Education

Writing Classroom Tests

Maria V. Nagel

University of Nebraska Medical Center, Omaha, Nebraska

A step-by-step approach for developing classroom tests is described for nuclear medicine technology educators.

Health personnel often become teachers in their specialty because they excel as practitioners. They do not have formal instruction in education but are asked to teach because of their practical expertise. The new or less experienced teacher faces many challenges. One challenge is writing classroom tests. Steps for writing, assembling, administering, and analyzing tests for nuclear medicine technology students will be described. These steps stem from a workshop presented at the 1982 annual meeting of the Society of Nuclear Medicine, using material developed by the University of Kentucky under a Kellogg Grant. The university's program is called TIPS (Teaching Improvement Project System) and is used at 16 sites in the United States, Canada, and South America that present workshops for teacher development.

Objectives

Adjectives such as *behavioral, enabling, planning,* or *instructional* are used to describe particular types of objectives. Basically, an objective is information telling students what they will be able to do after some identified instruction. Why are objectives used in teaching? The primary reason is to guide the instructor in teaching and students in learning the subject matter. Suggested components of an objective are (A) the *audience* for which the objective is intended, (B) the *behavior* to be measured, (C) the *conditions* under which the behavior is to be learned, and (D) the *degree* to which the behavior is to be learned. These are easily remembered as the ABCDs of an objective.

Consider this objective: "After class discussion, the nuclear medicine technology student will correctly identify the parts of a gamma camera as given in class." It contains the audience (A)—"the nuclear medicine technology student," the behavior (B)—"identify the parts of a gamma camera," the conditions (C)—"after class discussion," and the degree (D)—"correctly, . . . as given in class." It also gives the instructor guidance in what to teach and the student the important features of the instruction.

Nuclear medicine educators have many sources or goals on which to base objectives. These include the JRCNMT *Essentials* (*1*), the NMTCB Task Analysis (*2*), and the Nuclear Medicine Technologist Position Description (*3*). General goals for a particular college, university, or hospital and course descriptions in program curricula are other sources. Goals state general outcomes desired from an instructional unit, course, or program.

A starting place for each program's objectives is the *Curriculum Guide for Nuclear Medicine Technologists*, which gives suggested unit outlines and matches objectives to outlines (*4*).

Properly written objectives ask the student to exhibit more than the knowledge-level behaviors of *naming* and *listing*. The student should have to apply knowledge and solve problems. Objectives are built on a hierarchy beginning with knowledge and proceeding to application and problem-solving.

An example is: 1) Knowledge—after classroom discussion, the nuclear medicine technology student will correctly define all symbols in the decay equation:

$$A = A_O e^{\frac{-0.693 \cdot t}{T_{1/2}}}.$$

2) Application—given A_O, the radionuclide, and t, the elapsed time, the student will correctly solve for A. 3) Problem solving—given a particular nuclear medicine study, time of radiopharmaceutical injection, and assay, the student will correctly determine the amount of radiopharmaceutical to be drawn up at the preadministration time.

The above is an example of cognitive domain but there are also psychomotor and affective domains, which require the student to perform a task or have a certain feeling about a situation. Written tests predominantly use cognitive domain objectives.

Testing Purposes

Why test? The three purposes that prevail are instructional,

For reprints contact: Maria V. Nagel, University of Nebraska Medical Center, 42d and Dewey Ave., Omaha, NE 68105.

mastery, and measurement evaluations. Instructional evaluation usually takes the form of pop quizzes. They assist both instructor and student in identifying strengths and weaknesses and emphasize specific instructional areas.

Mastery evaluation assumes that the student must master basic knowledge before proceeding to the next instructional level. It is often called "criterion referencing" since the student's performance is compared with a pre-established set of criteria. A student is not compared to other students.

Measurement evaluation compares a student's performance to the group and is called "norm referencing." This type of evaluation is usually used to provide grades.

For a written test, however, it would not be evident which of the above evaluations was in use since these purposes do not relate to actual types of exam items but to the use of the exam.

Matching Objectives to Test Items

Test items are ideally developed after writing objectives but before teaching the subject. When constructing test items, the cognitive level of each objective is identified first and then a list is generated of the various types of questions that could measure it. The process is outlined in this example:

Topic: Decay Equation

Objective: After classroom discussion, the nuclear medicine technology student will correctly define all symbols in the decay equation

$$A = A_0 e^{\frac{-0.693 \cdot t}{T_{1/2}}}.$$

Cognitive level: Knowledge

Acceptable behaviors:

Write a sentence definition of each term.

Match each term with the appropriate definition.

Select the best definition from four or five alternatives.

Test item:

Match the definitions on the right with the appropriate decay equation symbol on the left. Each response may be used once.

Symbols	*Definitions*
1. A	a. Common logarithm of 2
2. A_0	b. Elapsed time
3. e	c. Half-life of radionuclide
4. t	d. Natural logarithm base
5. $T_{1/2}$	e. Natural logarithm of 2
6. 0.693	f. Radionuclide amount after certain time
	g. Radionuclide amount at start

All the above material may be placed on a card or entered into a computer, thus beginning a test item bank.

A test blueprint, also called a test plan or table of specifications, assists in giving an overall view of the test. The blueprint includes the purpose of the test, the content areas from a course outline, objectives matched with the content area, relative emphasis for each content area, the item types, the number of items, and the point value for each item. A blueprint may be altered as the material is taught and is not shared with students.

Constructing Test Items

Testing determines what information the student has obtained from instruction. It is not used to confuse or trick the student. Test items may be divided into selection and supply categories. The selection category includes true-false, matching, and multiple choice items. Completion and essay are in the supply category; they ask the student to write a response.

True-false (T-F) items allow rapid testing of a large amount of content and are easy to score. However, they are difficult to construct because the statements must be totally true or false. Other problems include the likelihood of testing trivia and the probability of guessing since the student has a 50% chance of getting the item correct. Guidelines for writing T-F items include:

1. Use and construct T-F only when there is a totally true or false answer.
2. Avoid using "always," "never," or similar words that signal the correct answer.
3. Construct questions in your own words rather than quoting directly from a textbook.
4. Use positive rather than negative statements to avoid misreading.
5. Have approximately the same number of true and false statements but avoid setting up an answer pattern.
6. Instead of asking one T-F item for a concept, construct up to ten.
7. Use novel ways for T-F such as providing a situation and giving T-F statements about it.

Matching items are easy to score but difficult to construct. These items consist of a set of directions followed by two columns containing homogeneous material to match with each other. They are usually used for knowledge level objectives. Guidelines for effective use include:

1. Use more than five items for the match.
2. Make the lists homogeneous such as all gamma camera parts and corresponding definitions.
3. Label each column.
4. Arrange the columns in a logical sequence, for example, alphabetically or numerically.
5. When each match is to be used once, have one list contain more items. An unequal match avoids answering by elimination and having two answers wrong.
6. Give explicit directions, such as whether the items may be used more than once.
7. Make the statements brief since each response must be read a number of times to complete the item.

Multiple choice items contain a stem or statement followed by alternatives or options from which the student selects. These items are the most difficult to construct; yet they are easy to score and have the advantage of measuring application and problem-solving levels. They are the item type used in the Nuclear Medicine Technology Certification Board exam and thus should be used to a large extent in nuclear medicine technology training programs to familiarize students with multiple choice exams (5). Effective multiple choice items can be written following these guidelines:

1. Have a single definite concept in the stem. A knowledgeable student should be able to answer the question without looking at the response alternatives. Stems such as: "The nuclear medicine technologist," "The gamma camera," or "In vitro" each followed by five alternatives will not give the student enough information to answer the items.
2. Common wording belongs in the stem. The student should not have to waste time rereading the same words in the alternatives before getting to the actual information.
3. Use positive statements in the stem; underline and capitalize negatives if it is necessary to use them.
4. The stem and all its alternatives should be grammatically consistent. Do not end a stem with "a" if vowels begin some alternatives. Do not make some alternatives nouns and others adjectives.
5. Do not give away the answer by using "sound-alikes" or the same word in the stem and an answer.
6. Have a single best answer. An example of more than one plausible answer is alternatives such as "(a) less than 20, (b) less than 40, (c) more than 20, (d) more than 40, (e) more than 60."
7. Give at least four alternatives to choose from. Guessing is increased as the number of alternatives is reduced.
8. Include plausible wrong answers. Obviously wrong answers increase the chances of guessing.
9. Use a logical sequence to arrange the alternatives. For example, arrange alternatives in increasing numerical order or alphabetically.
10. Construct alternatives that are not opposites. Opposites increase the chances of guessing because the student may eliminate the other choices and choose one of the opposing alternatives.
11. Make all alternatives the same length. The longest or shortest alternative is often chosen as the correct answer.
12. Write alternatives without using "always," "never," or similar words that tell the student to eliminate it.
13. Use "All of the above" and "None of the above" with caution, particularly in the same item. These are often used only to increase the number of alternatives.
14. Use "you should" instead of "you would" in the stem. "You would" asks for an opinion; thus, all responses are correct. "You should" stipulates the correct action.

Completion items include brief answers to questions and fill-in-the-blanks. These items are easy to construct but can be difficult to score since partial credit may be given to students who give an answer that is correct—but not the one intended. Guidelines include:

1. Construct the items to elicit only one answer.
2. Place the missing word or phrase at the end of the statement to avoid unnecessary rereading.
3. For multiple answers, have all the blanks the same length to avoid signaling the correct answer.

Essay items allow testing at the application and problem-solving levels but require careful construction. Scoring can be very difficult and time consuming. Always indicate exactly what should be discussed in the response. An essay item that states, "Compare and contrast nuclear medicine and ultrasound" is not explicit. The student could describe the instrumentation, the studies, the location of each in his institution, or any other comparison.

Before scoring essay items, the instructor should write down the key points. When scoring essays, remove student identification to help eliminate bias. Reading through all the essays will give the instructor an idea of the range of responses. The same essay should be read and scored on all papers before scoring the next essay. Writing your comments gives feedback. A final reading of the paper to check and correct anything missed in the first reading is advisable.

The choice of item type is the instructor's. To make the choice, consider first the cognitive level most appropriately tested by a particular item type. Ease in scoring and constructing items are two practical considerations that also enter into the decision. Frequent student practice on all item types will assure a wide range of test-taking skills.

Assembling the Test

After the objectives and test items are written, the test is assembled. For the student, testing is a high-anxiety situation, so the test should be assembled clearly to alleviate ambiguities. For the instructor, a clear test aids in scoring.

The instructor can group test items by objectives or item types. If more than one instructor is involved in constructing the test, item grouping by objectives taught by each instructor would be appropriate. Grouping by item type allows the student to complete each type instead of jumping back and forth and possibly misreading directions for a particular type.

The first items on a test appear to be the hardest for students and are most commonly missed because of beginning anxiety and tension. Placing easy items at the start of the test will ease the tension. If more than one item type is used, the number should be limited to two or three at most. The use of a separate answer sheet aids in scoring. Identifying the test—title, date, and instructor—is helpful. Directions should be specific for each item including how to record answers, and the time limits and point values of each item.

Giving the test to a nuclear medicine technologist or an advanced student will help the instructor find errors in directions and test items. The final copy should be carefully proofread for typographical errors. Make sure that an entire item is placed on one page, and that all information is included. The copies given to the student should be as clear as the original.

Administering the Test

The instructor should be aware of any distracting circumstances during the test. Factors in the testing situation should be as equal as possible for each student. One student may be able to concentrate on the test with a conversation being conducted in the hall while another student listens to the conversation, thus is distracted from the test.

There are other areas to be considered as well. Setting—Do the students have desks or table tops to write on? Is the lighting adequate and room temperature moderate? Equipment—Have students been asked to supply calculators or other special equipment or will the instructor supply them? Directions—Are test directions written explicitly? Oral instructions may cause misinterpretation. Any changes in the test are preferably given at the start of the exam by writing them on the blackboard. Teacher behavior—Is the instructor distracting students during the test?

Test reliability and validity are affected when students are distracted.

Analyzing Test Results

After the test is given, results are analyzed to assist the instructor in drawing conclusions. The instructor is the subjective portion of any objective test analysis. These analyses are intended to assist the instructor in improving teacher-made tests.

Going over the test with students gives both the instructor and student feedback on the items and instructional material. It is important that the instructor gather information to improve the learning situation without being defensive about the exam.

Statistical analyses also aid the instructor. Multiple choice, matching, and true-false items lend themselves to these types of analyses better than completion and essay. Many institutions will make their computers available for test scoring and analysis. Analysis often includes students' scores expressed as standard scores and percentiles. Item analysis, reliability coefficients, and standard errors of measurement for the entire test can also be computed. If you use a computer, refer to the descriptive information provided with the printout to help determine which parts of the analysis are useful.

For hand-derived analyses of norm-referenced tests, the instructor will want to know the "difficulty" and "discrimination" of each item. Item difficulty gives the instructor an idea of how difficult the item is for a selected group of students. To obtain the difficulty of each item, the instructor totals the scores of each paper and arranges the papers from high to low scores. If there are fewer than twenty students in the class, the papers are divided into two equal groups: one of high scores and the other of low scores. For each item in the exam, the number in each group who gave the correct answer is tallied. The difficulty index percentage is calculated by adding together the correct number of answers from each group, dividing by the total number of papers in both groups, and multiplying the result by 100. The closer the percentage is to 100, the easier the item. The instructor decides on the desired difficulty for future use of the test item and range of difficulty for the test. This decision may be based on the test type, the cognitive level of objectives and test items, and the test's purpose.

For classes with more than 20 students, the top 33–50% and bottom 33–50% of the scores are used for analysis.

Item discrimination determines how well an item distinguishes between high and low scorers. To calculate item discrimination, follow the steps for item difficulty through tallying the number who gave the correct answer in the two groups. Then subtract the number in the low group from the high group and divide by one-half the total number of papers. This gives the discrimination. The range for item discrimination is $+1.00$ to -1.00. ($+1.00$ is the highest discrimination index; it indicates that only the top scorers answered the item correctly.) Minus indices show poor discrimination and zero shows no discrimination. Use of positive-discrimination items is suggested. The instructor matches item discrimination and difficulty to determine which items to retain or rewrite. Similar methods apply when doing item discrimination of mastery tests.

A table can be constructed showing the tally, difficulty and discrimination indices of each test item. This can be kept with the test and transferred to each test item bank card.

Another helpful analysis for multiple choice items is to tally for each item the number of responses for each alternative in the high and low groups. For norm-referenced tests the instructor should retain alternatives that are answered at least once by both high and low groups and "distractors," which are chosen more often by low groups. The number of students in the class may preclude this.

Validity and reliability of a test are important. Determining these for each quiz is not realistic but for a course exam would be desirable. Validity is how close the test measured what it was intended to measure. To accomplish this the instructor and a colleague informally compare the objectives and blueprint with the test (*6*).

Reliability is how close the students' scores would be if they took the test on two separate occasions. Increasing the length of the test is one way of increasing reliability. If a test consisted of two items and the student missed one the first time he took it and none the second, the scores would be 50% and 100% respectively, showing poor reliability. Each item added to the test improves reliability. A reliability coefficient can be computed by either splitting the test into even and odd numbered items and applying the Spearman-Brown technique or using the Kuder-Richardson formula 21 (*6,7*).

There are other statistical analyses such as the mean, median, and standard deviation but they are not direct methods of improving a test.

In conclusion, constructing tests involves careful thought. When used appropriately, tests offer an easy and efficient means of generating feedback on teaching and learning outcomes.

Addendum

For information on TIPS and these recommended teaching modules, *Writing Instructional Objectives*, B. Daniell, principal developer, and *Developing Written Tests*, J. Worth, principal developer, contact Betty Bowling, Director, Teaching Improvement Project System, Center for Learning Resources,

College of Allied Health Professions, University of Kentucky, Lexington, KY 40536.

References

1. *Essentials of an Accredited Educational Program for the Nuclear Medicine Technologist*. Chicago: Committee on Allied Health Education and Accreditation, 1976, 4.

2. NMTCB Task Analysis of Nuclear Medicine Technology. *J Nucl Med Technol* 1979;7:102-07.

3. Position Description, Nuclear Medicine Technologist. *J Nucl Med Technol* 1979;7:178-81.

4. Boyd M, Conant R, McKeown J, Nagel M, eds. *Curriculum Guide for Nuclear Medicine Technologists*. New York: Society of Nuclear Medicine, 1982, 336.

5. *NMTCB Item Writer's Guide*. Stone Mountain, Georgia: Nuclear Medicine Technology Certification Board, 1981, 33.

6. Collins HW, Johansen JH, Johnson JA. Teacher-made tests. In: *Education measurement and evaluation: a worktext*. Glenview, Illinois: Scott Foresman and Co., 1969:53-108.

7. Gronlund NE. Simplified methods of treating test scores. In: *Constructing Achievement Tests*. Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1968:91-104.