

# **Writing Multiple Choice Questions for Continuing Medical Education Activities and Self-Assessment Modules**

Jannette Collins, MD, MEd, FCCP

## **Abstract**

The multiple-choice question (MCQ) is the most commonly used type of test item on radiologic graduate medical and continuing medical education examinations. Now that radiologists are participating in the maintenance of certification process, there is an increased need for self-assessment modules that include MCQs, and persons with item-writing skills to develop such modules. Although principles of effective item-writing have been documented, violations of these principles are very common in medical education. Guidelines for test construction are related to development of educational objectives, defining levels of learning for each objective, and writing effective MCQs that test that learning. Educational objectives should be written in observable, behavioral terms that allow for an accurate assessment of whether the learner has achieved the objectives. Learning occurs at many levels, from simple recall to problem solving. The educational objectives and the MCQs that accompany those objectives should target all levels of learning appropriate for the given content. Characteristics of effective MCQs can be described in terms of the overall item, the stem, and the options. Flawed MCQs interfere with accurate and meaningful interpretation of test scores and negatively impact student pass rates. Therefore, in order to develop tests that are reliable and valid, items must be constructed that are free of such flaws. This review provides an overview of established guidelines for writing effective MCQs, a discussion of writing appropriate educational objectives and MCQs that match those objectives, and a brief review of item analysis.

## **Introduction**

The multiple choice question (MCQ) is the most common type of written test item used in undergraduate, graduate and post-graduate medical education [1]. MCQs can assess a broad range of learner knowledge in a short period of time. Because a large number of MCQs can be developed for a given content area, which provides a broad coverage of concepts that can be tested consistently, the MCQ format allows for test reliability. If MCQs are drawn from a representative sample of content areas that constitute pre-determined learning outcomes, they allow for a high degree of test validity. Critics of MCQs argue that MCQs are unable to test higher level learning. However, this criticism is more often attributed to flaws in the construction of the items rather than to their inherent weakness. Appropriately constructed MCQs result in objective testing that can measure knowledge, comprehension, application and analysis [2]. Disadvantages of MCQs are that they test recognition (choosing an answer) rather than recall (constructing an answer), allow for guessing, and are difficult and time-consuming to construct.

The principles of writing effective MCQs are well documented in educational measurement textbooks, the research literature, and in test-item construction manuals designed for medical educators [3-5]. Yet, a recent study from the National Board of Medical Examiners showed that violations of the most basic item-writing principles are very common in medical education tests [6].

The number of radiologists who will be writing MCQs is expected to increase as more radiologists develop self-assessment modules (SAM) for the American Board of Radiology's maintenance of certification (MOC) program. In a ten-year period, enrollees in MOC must complete 20 SAMs that include MCQs [7]. All diplomates certified in 2002 and beyond are automatically enrolled in the MOC program, and the ABR is encouraging all diplomates to enroll in MOC. MCQs are difficult and time-consuming to construct, even for those who have been formally trained in the construction of MCQs. Professional item writers plan on one hour or more to write one good item [8]. At a recent meeting of the American Roentgen Ray Society (ARRS 105<sup>th</sup> Annual Meeting, New Orleans, LA, May 14, 2005), a half-day workshop on the development of MCQs for SAMs was attended by ARRS committee members who volunteered to write and edit SAMs for the ARRS. Such training on a wider scale may be needed if a large number of radiologists become involved in SAM development.

The purpose of this review is to provide guidelines that can be used by radiologists in writing MCQs for SAMs and other continuing medical education materials, as well as medical student clerkship tests, and radiology resident in-service and written board examinations. Three areas will be addressed: 1) writing educational objectives, 2) defining levels of learning for each objective, and 3) writing effective MCQs that test that learning. This is followed by a brief discussion of item analysis.

### **Writing educational objectives and defining levels of learning**

Good test question writing begins with identifying the most important information or skill that is to be learned. A direct relationship between instructional objectives and test items must exist. Thus, test items should come directly from the objectives [2] and focus on important and relevant content, avoiding testing the knowledge of medical trivia. Controversial items should be avoided, especially when the knowledge is incomplete or the facts are debated [9]. Determining the appropriate test questions can be facilitated by reviewing the major subtopics of the article or other content, and identifying sentences that summarize main ideas or principles. From this, key facts can be written down as declarative sentences, creating a clear picture of what the student should learn. It has been suggested that if the written idea as an explicit statement, proposition, or principle forms an important part of the instruction, it is worth testing [10].

Objectives should be written in terms of specific learner behavior and not what the program will teach. They should define important knowledge or skills and be supported by the instruction provided through the educational program. Observable, measurable objectives allow for accurate assessment of whether the learner has achieved an objective. Examples of measurable terms are state, explain, list, identify, and compare. Non-measurable terms include know, understand, learn, or become familiar with. For example,

#### *Unmeasurable objective*

“Understand the appearance of pneumothorax on a supine chest radiograph.” (It is not clear how the student will show that he/she “understands.”)

#### *Measurable objective*

“Describe five findings of pneumothorax that can be seen on a supine chest radiograph.”  
(It is clear how the student will demonstrate learning, and the qualifier of “five” indicates a specific level of knowledge.)

In 1959, Bloom published a taxonomy of cognitive learning as a hierarchy of knowledge, comprehension, application, analysis, synthesis and evaluation [11]. Educators have adopted Bloom’s taxonomy for test development [12, 13] and some have simplified and collapsed it into three general levels [14]. The three levels include the following categories 1) knowledge (recall or recognition of specific information), 2) combined comprehension and application (understanding or being able to explain in one’s own words previously learned information and using new information, rules, methods, concepts, principles, laws and theories), and 3) problem solving (transferring existing knowledge and skills to new situations). A MCQ should test at the same level of learning as the objective it is designed to assess. Table 1 shows examples of MCQs and objectives for each level of learning.

If the desired outcome of an educational program involves having participants do more than recall facts, the program should be designed to enable learners to apply knowledge or skills. The program’s objectives and test questions should reflect different levels of learning. Thoughtfully written objectives are critical to the construction of appropriate test questions, and in ensuring adequate assessment of intended learner competence. MCQs written to test knowledge (lower level learning) would not be appropriate to test competence of objectives reflecting comprehension (higher level learning). For example, a MCQ asking the learner to recognize benign dermal calcifications on a mammogram does not test the learner’s problem solving ability. A question that provides specific patient information and imaging data (patient vignette), and asks the learner to choose the most appropriate management is an example of an item that tests problem solving ability. Such patient vignettes offer several benefits in addition to assessing application of knowledge. Because they require problem solving, they increase the validity of the examination. Such items are more likely to focus on important information, rather than trivia. Lastly, they help identify examinees who have memorized facts but are unable to use the information effectively.

## Guidelines for writing MCQs

Several authors have outlined the elements of good MCQs [1, 9, 10, 13, 15]. The National Board of Medical Examiners has published on its website a manual on constructing written test questions for the basic and clinical sciences, reflecting what the authors had learned in developing items and tests over the past 20 years [16]. Published guidelines should be viewed as best-practice rules and not absolute rules. In some circumstances, it may be appropriate to deviate from the guidelines. However, such circumstances should be justified and occur infrequently.

Terms are applied to the different components of MCQs. The “item” is the entire unit and consists of a stem and several options. The “stem” is the question, statement or lead-in to the question. The possible answers are called “alternatives”, “options”, or “choices.” The correct option is called the “keyed response.” The incorrect options are called “foils” or “distractors.”

The stem is usually written first and is best written as a complete sentence or question. Direct questions (e.g., Which of the following is an imaging feature of benign pulmonary nodules?) are clearer than sentence completions (e.g., Benign pulmonary nodules...). Research has shown that the use of incomplete stems lowers the students' correct response rate by 10% to 15% [17]. A stem can incorporate maps, diagrams, graphs, or radiologic images, but should be accompanied by a complete statement. Ideally, the item should be answerable without having to read all of the options. The stem should include all relevant information, only relevant information, and contain as much of the item as possible. If a phrase can be stated in the stem, it should not be repeated in the options. For example,

*Phrase repeated in each option*

Which of the following would decrease radiation dose by  $1/2$ ?

- A. Decreasing mA by  $1/4$
- B. Decreasing mA by  $1/3$
- C. Decreasing mA by  $1/2$
- D. Decreasing mA by  $3/4$

*Item that includes all relevant information in the stem*

By what fraction would mA need to be decreased to lower the radiation dose by  $\frac{1}{2}$ ?

- A.  $\frac{1}{4}$
- B.  $\frac{1}{3}$
- C.  $\frac{1}{2}$
- D.  $\frac{3}{4}$

The stem should be kept as short as possible and include only the necessary information. It should not be used as an opportunity to teach or include statements which are informative but not needed in order to select the correct option. Stems should not be tricky or misleading, such that they might deceive the examinee into answering incorrectly. The level of reading difficulty should be kept low using simple language so that the stem is not a test of the examinee's reading ability. As a general guide, students can complete between one and two multiple-choice items per minute [18, 19]. Items that significantly exceed this time to complete should be closely examined as to whether they are unnecessarily verbose or confusing.

The stem is generally longer when application of knowledge is being tested as opposed to recall of an isolated fact. To test application of knowledge, clinical vignettes can provide the basis for the question, beginning with the presenting problem of a patient, followed by the history (duration of signs and symptoms), physical findings, results of diagnostic studies, initial treatment, subsequent findings, etc. Vignettes do not have to be long to be effective, and should avoid verbosity, extraneous material and "red herrings." In a study that compared non-vignette, short vignette and long vignette MCQs [5], designed to require increasing levels of interpretation, analysis and synthesis, items were shown to be more difficult as patient findings were presented in a less interpreted form. However, the differences in discrimination were not statistically significant. Regardless of these psychometric results, vignette items are generally felt to be more appropriate because they test application of knowledge and thus improve the content validity of the examination [5]. For example,

*Item measuring recall*

“Which of the following presents as chronic (longer than 3 months) airspace disease on a chest radiograph?”

- A. Streptococcal pneumonia
- B. Adult respiratory distress syndrome
- C. Pulmonary edema
- D. Pulmonary alveolar proteinosis

*Item with a vignette measuring application of knowledge*

“A 30-year-old-man presented with a 4-month history of dyspnea, low-grade fever, cough and fatigue. Given the following chest radiograph, what is the most likely diagnosis?”

- A. Adult respiratory distress syndrome
- B. Pulmonary edema
- C. Streptococcal pneumonia
- D. Pulmonary alveolar proteinosis

The stem should be stated so that only one option can be substantiated and that option should be indisputably correct. It is wise to document (for later recall) the source of its validity. If the correct option provided is not the only possible response, the stem should include the words “of the following.” When more than one option has some element of truth or accuracy but the keyed response is the best, the stem should ask the student to select the “best answer” rather than the “correct answer.”

Questions should generally be structured to ask for the correct answer and not a “wrong” answer. Negatively posed questions are recognizable by phrases such as “which is not true” or “all of the following except.” Negative questions tend to be less effective and more difficult for the examinee to understand [9]. Negative stems may be good choices in some instances, but should be used selectively. When negative stems are used, the negative term (e.g., “not”) should be underlined, capitalized or italicized to make sure that it is noticed. For example,

*Negatively worded stem*

“Which of the following is NOT a characteristic CT finding of small airway disease?”

*Positively worded stem*

“Which of the following best distinguishes small airway disease from interstitial lung disease on chest CT?”

Absolute terms, such as “always”, “never”, “all” or “none” should not be used in the stem or distractors. Savvy examinees know that few ideas or situations are absolute or universally true [20]. The terms “may”, “could”, and “can” are cues for the correct answer, as testwise examinees will know that almost anything is possible. Imprecise terms such as “seldom”, “rarely”, “occasionally”, “sometimes”, “few”, and “many”, are not uniformly understood and should be avoided. In a study conducted at the National Board of Medical Examiners [5], 60 members of eight test committees who wrote questions for various medical specialty examinations reviewed a list of terms used in MCQs to express some concept related to frequency of occurrence and indicated the percentage of time that was reflected by each term. The mean value plus or minus one standard deviation exceeded 50 percentage points for more than half of the phrases. For example, on average, the item writers believed the term “frequently” indicated 70% of the time; half believed it was between 45% and 75% of the time; actual responses ranged from 20% to 80%. Of particular note is that values for “frequently” overlapped with values for “rarely.” Absolute numbers are better. For example, “In less than 15% of the population” is better than “rarely.”

Eponyms, acronyms or abbreviations without some qualification after each term should be avoided. Examinees may be unfamiliar with such terms, or the terms may have more than one meaning. In such cases, the item becomes a test of whether the examinee understands the meaning of a term, or the item is faulty because a term can be interpreted in more than one way.

The most challenging aspect of creating MCQs is designing plausible distractors. The ability of an item to discriminate (i.e., separate those who know it from those who don't) is founded in the quality and attractiveness of the distractors. The best distractors



are statements that are accurate but do not fully meet the requirements of the problem, and incorrect statements that seem right to the examinee [20]. Each incorrect option should be plausible but clearly incorrect. Implausible, trivial, or nonsense distractors should not be used. Ideal options represent errors commonly made by examinees. Distractors are often conceived by asking questions such as, “What do people usually confuse this entity with?”, “What is a common error in interpretation of this finding?” or “What are the common misconceptions in this area?”

The best number of options is three to five. Research has shown that three-option items are as effective as four-choice options [21]. Constructing more than five is burdensome and often leads to faulty options while increasing the reading demands of the student. Furthermore, there is no hard and fast rule that the number of options needs to be uniform [18]. In one examination, some items may have four options and some may have five.

Distractors should be related or somehow linked to each other. That is, they should fall into the same category as the correct answer (e.g., all diagnoses, tests, treatments, prognoses, disposition alternatives). For example, all options might be a type of pneumonia or radiation dose.

The distractors should appear as similar as possible to the correct answer in terms of grammar, length, and complexity. There is a common tendency to make the correct answer substantially longer than the distractors. For example,

*Item with a keyed response longer than the other options*

An otherwise healthy 28-year-old woman presented with a two-day history of cough, fever and shortness of breath, and the following chest radiograph. What is the most likely diagnosis?

- A. Tuberculosis
- B. Community-acquired streptococcal pneumonia
- C. Varicella pneumonia
- D. Blastomycosis

*Item with options of similar length*

An otherwise healthy 28-year-old woman presented with a two-day history of cough, fever and shortness of breath, and the following chest radiograph. What is the most likely diagnosis?

- A. Tuberculosis
- B. Streptococcal pneumonia
- C. Varicella pneumonia
- D. Blastomycosis

Options should not include potentially offensive or unfair material to selected groups of examinees. Therefore, references to gender or race should be made only when necessary.

The options should not stand out as a result of their phrasing . Grammatical cues, such as when one or more options don't follow grammatically from the stem, lead the examinee to the correct option. If the stem is in past tense, all the options should be in past tense. If the tense calls for a plural answer, all the options should be plural. Stem and options should have subject-verb agreement. Because an item writer tends to pay more attention to the correct option than to the distractors, grammatical errors are more likely to occur in the distractors. This is not an issue when the stem is written as a question. For example,

*Option C does not follow grammatically from the stem*

“A chest radiographic finding of left upper lobe collapse is \_\_\_\_\_.”

- A. Hyperlucency of the upper and lower left hemithorax.
- B. Elevation of the left diaphragm.
- C. When the mediastinum shifts to the right.
- D. Posterior displacement of the minor fissure.

*All options follow grammatically from the stem*

“A chest radiographic finding of left upper lobe collapse is \_\_\_\_\_.”

- A. Hyperlucency of the upper and lower left hemithorax.

- B. Elevation of the left diaphragm.
- C. Mediastinal shift to the right.
- D. Posterior displacement of the minor fissure.

Options should not include “none of the above” or “all of the above.” None of the above is problematic in items where judgment is involved and where the options are not absolutely true or false. If the correct response is intended to be one of the other listed options, knowledgeable examinees can be faced with a dilemma because they have to decide between a very detailed perfect option and the one that is intended as correct. Examinees can often construct an option that is more correct than the one intended to be correct. Use of “none of the above” turns the item into a true/false item; each option has to be evaluated as more or less true than the universe of unlisted options [16]. “None of the above” only informs about what the examinee knows is not correct and not what is correct. The examinee only needs to recognize that two of the options are correct for “all of the above” to be the correct option.

Options should be placed in logical order, if there is one. For example, if the answer is a number, the options should begin with the smallest and proceed to the largest (it is also acceptable to begin with the largest and proceed to the smallest). If the options are dates, they should be listed in chronological order. Options should be independent and not overlap with each other. For example,

*Item with overlapping options*

What is the average effective radiation dose from chest CT?

- A. 1-8 mSv
- B. 8-16 mSv
- C. 16-24 mSv
- D. 24-32 mSv

*Item without overlapping options*

What is the average effective radiation dose from chest CT?

- A. 1-7 mSv

- B. 8-15 mSv
- C. 16-23 mSv
- D. 24-32 mSv

Items must be independent of one another, such that one item does not reveal information that allows the examinee to automatically know the correct answer to another item. This is referred to as “cueing”, when an option in one item provides a hint to the answer to another item. It is also important to avoid “hinging”, where questions require that students know the answer to one item in order to answer another item.

The position of the keyed response should vary from the A, B, C and D positions. Research shows that the B or C position is overused [21]. Testwise examinees, familiar with this tendency, will choose B or C to increase the likelihood of getting the answer right when they don't know the answer and are forced to guess.

### **Item analysis**

Items that attempt to assess critically important topics cannot do so unless they are well-structured. Flaws that benefit the testwise examinee (e.g., grammatical cues; use of terms such as “always” or “never”; and the correct answer being longer than the other options), and items with irrelevant difficulty (e.g., long or complicated options; inconsistently stated numerical data; use of vague terms such as “rarely” or “usually”; use of “none of the above”; and tricky or unnecessarily complicated stems) must be avoided in order for MCQs to generate valid scores. Several item-writing principles have been investigated for their effects on test psychometric indices [4]. Most studies evaluate the effect of a single-item flaw, such as negative stems [6] and none of the above option [22]. Downing [22] evaluated the validity of a classroom achievement test in medical education that contained flawed test items. Eleven (33%) of the 33 items were classified as flawed (unfocused item stems, use of “none of the above” and “all of the above” and negative stem). He found that flawed items failed nearly one fourth more students than non-flawed items. The increased test and item difficulty associated with the use of flawed items is an example of construct-irrelevant variance, because poorly crafted test questions add artificial difficulty to the test scores. This variance interferes with the

accurate and meaningful interpretation of test scores and negatively impacts students' passing rates, particularly for passing scores at or just above the mean of the test score distribution.

Authors of MCQs should review their items for accuracy and appropriate formatting. However, just as with any editorial work, internal review may not reveal all errors. It can be very beneficial to have a colleague read and respond to the MCQs, and offer feedback. Many institutions have testing services that can analyze the quality of test items for faculty. As MCQs become more widely utilized in the MOC process, organizations that provide CME activities and SAMS should consider providing professional assistance with item writing and item analysis. Figure 1 provides a list of guidelines for writing effective MCQs that can be referenced when proofing test items.

MCQs can be evaluated according to their reliability, validity, and resource intensiveness [23, 24]. Reliability provides a measure of an item's generalizability. Items in a test represent a small sample of all the possible MCQs that could be asked, and the test score should be indicative of the score of the same student on any other set of relevant items. Validity refers to the extent that a test measures what it claims to measure. Resource-intensiveness is determined by the costs of constructing and grading items. MCQs are relatively easy to grade, especially with computer assistance, but are difficult and time-consuming to construct.

Item analyses provide a numerical assessment of item difficulty and item discrimination. To determine item difficulty, the percentage of students who answered each item correctly is calculated. The goal is to construct a test that contains only a few items that more than 90% or less than 30% of students answer correctly [20]. Optimally, difficult items are those that about 50 to 75 percent of the students answer correctly. Items are considered low to moderately difficult if between 70 and 85 percent of the students select the correct response.

Item discrimination refers to the percentage difference in correct responses between two groups of students (generally referring to students in the top 25% and the lower 25%). The discrimination ratio for an item will fall between -1.0 and +1.0. The closer the ratio is to +1.0, the more effectively that item distinguishes students who know the material (the top group) from those who don't (the bottom group). Ideally, each item

will have a ratio of at least +.5 [20]. An item with a discrimination of 60% or greater is considered a very good item, while a discrimination of less than 19% indicates a low discrimination item that needs to be revised [15]. An item with a negative index of discrimination indicates that the poor students answer correctly more often than do the good students, and such items should be avoided.

## **Summary**

As the demand for continuing medical education materials and SAMs increases, so does the need for individuals skilled in item-writing. Radiologists, typically not trained in item-writing, will be one group of individuals called upon to develop these materials. Radiologists are generally not familiar with how to write measurable educational objectives and MCQs that match those objectives in terms of the level of learning involved. Beyond that, effective item construction requires a knowledge of established item-writing principles. Figure 1 provides a list of guidelines for overall item-writing and for writing effective stems and options. This list can be referenced by radiologists who are writing MCQs for students at all levels (i.e., medical students, residents and practicing radiologists). It is important for test developers to be skilled in effective item-writing to ensure that the materials used to evaluate learners are valid assessments of a learner's knowledge. Measurement of a learner's knowledge is an important step in the educational process that should be afforded the same attention as does the development and implementation of curricula. The results of measurements of learning are used in establishing future learning goals, which completes the continuous cycle of learning.

## References

1. Farley JK. The multiple choice test: writing the questions. *Nurse Educator* 1989 14:10-12, 39
2. Kemp JE, Morrison GR, Ross SM. Developing evaluation instruments. In: *Designing Effective Instruction*. New York, NY: MacMillan College Publishing Company, 1994:180-213
3. Gronlund NE. *Assessment of student achievement*. Boston, MA: Allyn and Bacon, 1998
4. Haladyna TM, Downing, SM, Rodriguez MC. A review of multiple-choice item-writing guidelines. *Applied Meas Educ* 2002; 15:309-333
5. Case SM, Swanson DB. *Constructing written test questions for the basic and clinical sciences*. Philadelphia, PA: National Board of Medical Examiners, 1998
6. Jozefowicz RF, Koeppen BM, Case S, Galbraith R, Swanson D, Glew H. The quality of in-house medical school examinations. *Acad Med* 2002; 77:156-161
7. ABR-MOC. [http://www.theabr.org/MOCrebuild\\_overview.htm](http://www.theabr.org/MOCrebuild_overview.htm). Accessed 5-27-05
8. Van Hoozer H, et al. *The teaching process: theory and practice in nursing*. Norwalk, CT: Appleton-Century-Crofts, 1987:279-280
9. Braddom CL. A brief guide to writing better test questions. *Am J Phys Med Rehabil* 1997; 76:514-516
10. Cox KR, Bandaranayake R. How to write good multiple choice questions. *Med J Aust* 1978; 2:553-554
11. Bloom BS, ed. *Taxonomy of educational objectives. Vol. I: Cognitive Domain*. New York: McKay, 1956
12. Fuhrmann BS, Grasha AF. *A practical handbook for college teachers*. Boston: Little, Brown, 1983, p. 170
13. Schultheis NM. Writing cognitive educational objectives and multiple-choice test questions. *Am J Health-Syst Pharm* 1998; 55:2397-2401
14. Crooks TJ. The impact of classroom evaluation practices on students. *Review of educational research*. 1988 58:438-481

15. Vydareny KH, Blane CE, Calhoun JG. Guidelines for writing multiple-choice questions in radiology courses. *Invest Radiol* 1986; 21:871-876
16. The National Board of Medical Examiners. Constructing written test questions for the basic and clinical sciences.  
<http://www.nbme.org/about/itemwriting.asp>. Accessed 5/25/05
17. Kent TH, Jones JJ, Schmeiser CB. Some rules and guidelines for writing multiple choice test items. Iowa City: University of Iowa College of Medicine and American College Testing Program. 1974:6
18. Srinivasa DK, Adkoll BV. Multiple choice questions: how to construct and how to evaluate? *Indian J Pediatr* 1989; 56:69-74
19. Lowman J. *Mastering the techniques of teaching*. San Francisco: Jossey-Bass, 1984
20. Davis BG. *Tools for teaching*. Jossey-Bass Publishers. San Francisco, CA 1993, chapter 30
21. McKeachie WJ. *Teaching Tips*. 8<sup>th</sup> ed. Lexington, Mass.: Heath, 1986
22. Downing SM, Baranowski RA, Grosso LJ, Norcini JJ. Item type and cognitive ability measured: the validity evidence for multiple true-false items in medical specialty certification. *Appl Meas Educ* 1995; 8:189-199
23. Schuwirth LWT, van der Vleuten CPM. Different written assessment methods: what can be said about their strengths and weaknesses? *Medical Education* 2004; 38:974-979
24. van der Vleuten CPM. The assessment of professional competence: developments, research and practical implications. *Adv Health Sci Educ* 1996; 1:41-67



**Table 1.** Examples of Objectives and MCQs for three level of learning.

<b>Level</b>	<b>Objective</b>	<b>Question</b>
<i>Knowledge</i> (learner must recall memorized information but not explain or apply it)	State the average effective radiation dose from chest CT.	What is the average effective radiation dose from chest CT? A. 1 mSv B. 8 mSv C. 16 mSv D. 24 mSv
<i>Combined comprehension and application</i> (learner must demonstrate an ability to use, not just explain, new information, applying rules, methods, concepts, principles, laws or theories)	Compare the radiation exposures from different radiologic examinations.	Which of the following imaging examinations is associated with the highest effective radiation dose? A. Abdomen/pelvic multidetector CT B. Coronary artery multidetector CT C. Conventional pulmonary angiogram D. Digital pulmonary angiogram
<i>Problem solving</i> (learner must understand a concept's components and their relationships to each other, and analyze information)	Explain the effects that various factors have on radiation dose from chest CT.	Which of the following would decrease the radiation dose from chest CT the least? A. Decreasing mA from 250 to 125 B. Decreasing kVp from 140 to 120

		C. Decreasing the pitch from 2 to 1 D. Decreasing scan time from 1 to 0.5
--	--	--

**Figure 1.** Guidelines for writing effective MCQs

*Items*

- Relate items to instructional objectives
- Test at the same level of learning as the objectives are designed to assess
- Write items to reflect different levels of learning

*Stems*

- Provide a complete statement
- Include only relevant information
- Contain as much of the item as possible in the stem
- Keep stems as short as possible
- Ask for the correct, not “wrong” answer
- Avoid absolute terms such as “always”, “never”, “all”, or “none”
- Avoid imprecise terms such as “seldom”, “rarely”, “occasionally”, “sometimes”, “few”, or “many”
- Avoid cues such as “may”, “could” or “can”
- Define eponyms, acronyms, or abbreviations when used

*Options*

- Keep options grammatically consistent with the stem
- Write incorrect options to be plausible but clearly incorrect
- Link options to each other (e.g., all diagnoses, tests, treatments)
- Write distractors to be similar to the correct answer in terms of grammar, length, and complexity
- Avoid “none of the above” or “all of the above”
- Place options in logical order (e.g., numerical, chronological)
- Write options to be independent and not overlapping

- Vary position of keyed responses