multiple bone metastases, hemibody irradiation can relieve pain in 73% of treated patients.
C. The most effective and safest doses of hemibody irradiation are 6 Gy in one fraction to the lower hemibody and 8 Gy in one fraction to the upper hemibody.
D. A comprehensive premedication program has been proven to decrease the acute radiation syndrome to acceptable levels.
E. Patients with metastatic breast or prostate cancer experience pain relief in more than 80% of cases when treated with hemibody irradiation.

(Key is C)

Better:

Positive stems from options:

1. A bone metastasis compromises the strength of the cortex of a supporting bone. The bone is surgically stabilized. If postoperative irradiation is administered, what is the most likely effect on functional status?

2. Of patients with multiple brain metastases, approximately what percentage experience pain relief after hemibody irradiation?

3. To decrease acute radiation syndrome to acceptable levels for patients with bone metastases, what is the most appropriate treatment plan?

4. More than 80% of patients with metastatic breast or prostate cancer experience pain relief when treated with which of the following?

In this case, rewritting the wrong-answer-but-correct-option (key) could require two answers to be filled in, which is not a psychometrically-sound format. However, two stems could be constructed:

1. What is the safest effective dose of hemibody irradiation to the lower hemibody?

2. What is the safest effective dose of hemibody irradiation to the upper hemibody?

Further revisions for clarity and specificity may be needed.

Negatively-worded, unfocused stem Example 3

Weak:

Which of the following statements about malignant tumors of the heart and pericardium is **FALSE**?

A. Angiosarcoma is the most common tumor.
B. Rhabdomyosarcoma is the most common tumor.
C. Radiation therapy is of limited value for angiosarcoma of the heart.
D. Radiation therapy is of limited value for rhabdomyosarcoma of the heart.

(Key is A)

Better:

What is the most common malignant tumor of the heart and pericardium?

A. Angiosarcoma
B. Rhabdomyosarcoma
C. Metastatic melanoma
D. Myxoma sarcomatosum

Negatively-worded, unfocused stem; mutually-exclusive options. Example 4

Weak:

Which of the following is **NOT** true regarding carcinoma of the male urethra?

A. Primary lymph node drainage is to the inguinal nodes.
B. The most common histology is squamous cell carcinoma.
C. Lesions of the distal urethra are associated with best prognosis.
D. Lesions of the bulbomembranous urethra are associated with best prognosis.

(Key is D)

Negative stem aside, this is a weak item because Options C and D are mutually exclusive. Both C and D cannot be true statements. Because A and B are not involved in the exclusivity, the pattern immediately renders them invalid. Therefore, either C or D must be the correct (i.e., false) statement. The structure should not hint at the correct option, and all options should be feasible alternatives.

In this case, the best solution is to create multiple positive items.

Better:

1. In carcinoma of the male ureter, primary lymph node drainage is to which of the following?

2. In carcinoma of the male ureter, what is the most common histology?

3. In carcinoma of the male ureter, what is the prognosis if there are lesions in the distal urethra?

4. In carcinoma of the male ureter, what is indicated by lesions in the bulbomembranous urethra?

Double negative stem. Example 5

Weak:

False statements about treating mesothelioma with radiotherapy include all of the following **EXCEPT**:

Better:

Positive delivery.

Which of the following is the most appropriate method for treating mesothelioma with radiotherapy?

Unfocused stem; mixed options. Example 6

Weak:

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

(Key is B)

Better:

Make multiple, focused items.

A patient has metastatic lung disease. Which of the following types of cancer is most effectively treated by resection?

A. Breast cancer
B. Bone sarcomas
C. Ovarian cancer
D. Brain cancer

(Key is B)

In patients with pulmonary metastases who are candidates for resection, what is the most effective radiology study for tracking their cases?

etc.

Unfocused stem; does not pass the cover test; mixed options.
Example 7

Weak:

The depth-dose curve for a monoenergetic neutron beam:

A. has a Bragg peak at the end of the neutron range.
B. shows little penumbra broadening due to scattering.
C. is unaffected by the atomic composition of the tissue through which it travels.
D. is essentially exponential.

(Key is D)

Why:

What about it? Looking just at the stem, it is unclear what we are testing.

Better:

At what rate does the depth-dose curve for a monoenergetic neutron beam increase?

Unfocused stem; does not pass the cover test; mixed options.
Example 8

Weak:

Megavoltage photon beam off-axis factors:

A. increase with field size.
B. increase with depth.
C. depend on the flattening filter.
D. depend on the target.

(Key is C)

Why:

What about them? This stem does not pass the cover test.

Better:

What do megavoltage photon beam off-axis factors depend on?

Nonlinear delivery of information.
Example 9

Weak:

What position should you put a patient in for treatment when you suspect air embolism to the heart and great vessels?

Better:

A patient with suspected air embolism to the heart and great vessels should be placed in what position for treatment?

A. Supine
B. Erect
C. Right lateral decubitus
D. Left lateral decubitus
E. Prone

(Key is D)

Note: We do not use items written in the second person, as in the "weak" example here.

Nonlinear delivery of information. Example 10

Weak:

The "temperature corrected" reading of an ionization chamber at 75° F exposed to the atmosphere is calculated by multiplying the reading by:

A. 0.9936
B. 1.0032
C. 1.0064
D. 1.0859

(Key is C)

Why:

The information is presented in an unorderly fashion, which complicates the thought process. Information should be presented as:

background info + situational info + request for solution

Better:

At 75° F, an ionization chamber exposed to the atmosphere should have its reading "temperature corrected" by multiplying the reading by which of the following?

Nonlinear delivery of information. Example 11

Weak:

What is the percentage of 80-keV photons attenuated by a 4-cm-thick slab of bone when the mass attenuation coefficient of bone with a density of 1.8 g/cm^3 is 0.2 cm^2/g for an 80-keV gamma ray?

A. 36%
B. 45%
C. 55%
D. 64%
E. 76%

(Key is E)

Better:

The mass attenuation coefficient of bone with a density of 1.8 g/cm^3 is 0.2 cm^2/g for an 80-keV gamma ray. For a slab of bone 4 cm thick, what is the percentage of 80-keV photons attenuated?

A. 36%
B. 45%
C. 55%
D. 64%
E. 76%

(Key is E)

Nonlinear delivery of information; options do not properly complete stem.
Example 12

Weak:

The differentiation between various tissues is set by a basic factor in computed tomography, which is the:

A. providing x-ray beams with sufficient collimation.
B. effect of scattered radiation in reducing contrast.
C. accurate computer algorithms for reconstruction.
D. ability to design x-ray tubes with sufficiently high outputs.
E. dependence of statistical variations in photon transmission on patient dose.

(Key is E)

Better:

In computed tomography, what is the basic factor that sets a limit to differentiation between various tissues?

A. Ability to provide x-ray beams with sufficient collimation
B. Effect of scattered radiation in reducing contrast
C. Accuracy of computer algorithms for reconstruction
D. Ability to design x-ray tubes with sufficiently high outputs
E. Dependence of statistical variations in photon transmission on patient dose

Nonlinear delivery of information; run-on sentence.
Example 13

Weak:

A patient is experiencing persistent extravasation of contrast from a prepyloric ulcer after embolization of the gastroduodenal artery (GDA) for treatment of bleeding, which shows on a common hepatic arteriography that also shows occlusion of the GDA. The source most likely is:

Better:

A patient with a bleeding prepyloric ulcer undergoes embolization of the gastroduodenal artery (GDA). Common hepatic arteriography shows occlusion of the GDA but persistent extravasation of contrast from the ulcer. What is the most likely source?

A. Left gastric artery
B. Right gastric artery
C. Left gastroepiplotic artery
D. Right gastroepiplotic artery

(Key is B)

Ideas to Get You Started

The following may help as you start formulating items, and as you "polish" items that are not quite what you want them to be. These are reprinted from *Developing and Validating Multiple-Choice Test Items* by Thomas M. Haladyna with permission from publisher Lawrence Erlbaum Associated (Mahwah, NJ, 1999).

Understanding—Concepts
Which is the best definition of this (concept)?
Which is the meaning of this (concept)?
Which is synonymous with this (concept)?
Which is like this (concept)?
Which is a characteristic of this (concept)?
Which is an example of this (concept)?

Understanding—Principles
Which is the best definition of…?
Which statement below exemplifies the principle of…?
Which is the reason for or cause of…?
Which is the relationship between … and…?
Which is an example of the principle of…?

Critical Thinking—Predicting Using a Principle
What would happen if…?
If (there is an action), then what would happen?
What is the consequence of (an action)?
What is the cause of (a result)?
Information given. What is the expected result?
Which distinguishes (one concept from another concept)?

Critical Thinking—Evaluating Using Facts and Concepts
Which is the most or least important, significant, effective…?
Which is better, higher, lower, farther, nearer, heavier, lighter…?
Which is most like …?
What is the difference between … and…?
What is the similarity between … and…?

Critical Thinking—Evaluating Using a Principle
Which of the following principles applies to evaluating something?
What is the most important factor contributing to…?

Problem-Solving—Concepts, Principle, Procedures
Problem presented. What is the best way to solve this problem?
Problem presented. What is the solution?

Defining
What are the main symptoms of…?
What is the most common (cause, complication, symptom, or consequence) of (a procedure, a disorder, or an action)?

Predicting
What is the most common (cause or symptom) of (a patient problem)?

Evaluating
Patient illness is diagnosed. Which treatment is likely to be most effective?

Applying
Information [about a patient problem] is presented. How should the patient be treated?

Bibliography

Case SM, Swanson DB. *Constructing Written Test Questions for the Basic and Clinical Sciences.* 3rd ed. Philadelphia, PA: National Board of Medical Examiners; 2002. Available at http://www.nbme.org/PDF/ItemWriting_2003/2003IWGwhole.pdf

Collins J. Education techniques for lifelong learning: writing multiple-choice questions for continuing medical education activities and self-assessment modules. *RadioGraphics* 2006;26:543–551.

Downing SM. Construct-irrelevant variance and flawed test questions: do multiple-choice item writing principles make any difference? *Academic Medicine* 2002;77:s103–104.

Haladyna TM. *Developing and Validating Multiple-Choice Test Items*. Mahwah, NJ; Lawrence Erlbaum Associated; 1999.

Harasym PH, Leong EJ, Violato C, Brant R, Lorscheider FF. Cuing effect of "all of the above" on the reliability and validity of multiple-choice test items. *Evaluation and the Health Profession* 1998;21:120–133.

Jozefowicz RF, Koeppen BM, Case S, Galbraith R, Swanson D, Glew H. The quality of in-house medical school examination. *Academic Medicine* 2002;77:156–161.

Raymond M, Neiiers B, Reid JB. *Test-Item Development for Radiation Technology*. Minneapolis, MN; The American Registry of Radiation Technologists; 2003.

Rodriguez MC. Three options are optimal for multiple-choice items: a meta-analysis of 80 years of research. *Educational Measurement: Issues and Practice* 2005;24:3–13.

Acknowledgment

Special thanks to Dr. David H. Hussey for his extensive preparation of the base document.