The Art of Writing a Test
A Skill Never Learned

WILLIAM G. SPARKS III

Young teachers are not taught how to write a good test. In examining the curriculum of many teacher education institutions, instruction in test making appears to be a major deficiency. Therefore, how can teachers learn the art of test construction except through trial and error? During this learning process their students are often the victims of poorly designed tests. Test results are not an accurate measure of what the students have learned; therefore, the evaluation of student performance may not be valid.

According to Gronlund (1981), “Classroom tests play a central role in the evaluation of pupil learning. They provide relevant measures of many important learning outcomes and indirect evidence concerning others. The validity of the information they provide, however, depends on the care that goes into the planning and preparation of the tests” (p. 123).

What is a good test and how is it designed? A good test is a result of planning. A test must be systematically designed and reflective of the objectives of the unit. The likelihood of preparing a valid, reliable, and useful classroom test is greatly enhanced if a series of basic steps is followed. These steps should be systematic and sequential. They might include the following aims:

1. to determine the goals of the test
2. to develop the test specifications
3. to select appropriate test items
4. to prepare relevant test items
5. to assemble the test
6. to write clear directions
7. to learn how to administer the test
8. to interpret the results
9. to use the results for specific purposes

Determining Test Goals

In determining the goals of the test, clear purposes must exist. In evaluating the students, what information is desired? Usually, this information can be placed in one of the following four categories.

Placement Evaluation

These are pretests designed to measure (a) whether pupils possess the prerequisite skills needed to succeed in a unit or course or (b) to what extent pupils have already achieved the objectives of the planned instruction.

Formative Evaluation

These include tests that are given periodically during instruction to monitor pupil learning progress and to provide systematic feedback to pupils and the teacher. These tests might also be called learning tests, quizzes, unit tests, etc.

Diagnostic Evaluation

This type of test begins where the formative test stops. If pupils do not respond to the feedback provided by formative testing, a more comprehensive analysis of learning deficiencies is necessary. These tests are generally confined to a limited area of instruction, and the test items tend to have a relatively low level of difficulty.

Summative Evaluation

This test is given at the end of a course or unit of instruction, and the results are used primarily for assigning grades or for certifying pupil mastery of the instructional objectives. This may also be used for evaluating the effectiveness of instruction.

Control Statement

After the goals of the test have been determined, the next step is to integrate those goals into a brief “control

Dr. Sparks is the coordinator of student teaching at Emporia State University in Emporia, Kansas.
statement” that provides direction to the overall design of the test. Its purpose is to identify the overall theme of the unit, establishing a focus and direction as each question is written in order to stay on task. For example, if a unit of instruction dealt primarily with the idea that a war stimulates an economy through increased production, then that idea should be written down as the control statement of the test. Specific questions on the test can then reflect the theme of the control statement.

**Developing Test Specifications**

The only assurance we have that a classroom test is a valid measure of the instructional objectives and course content is to design a procedure for obtaining a representative sample of pupil performance in each area to be measured on a test. One method created for that purpose is a two-way chart referred to as a table of specifications. The chart relates the instructional objectives to the course content and, in the process, specifies the relative emphasis to be given to each type of learning outcome. The table of specifications should also be weighted according to the amount of instructional time actually spent in class teaching selected concepts. The more time spent in class teaching a concept, the greater the emphasis it is given on the test.

The building of a table of specifications includes (1) obtaining a list of instructional objectives, (2) outlining the course content, and (3) preparing the two-way chart. A brief description of each follows.

**Obtaining a List of Instructional Objectives**

A variety of learning outcomes can be measured through testing. These outcomes should be a reflection of the instructional objectives established for a unit. By establishing these objectives initially, a teacher can determine the extent of learning as well as the effectiveness of the instruction.

Instructional objectives are excellent tools for planning a unit. In selecting instructional objectives, learning outcomes as well as thinking skills need to be analyzed. Bloom’s (1956) taxonomy of educational objectives provides a good reference.

1. Knowledge
   1.1 terminology
   1.2 specific facts
   1.3 concepts and principles
   1.4 methods and procedures
2. Comprehension (Understanding)
   2.1 concepts and principles
   2.2 restate, relate, describe
   2.3 understanding material without relating it to other material
   2.4 summarize and generalize
3. Application
   3.1 use of information in different situations

3.2 concepts and principles
3.3 methods and procedures
3.4 problem-solving skills

4. Analysis
   4.1 comparing and proposing
   4.2 contrasting, categorizing, classifying
   4.3 comparisons

5. Synthesis
   5.1 problem solving
   5.2 predicting, modifying, imagining
   5.3 inventing and improving

6. Evaluation
   6.1 judging and deciding
   6.2 reaching conclusions
   6.3 critiquing and appraising

A glance at this list reveals the wide variety of learning outcomes that can be considered when one is developing a list of instructional objectives for a particular course. In teaching a unit, all areas would not necessarily be covered. The age level of the pupils, the subject matter area, and the philosophy of the school will help determine the learning outcomes to be stressed in a particular set of instructional objectives.

**Outlining Course Content**

Whereas the list of instructional objectives describes the types of performance the pupils are expected to demonstrate, the course content should indicate the major areas in which each type of performance is to be taught. Therefore the second step in preparing the test specifications is to make an outline of the course content. The outline could simply list major topics to be covered during the course or could include more detailed topics and subtopics.

**Preparing a Two-Way Chart**

The final step in building a table of specifications is to prepare the two-way chart that relates the instructional objectives to the course content. By creating this relationship, the nature of the test sample can be formulated. An example of a two-way chart for a summative test in a beginning racquetball class is described in Table 1.

**Selecting Appropriate Test Items**

After the outline of the test or the table of specifications has been created, the next step is the selection of appropriate test items. Item types generally are divided into two categories. The objective item is a precise, highly structured question that requires the students to supply a word or two, or to select the correct answer from among a limited number of alternatives. Examples of these types of questions include supply-type items, which are usually short-answer or completion questions, or selection-type questions, which generally include
true-false, matching, and multiple-choice questions. A second kind of item is the essay question. This is a single-item type that allows freedom of response from the students. The essay question can be broken down further into the extended response variety, which gives students almost complete freedom in responding to the question, and the restricted-response type, which limits the nature, length, or organization of the response. For example, an extended-response type question might ask students to describe the role of the federal government in administering block educational grants to the states, including specific examples and programs and supporting their examples with previous legislative actions. On the other hand, the restricted-response type of essay question might ask students to state two advantages and disadvantages of maintaining a uniform policy for the federal government to administer block grants to the states. Both the objective item and the essay question can provide valuable evidence of pupil achievement. Each has its unique advantages as well as limitations. Selection of the most appropriate test item should be clearly analyzed and reviewed, and questions should then be designed based upon the type of thinking and feedback desired.

Establishing Validity

In the preparation of questions, it is also important to establish content validity, that is, to determine that the test actually examines what it is supposed to examine. To do so, a specific two-step process should be followed. First, the objectives of the unit should be analyzed to ensure that the questions are relevant. Second, it is important that an appropriate number of test items reflect those objectives on the test. To establish content validity, the designer of the test should have experts in the field analyze and rate the questions. If experts are not available in the school, another staff member in the same department or another teacher on the faculty could serve as the expert.

Test Length

Length is usually determined at the time the set of specifications is built and depends on such factors as the purpose of testing, the types of test items used, the age of the pupils, and the level of reliability needed for effective test use. These factors need to be thoroughly analyzed and reviewed in developing test length.

Assembling the Test

Remmers and Gage (1965) suggest certain principles to follow in developing test questions. Some of these principles include

1. avoiding trivial, meaningless items
2. observing the rules of good grammar and punctuation
3. avoiding items that have answers that all experts will accept as correct
4. avoiding trick items or items so phrased that the correct answer depends upon a single obscure word
5. avoiding items that furnish answers to other items
6. avoiding items that contain irrelevant cues and are phrased in such a way that the correct answer can be easily guessed and does not require knowledge of the concept

Time and thought must be applied when writing specific questions. Tests are important means of the total
evaluation process. Therefore each question should be analyzed thoroughly before inclusion on a test.

It is the teacher's responsibility to minimize distractions so that the test accurately represents the extent of learning.

Directions and Scoring

After arranging the test items, the next consideration is to write clear, precise, easily understood directions. The directions should say what you want them to say.

In determining the method for scoring the test, it is important to think through the point value of each question as well as the overall value of the test. If the test ends a unit of study, it should have more value than a daily test. An instructor should ensure that the test is weighted according to the time spent on the unit. The test should blend into the overall scheme of evaluation for the course but reflect the importance of the unit.

Testing Atmosphere

Another consideration is the testing atmosphere, which should be quiet so the students can think without being distracted. It is the teacher's responsibility to control the environment so all students have an equal chance for success. If the testing environment is not controlled, the results of the test could reflect the conditions at the time of the test, and therefore the test would be invalid. A quiet room does not necessarily ensure ideal testing conditions, but it contributes to each student's having an equal opportunity to succeed.

Reducing Test Anxiety

At the start of the test, the teacher should read the directions aloud. This provides an opportunity for the teacher to structure any difficult or confusing section. It also allows students to ask questions without interrupting others' train of thought. This also can be a time to clear up misconceptions about how to answer a particular question. Other students might also be confused but hesitant to ask. Most important, this is a time for the teacher to help the students relax, thereby reducing some of the test anxiety most students feel and helping them to think clearly and draw upon stored knowledge to answer questions. Remember, the purpose of a test is to gauge what a student has learned. It is the teacher's responsibility to minimize distractions so that the test accurately represents the extent of learning.

Interpreting the Results

Once the test has been taken and the results have been tabulated, it is important to consider what the results mean and how they are going to be used. According to Safrit (1981), some of the more common uses of test results are:

1. to diagnose weaknesses
2. to classify according to ability
3. to exempt from aspects of the program
4. to predict future performance levels
5. to determine achievement levels
6. to indicate specific amounts of improvement
7. to motivate
8. to evaluate teaching effectiveness
9. to determine grades
10. to justify programs to administrators

There are many ways to use the results of a test. It is important to be sure the test is a valid one and is used for the correct purpose. A test is only a tool to assist in analyzing a student's progress, and it is only one indication of a student's ability. A test must be analyzed along with other criteria in identifying patterns or tendencies.

Summary

The design and administration of a test should be a planned process. The test should be weighted according to the importance of the unit. It should also blend in appropriately with the overall grading scheme. Unfortunately, new teachers are seldom taught how to design and administer a test. Historically, teachers have learned through trial and error. This should be corrected so that test design and administration is a required part of all curricula that prepare teacher educators. This will ensure that as students pass from one grade level to the next, their final evaluation will better represent their achievement and degree of knowledge.

REFERENCES


