DOCUMENT RESUME

ED 349 726 EC 301 471

AUTHOR Williams, Jane M.
TITLE Writing Quality Teacher-Made Tests: A Handbook for Teachers.
PUB DATE Sep 91
NOTE 48p.
AVAILABLE FROM Jane M. Williams, Wheaton High School, 12601 Dalewood Dr., Wheaton, MD 20906-4168 ($5).
PUB TYPE Guides - Classroom Use - Teaching Guides (For Teacher) (052)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Cognitive Processes; *Disabilities; Elementary Secondary Education; Item Analysis; *Teacher Made Tests; *Test Construction; Test Format; Test Items; *Thinking Skills; Verbs

ABSTRACT

This manual is designed to assist both special and regular educators with mastering the skills for developing quality teacher-made tests consistent with content-oriented instruction. The manual presents tips for constructing both objective and subjective, supply and select test questions--namely, short answer, essay, fill in the blanks or completion, true-false, matching, and multiple choice. The manual presents suggestions for using a table of specifications and item analysis to assure content validity of the tests and for developing multiple choice test questions which tap the higher order thinking skills of students. The manual also proposes solutions for eliminating formatting and construction errors and highlights pitfalls of each type of test question. A checklist for writing quality teacher-made tests is provided. Appendices include: (1) a list of verbs used in teacher-made tests for Bloom's Taxonomy of Education Objectives--Cognitive Domain; (2) a student guide to understanding words used in essay questions; and (3) a list of sample item stems for higher order cognitive questions. (Contains approximately 50 references.) (JDD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it
☐ Minor changes have been made to improve reproduction quality
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy

ED349726

WRITING QUALITY TEACHER-MADE TESTS:

A HANDBOOK FOR TEACHERS

Jane M. Williams, Ph.D.
Resource Teacher in Special Education
Wheaton High School
12601 Dalewood Drive
Wheaton, Maryland 20906-4168

September, 1991

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Jane M Williams
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

©1991 Jane M. Williams

EC 301471

BEST COPY AVAILABLE

Foreword

Writing quality teacher-made tests is a skill requiring mastery by both regular and special educators whether they utilize curriculum-based assessment (Brannon, Day, & Maley, 1978; ERIC Clearinghouse on Handicapped and Gifted Children, 1988; Howell & Morehead, 1987; Idol, Nevin, & Paolucci-Whitcomb, 1986; Marston & Magnusson, 1985; Tucker, 1985); criterion-referenced testing (Ediger, 1986; Fraenkel, 1980; Gage & Berliner, 1988; Gilman, 1988; Gronlund, 1977,1981; Salvia & Ysseldyke, 1981); or mastery learning techniques (Bloom, 1976; Guskey, 1985; Herman, 1984; Written Test Construction, 1985). However, the literature indicates that while most teachers rely on a student's performance on teacher-made tests to determine a student's grade (Barnes, 1985; Griswold, 1988; Kirby & Oescher, 1987; Marso, 1985; Office of Educational Research and Improvement, 1987; Stiggins & Bridgeford, 1985), many do not feel competent in developing valid test questions (Barnes, 1985; Griswold, 1988; Marso & Pigge, 1989; Stiggins & Bridgeford, 1985), and few believe they have received sufficient pre-service coursework in test construction (Barnes, 1985; Gullickson, 1986; Gullickson & Ellwein, 1985; Kirby & Oescher, 1987; Stiggins, 1985,1988; Stiggins & Bridgeford, 1985). In addition, when test questions from teacher-made tests are analyzed, most questions assess lower order cognitive skills such as knowledge or comprehension, rather than application, analysis, synthesis, or evaluation (Bloom, 1976; Carter, 1984; Kirby & Oescher, 1987; Marso & Pigge, 1989), and many questions contain grammatical, formatting, and construction errors (Kirby & Oescher, 1987; Marso & Pigge, 1989; Pigge & Marso, 1985).

For many special educators, like their regular education counterparts at the secondary level, designing first-rate teacher-made tests may be a problematic, yet necessary skill, since secondary special education students are increasingly receiving their content-oriented instruction within regular education classrooms or via a parallel curriculum approach from special educators (Carlson, 1985; Halpern & Benz, 1987; Schumaker & Deschler, 1988; Schumaker, Deshler, & Ellis, 1986; Seidenberg & Koenigsberg, 1990; McKenzie, 1991; Smith, 1987; Tindal, Parker, & Germann, 1990; U.S. Department of Education, 1990, 1991; Wagner, 1990; Wang & Birch, 1984), and these students must demonstrate proficiency with the curricular material in order to earn credits and graduate. Both of these program models are consistent with the regular education initiative ("Issues in the Delivery", 1987; Kaufman, 1988; "Regular Education", 1986; Reynolds, 1988; Schumaker & Deschler, 1988; Will, 1986) and mandates of the least restrictive environment requirements of Public Laws 94-142, 98-199, and 101-476.

Writing Quality Teacher-Made Tests is designed to assist both special and regular educators with mastering the skills for developing quality teacher-made tests consistent with content-oriented instruction. The manual presents tips for constructing both supply and select test questions--namely, short answer, essay, fill in the blanks or completion, true-false, matching and multiple choice. The handbook presents suggestions for using a table of specifications and item analysis to assure content validity of their tests and for developing multiple choice test questions which tap the higher order (application, analysis, synthesis, and evaluation) thinking skills of students. The manual also proposes solutions for eliminating formatting and construction errors and highlights pitfalls of each type of test question.

Table of Contents

Writing Quality Teacher-Made Tests: A Handbook for Teachers

Introduction

Writing teacher-made tests is a task required of all competent practitioners. Research indicates that classroom teachers spend approximately 15% of their time testing, administering a teacher-made test approximately once every two weeks (Gullickson, 1984). In fact, Stiggins (1988) found that teachers spend between 25% and 33% of their time measuring student achievement through evaluation techniques such as classroom testing, class participation, and observation. Moreover, Stiggins(1988) found that teachers make instructional decisions based on their assessment of student performance at the rate of once every two to three minutes. Most teachers advocate short, criterion-referenced tests (Gullickson, 1984), and Kirby and Oescher (1987) reported that teachers write 65.6% of their own test items, obtaining the remaining items from test guides, workbooks, and textbooks. Teachers view tests as important instructional tools worthy of the time and effort required for their use (Gullickson, 1984). Teachers believe that tests increase student effort, affect student self-concept, create competition, improve student interaction, and, in general, improve the learning environment (Gullickson, 1984).

While many teachers believe that tests should not serve as the sole basis for grades, teachers use student performance on their classroom tests as the primary measure of student learning (Gullickson, 1984) and as the major contributor to a student's grade (Barnes, 1985; Griswold, 1988; Kirby & Oescher, 1987; Marso, 1985; Office of Educational Research and Improvement, 1987; Stiggins & Bridgeford, 1985). Despite the critical role of teacher-made tests, many teachers report concerns regarding their pre-service preparation in test construction (Barnes, 1985; Griswold, 1988; Gullickson, 1986; Gullickson & Ellwein, 1985; Kirby & Oescher, 1987; Stiggins, 1985,1988; Stiggins & Bridgeford, 1985). According to Stiggins (1988), fewer than half the colleges and universities which belong to the American Assocation of Colleges of Teacher Education require training in student assessment as a condition of graduation, and most states require no training in assessment in order for teachers to be certified. Many teachers report concerns regarding the content validity, or the extent to which a test measures the topics taught (Nimmer, 1984), and the correlation between their tests and the curriculum.

Designing teacher-made tests relies on the premises established by curriculum based instruction and assessment, diagnostic testing, criterion-referenced testing, and mastery learning.

Curriculum based assessments can be defined as teacher constructed tests designed to measure directly students' skill achievements at specified grades; the assessments are criterion-referenced and their content reflects the curricula used in general education classrooms (Idol, Nevin, & Paolucci-Whitcomb, 1986).

Criterion-referenced tests measure an individual's ability with respect to some criterion or standard. Teacher-made criterion-referenced tests evaluate a student's achievement of a teacher's instructional objectives. They are not norm-referenced since their purpose is not to reveal differences among students, but to see what a particular student can do relative to a teacher's instructional objectives (Gage and Berliner, 1988).

Criterion-referenced testing requires a teacher to complete the following steps:

- state the instructional objectives
- design the criterion referenced mastery instrument (CRM)--i.e., the test
- teach to accomplish the objectives
- administer the CRM instrument
- score the CRM instrument
- evaluate the results: If student scores above a prescribed percentage (e.g., 70%), s/he has mastered the objectives. If a prescribed percentage of the students score above a certain level (e.g., 75% of the students score above 70%), the instruction has been effective. If either of these criterion has not been met, the teacher can decide if a change is needed in the objectives, in the instruction, or in the CRM instrument (Gilman, 1985).

Criterion-referenced testing can be viewed as synonymous with diagnostic testing, defined as "any test systematically designed to provide information about skills that students have or have not mastered" (Herman & Winters, 1985).

Mastery learning is an instructional model which calls for clarity about learning outcomes expected from instruction. The use of formative tests provides information for both students and teachers on a student's progress toward outcome attainment. Corrective instruction should be provided to students whose progress is unsatisfactory and "enriching" instruction should be provided for those students who master material (Bloom, 1976; Guskey, 1985; Ryan & Schmidt, 1979).

Developing a Table of Specifications

According to Gronlund (1981), effective classroom testing begins with a test plan that describes, in specific terms, the instructional objectives, content to be measured, and the relative emphasis to be given to each intended outome. Many authors suggest developing a table of specifications to accomplish this task. In developing a table of specifications, the following steps should be followed:

1. Determine what new material or content has been introduced in the learning unit--that is, list the new terms, facts, relations, or procedures which were explained, defined, illustrated, or presented.
2. Determine the student behaviors that should be paired with the new material--that is, is the student expected to identify the meaning of terms, make associations between old and new information, analyze or synthesize data.

In order to design a table of specifications a teacher must be familiar with Bloom's six cognitive processes: namely, knowledge, comprehension, application, analysis, synthesis, and evaluation (Bloom, 1976).

1. Knowledge is defined as recalling information as it was learned. For example,

_____. Who was the second President of the United States?
*1. John Adams
2. Thomas Jefferson
3. James Monroe
4. George Washington (Written Test Construction, 1985)

2. Comprehension is defined as reporting information in a way other than how it was learned in order to show it has been understood. In other words, comprehension can be demonstrated when one interprets information using one's own words or extrapolating from it new but related ideas and implications. For example,

_____ Which of the following coefficients of correlation has the highest predictive value?
1. -.30
2. -.94
3. .50
*4. .85 (Written Test Construction, 1985)

3. Application can be shown by using learned information to solve a problem. It is carrying knowledge of facts or methods learned in a specific context over to completely new contexts. For example,

_____ If lumber is priced at $0.50 bd. ft. and you need 60 linear feet of 1'x 4', how much will you pay?
1. $10
2. $15
*3. $30
4. $60 (Written Test Construction, 1985)

4. Analysis requires taking learned information apart--figuring out a subject matter's most elemental ideas and their interrelationships. For example,

_____ Warp is to Wood as Blister is to:
1. Metal
*2. Paint
3. Rattan
4. Tile (Written Test Construction, 1985)

5. Synthesis involves creating something new based on some criterion. For example,

_____ If you were preparing chocolate pudding using very high heat, no stirring, and unbeaten eggs, the result would be:
1. Curdling
*2. Lumpy texture
3. Smooth texture
4. Soft consistency (Written Test Construction, 1985)

6. Evaluation is judging the value of something based on one's own criteria or the well-understood criteria of another. For example,

_____ You are planning to ascend Mt. Hood starting from the main lodge at 2:00 A.M. and returning by 4:00 P.M. It is early spring and the weather calls for clear skies, highs in the mid-50s, and hard packed snow. Which shoes would best serve your needs?
1. Low top "tennis" shoes
2. High top "tennis" shoes
*3. Half shank boots
4. Full shank boots (Written Test Construction, 1985)

Guskey (1985) presented the following format for developing a table of specifications:

Table of Specifications

Knowledge of						
Terms	Facts	Rules and Principles	Processes and Procedures	Translations	Applications	Analyses and Syntheses

Guskey(1985) defined each of the categories as follows:

1. Knowledge of terms--terms are defined as new words or phrases; a student is required to define terms, recognize illustrations of them, determine when they are used correctly, and recognize synonyms. Knowledge can be recognition or recall. For example,

   What is the name of lines on a weather map? (Isobars; Gronlund, 1977)

2. Knowledge of facts--facts are defined as specific types of information students are expected to remember--e.g., names of persons, events, operations. For example,

   How long is the term of a United States Senator? (6 years)

3. Knowledge of rules and principles--These concern specific patterns or schema that are used to organize major ideas of a subject. They include interrelationships among a number of specifics. For example,

   If the temperature of a gas is held constant while the pressure applied to it is increased, what will happen to its volume? (Decrease; Gronlund, 1977)

4. Knowledge of processes and procedures--Students need to know particular steps in a process. For example,

   Name the steps which must be followed in order for a bill to become law.

5. Ability to make translations--involves transformation of a term, fact, rule, or process from one form to another. A student may be asked to express ideas in new way or to take phenomena or events in one form and represent them in an equivalent form. For example,

   Write an original sonnet.

6. Ability to make applications--using terms, facts, principles or procedures to solve problems in new or unfamiliar situations. For example,

   Solve the following mathematics equation:

7. Ability to analyze data--Analysis is breaking concepts into constituent parts and the detecting relationships among those parts. Distinguishing fact from opinion requires analysis.

8. Ability to synthesize data--Synthesis requires putting together elements or concepts in such a way as to develop a meaningful pattern or structure. Generating a conclusion and/or a supporting statement requires synthesis.

Nimmer (1984) suggested a somewhat simplified version of a table of specifications for teachers to use when developing tests. He suggested the following steps:

1. List all the content topics taught from the class lectures, activity guides, assignments, lab experiments, and textbook readings from the instructional unit.
2. Assign the relative emphasis desired for content topic--that is, estimate the appropriate percentage of the total instructional effort that was devoted to each content or topic.
3. Determine the total test length--e.g., 60 points.
4. Determine the number of test items or points per content topic desired; that is, multiply the total number of points by the relative emphasis of each content topic.

For example, for a test designed to test a student's knowledge about coniferous trees, he lists the following instructional objectives:

a. Define "coniferous tree."
b. Describe the structural parts of a coniferous tree and their functions.
c. Explain the reproductive cycle of a coniferous tree.
d. Name the coniferous trees native to this state.
e. Identify common coniferous trees by their cones and needles.
f. Identify common coniferous trees in photographs and slides.
g. Explain the economic uses of coniferous trees.

The table of specifications using his paradigm is constructed in the following manner:

Topic	Amount of Emphasis	Number of Points
Define "coniferous tree"	5%	3
Describe structural parts and functions	20%	12
Describe reproductive cycle	10%	6
Identify coniferous trees of Oklahoma	10%	6
Identify coniferous trees by cones and needles	30%	18
Identify coniferous trees in slides	10%	6
Identify economic uses of coniferous trees	15%	9
Total	100%	60

Whichever format a teacher decides to use to identify the instructional objectives of his or her instructional unit, the teacher's task is now to write test items for each content topic.

Types of Test Questions

Test questions have been categorized as objective and subjective by many authors including Gronlund (1981). For objective questions, there is only one correct answer, with no judgement entering into the correctness of the answer. Within the objective category, a subcategorization of select and supply has been developed. The types of objective select questions are multiple choice, true-false or alternative response, and matching. In an objective select question, the student selects the correct answer from among given alternatives. The types of objective supply questions are fill in the blanks or completion and short answer or short response.

Objective select type questions require recognition of material; objective supply type questions require recall of information by the student, often a more difficult cognitive task.

An essay question is the only form of a subjective test question. Essay questions can be further categorized as restricted response or extended response.

Guidelines for Developing Test Questions

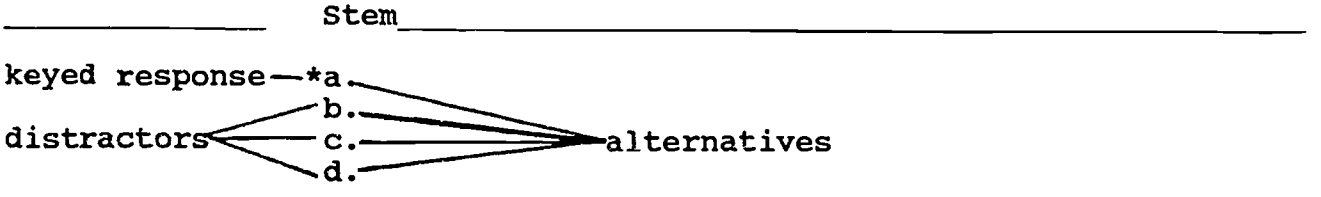
Objective Select Test Questions

Multiple Choice

A multiple choice test question has the following parts:

1. Stem--a direct question or an incomplete statement which precedes the answers and clearly states the topic or problem with which the item is concerned.
2. Alternatives--the answers. The keyed response is the correct answer; distractors are the incorrect answers.

Graphically, the format of a multiple choice item is as follows:



The general rules for constructing a multiple choice question are as follows:

1. A multiple choice test item should be utilized when the instructional objective requires the student to select from alternatives, or recognize the correct answer, not recall information. Multiple choice questions are best suited for measuring learning outcomes that require interpretation, understanding, or application of factual information. In other words, if the instructional objective indicates a student will be able to choose, select, or identify terms, facts, or relations, a multiple choice test item is the most appropriate type of test item.

2. The stem and the distractors should be contained on the same page, with the alternatives aligned vertically under the stem, not presented horizontally. For ease of scoring, the letter of the correct answer should be written on a line to the left of the stem. Students may circle the correct answer, if necessary, because of learning disabilities, but the practice of writing the correct answer (letter) to the left of the question will aid the teacher in scoring the test and prepare the student for Scantron sheets used in regular education classes as well as for standardized testing situations and conditions.

3. Directions should precede each set of multiple choice questions. For example, "Read each question. Choose the single best answer and write the letter of that answer in the blank to the left of the question."

4. Multiple choice test items should first be written as a direct question and changed to an incomplete statement only when greater conciseness is possible and the clarity of the question can be retained. For example, the direct question format for the following stem may be written as follows:

   _____ In which one of the following cities is the capital of California located?
   a. Los Angeles
   * b. Sacramento
   c. San Diego
   d. San Francisco

   Writing the stem as "The capital of California is located in..." achieves greater conciseness while retaining clarity.

5. The stem of a multiple choice item must be unambiguous and complete enough to present a problem. For example, the following stem is not meaningful by itself and does not present a clear question:

_____ South America
a. is a flat arid country
b. imports coffee from the United States
c. has a larger propulation than the United States
* d. was settled by colonists from Spain

A better phrasing of the stem would be as follows:

_____ Most of South America was settled by colonists from
a. England
b. France
c. Holland
* d. Spain

6. The stem of a multiple choice item should include as much of the item as possible, rather than repeating information in the alternatives. For example, in the following question, the alternatives repeat information that should be part of the stem:

_____ Why did Spanish colonists settle most of South America?
a. They were adventurous.
b. They wanted lower taxes.
c. They were seeking religious freedom.
*d. They were in search of wealth.

A better phrasing of the question would be the following:

_____ Spanish colonists settled most of South America because they were in search of:
a. adventure
b. lower taxes
c. religious freedom
*d. wealth

Placing as much of the wording as possible in the stem helps clarify the question, avoids unnecessary repetition of material, and reduces the time needed by students to read the alternatives.

7. The stem should not include clues to the correct answer. For example, grammatical clues such as a/an and singular and plural nouns and verbs should be avoided. In other words, articles, tenses, and syntax must be consistent between the stem and the alternatives; otherwise, students can select the correct answer using these as clues. For example, the use of "an" in the stem of the following question determines that alternative "a" is correct:

_____ Galileo could be best described as an
*a. astronomer
b. biologist
c. mathematician
d. physicist

Using "a/an" at the end of the stem prevents the article from serving as a clue to students.

In the following example, the use of the plural "presidents" and the plural verb "are" dictate the correct answer is "d":

_____ Which of the following former presidents of the United States are still living?
a. Dwight David Eisenhower
b. Lyndon Baines Johnson
c. John F. Kennedy
*d. Jimmy Carter, Gerald Ford, Richard Nixon, and Ronald Reagan

A better phrasing of the question would be as follows:

_____ All of the following former presidents of the United States are still living EXCEPT:
a. Jimmy Carter
*b. Dwight David Eisenhower
c. Gerald Ford
d. Richard Nixon
e. Ronald Reagan

Phrasing the question in the above manner achieves the same intent--that is, having the students recognize the four living former presidents of the United States, but its format does not contain a clue to the correct response.

Similarly, the student should not be able to determine the correct answer by using verbal associations, similarities in word meaning or in word resemblance. In the following example, the students can use the verbal association between the word "mystical" in the stem and "mysterious" in the alternatives to determine that "c" is the correct answer:

_____ The mystical tone established in this excerpt is best described as
a. humorous
b. ironic
*c. mysterious
d. sarcastic

8. The stem should be phrased in positive, rather than negative terms, unless there is a valid instructional reason for using a negative in the stem. For example, in the following question, the task of the student should be to identify the formula for finding the area of an ellipse, if that is the learning outcome desired by the teacher, not indirectly recognizing that the formula for calculating the area of a square, rectangle, and triangle is "length times width." Identifying answers that do not apply does not guarantee student knowledge or comprehension of information being requested.

_____ The formula "a = l x w" is not applicable in finding the area of a/an:

*a. ellipse
b. rectangle
c. square
d. triangle

There may be occasions for using a negative in a stem. For example, the following question asks for valid information:

_____ Which one of the following is not a safe driving practice on icy roads?

a. accelerating slowly
b. holding the wheel firmly
*c. jamming on the brakes
d. slowing down gradually

The question should be rewritten into the following format, however, to assure that the students read the question correctly:

_____ All of the following are safe driving practices on icy roads EXCEPT:

a. accelerating slowly
b. holding the wheel firmly
c. jamming on the brakes
d. slowing down gradually

With the EXCEPT capitalized, underlined, and placed at the end of the stem, students are less likely to overlook the negative format.

Carter(1986) also suggested avoiding the use of negative versus positive alternatives as used in the following example:

_____ The setting of this story, a stormy night,

a. is unimportant because you can read on any kind of night
b. is unimportant because the next door neighbors are at home
*c. is unimportant because Debbie does not fear storms
d. is important because it adds to the things that are frightening to Debbie

In Carter's (1986) study, 79.81% of respondents choose alternative "d," though alternative "c" is the correct answer. Because of the deviation in format of alternative "d," the students were convinced it was the correct answer.

9. All distractors must be plausible, familiar to the students, and related to the content studied. In addition, only one alternative must be correct. For example, in the following item, both alternatives "b" and "c" are correct.

_____ The state of Michigan borders on
a. Indiana
*b. Illinois
*c. Lake Huron
d. Lake Ontario

To improve this question, both the stem and alternatives should be rewritten to pose a specific problem and make the alternatives homogeneous. For example,

_____ Which of the following states borders Michigan?
a. Illinois
*b. Indiana
c. Pennsylvania
d. Wisconsin

10. All the alternatives should be of similar length and complexity. For example, in the following question, the length and detail provided in alternative "c" provides a clue to the students as to its correctness:

_____ The cell membrane performs which of the following functions for the cell:
a. controls cell reproduction
b. produces chlorophyll
*c. selectively controls the passage of some materials into and out of the cell
d. stores food and gases

Research has indicated that students generally select the alternative that is the longest and most complex, and teachers most often write more information in the keyed response than in the distractors. The length and complexity of alternative "c" in the preceding example may be considered by the student in selecting the correct response.

11. Avoid using "all of the above" or "none of the above" as alternatives. Students can guess if either is the correct response if they know two alternatives are correct or incorrrect. Similarly, avoid complex alternatives such as "a and b, but not c." This alternative requires a cognitive skill which detracts from assessing if the student can select the correct answer.

12. Do not use absolute terms such as "all," "never," "only," and "none" in the alternatives. These qualifiers will most often make the alternative incorrect, and hence the student will eliminate the alternative from consideration.

13. Do not use verbatim or paraphrased items directly from the textbook. This practice encourages memorization rather than testing for comprehension of the content.

14. Be sure each item is independent of other items--that is, do not make answering an item correctly dependent on correctly answering a preceding item. Interlocking items are not fair to students and their use will not provide an accurate representation of student knowledge.

15. Write distractors in such a way as to gain diagnostic information from incorrect responses when an item analysis is performed (See page 28). This diagnostic information can provide clues as to the material students are learning as well as improvements needed in the instruction provided.

16. Randomly assign the keyed response to each of the 4 (or 5) alternatives. Research has shown that alternative "c" is most often selected by students as the correct answer and utilized as the keyed response by teachers. Vary the placement of the correct answer by using a book. For each test item, Gronlund (1981) suggests opening the book to an arbitrary position, noting the number on the right hand page, and placing the correct answer for that test item as follows:

If page number ends in	Place correct answer
1	1st
3	2nd
5	3rd
7	4th
9	5th

Gronlund (1981) also suggests placing all verbal alternatives in alphabetical order and placing all numerical answers in numerical order.

17. Write multiple choice items to test higher order thinking skills when the learning outcomes dictate. Pigge and Marso (1988) suggest posing hypothetical situations or problems to increase the cognitive level of questioning, presenting questions with novel or new examples, and preparing questions which require best judgment selections based upon predictions, applications, or principles and laws. Pigge and Marso (1988) suggest using the stems "What would happen if...?" and "How can this be corrected?" to assess comprehension and/or application.

Gronlund (1981) suggests using the following stems for assessing <u>application</u> learning outcomes:

- What method would be best for ...?
- What steps should be followed to construct ....?
- Which of these indicates application of ...?
- Which of these solutions is correct for ...?

Ability to interpret cause and effect relationships:

_____ Bread will not become moldy as rapidly if placed in the refrigerator because
a. cooling prevents the bread from drying out as quickly
*b. cooling retards the growth of fungi
c. darkness retards the growth of mold
d. mold requires both heat and light for best growth

Ability to apply facts and principles:

Directions:

In each of the following sentences circle the word that makes the sentence correct.

1. This is the boy who whom that asked the question.

2. This is the dog who whom that he asked about.

Ability to justify methods and procedures:

_____ Why is lighting necessary in a balanced aquarium?
a. fish need light to see their food
b. fish take in oxygen in the dark
*c. plants expel carbon dioxide in the dark
d. plants grow too rapidly in the dark

Analysis:

_____ What part of speech is the underlined word in the following sentence?

John eagerly played ball.
a. adjective
*b. adverb
c. noun
d. verb

Evaluation:

_____ Which of the sketches drawn on the chalkboard portrays the best informal balance?
a. sketch 1
b. sketch 2
c. sketch 5
d. sketch 6

Appendix A contains a list of verbs to use for writing test questions for each of Bloom's cognitive levels.

Objective Select Test Questions

True-False or Alternative Response

A true-false or alternative response question is a declarative statement which the student must mark as true or false, right or wrong, correct or incorrect, yes or no, agree or disagree, or fact or opinion. This type of test question measures a student's ability to identify the correctness of a statement of fact, of a definition, or of a principle. It is best used when the learning outcome requires knowledge or comprehension of factual information.

Graphically, the format for a true-false or alternative response question is as follows:

T F Declarative Statement

The general rules for constructing a true-false question are as follows:

1. Directions should be included for each set of true-false or alternative response questions. For example,

   Directions: Read each statement carefully and determine if the statement is true or false. If the statement is true, circle the "T" in front of the statement. If the statement is false, circle the "F" in front of the statement.

   Students should not be required to write responses instead of circling the correct response. This practice is more time consuming than circling a response, and circling a response eliminates possible difficulties a teacher may encounter deciphering the student's handwriting.

2. The statement must include only one significant idea which is worded clearly and precisely and is either true or false without qualification. Ambiguous, broad general statements should not be used. For example, the following question is poor because it is a broad generalization:

   T F The president of the United States is elected.

   Specificity, rather than a qualifier such as "usually" would improve the question. For example,

   T F Election to the presidency of the United States requires a majority vote by the electoral college.

3. Avoid using specific determiners which may make the statement true or false. For example, the following words usually make a statement true:

as a rule	most
could	often
customarily	several
few	some
generally	sometimes
may	usually
maybe	

The following words usually make a statement false:

absolutely	fully
all	never
alone	none
always	nothing
completely	only
entirely	solely
exactly	totally
exclusively	

4. Do not include two ideas within one statement, either of which may be true or false. For example, in the following question, either proposition could be true or false.

T F A worm cannot see because it has simple eyes.

5. Avoid long, compound, and complex sentences which may assess reading comprehension, or trivial information, not knowledge or comprehension.

6. Do not use single or double negatives in statements. If a negative is used, <u>underline</u> it or put it in italics.

7. Do not use statements verbatim from a student's text and add the word "not." This practice encourages poor study habits and may lead to distrust among students.

8. Do not use trick questions. For example, the answer to the following question is false because the votes of the electoral college, not the vote of the people, determine the election of the president of the United States.

   T F George Bush was elected to the presidency of the United States by popular vote and by the electoral college.

   If the teacher wishes to test the student's knowledge of this procedure, the question should be phrased in either of the following formats:

   T F Election to the presidency of the United States requires a majority vote by the people of the United States and by the electoral college.

   T F Election to the presidency of the United States requires a majority vote by the electoral college.

9. Do not test trivial bits of information. For example, in the following question, the year is incorrect:

   T F Japan attacked Pearl Harbor on December 7, 1942.

   If recognition of the correct year is the intended outcome, either a multiple choice or a matching question may be more appropriate. If recall of the year is the intended outcome, a fill in the blanks or completion question may be preferable to making the question false because of a tiny bit of information.

10. Avoid opinion statements unless the source is identified.

11. Make true and false statements approximately the same length and complexity. True statements have a tendency to be longer than false statements because of the specificity required to meet the criteria of absolute truth. Gronlund (1981) suggests, if necessary, lengthening false statements.

12. Include an approximately equal number of true and false statements in the test.

13. Randomly determine the placement of true and false questions in order to avoid an answer pattern, detectable by the students--e.g., T T F F T T .... Flip a coin to randomize the placement of questions.

14. Remember that the student has a 50/50 chance of guessing the correct answer. Hence, test very discreet bits of information and write concise and precise statements.

15. Write true-false questions only when the learning outcome dictates demonstration of knowledge or comprehension of factual information.

Objective Select Test Questions

Matching

Matching test questions involve two parallel columns with each word, date, or symbol in one column being matched to a word, phrase, or sentence in the other column. The two columns are called premises and responses. The premises are the items in the column for which a match is sought; premises are presented on the left, numbered consecutively with the test. Responses are the items from which a selection is made; responses are on the right, preceded by a letter. A blank should be provided to the left of of each premise on which the letter of the correct response is recorded by the student.

Matching test questions are best utilized to assess a student's knowledge or comprehension of terms or facts. Marso and Pigge (1989) suggest using matching test questions to test comprehension of classifications, original examples, and predicted consequences.

Graphically, a matching test question is presented as follows:

	Column A--Premises		Column B--Responses
_____	1.	_____	a. _____
_____	2.	_____	b. _____
_____	3.	_____	c. _____

Suggestions for writing quality matching test questions are as follows:

1. Present directions for each set of matching questions. The directions should contain the following 3 parts:

   a. basis for the match
   b. directions for responding to the premises
   c. directions regarding the use of responses

   For example,

   In Column A below are descriptions of some late 19th century American painters (basis for match). For each description, choose the name of the painter being described from Column B and write the letter identifying the painter on the line preceding the correct description (directions for responding). Each name in Column B may be used once, more than once, or not at all (directions regarding the use of responses).

2. Label or title each column. For example, in the exercise described above, Column A could be titled "Description of Painter" and Column B could be labeled "Name of Painter." Titling premises and responses aids the student with understanding the task required.

3. Make the premises longer than the responses. This practice assists the student with completing the task. S/he can scan the responses, arriving at the correct answer quickly. Making the responses longer than the premises slows the process for each student, since s/he is taught to read the responses each time s/he is responding to each premise.

4. Make the premises and responses homogeneous--for example, 19th century American painters, not 19th century painters.

5. Premises should be sufficiently long to be clear and should contain enough information for the student to construct an interrogative question from the material. For example, "When was the 14th amendment ratified?" In responding to the question, the student will be able to put the words into a declarative sentence such as "The 14th amendment was ratified in 1868." The student will then be able to check for accuracy by asking him/herself the question, "Does the sentence make sense and is it correct?" If the answer is "yes" to both parts of the question, the match is probably correct; if the answer is "no" to either part of the question, the match is probably incorrect.

6. Provide no more than 10 premises. Otherwise, the matching exercise becomes one of reading comprehension and stamina, not knowledge.

7. Provide an unequal number of premises and responses. Presenting more premises than responses allows the responses to be used more than once, eliminates guessing, and prevents pupils from matching the final pair of items based on the process of elimination.

8. Provide several plausible responses for each premise. If a response is inappropriate, students will eliminate it from consideration and thereby increase their chances of guessing the correct answer.

9. Arrange the responses in a logical order. For example, arrange the responses in alphabetical, chronological, or numerical order. Placing responses in a logical order will aid the student in locating the correct response quickly.

In the following example, the directions present the task clearly, the columns are labeled, the premises are homogeneous and complete, more premises than responses are provided, the premises are longer than the responses, and the responses are arranged chronologically.

Directions:

On the line to the left of each historical event in Column A, write the letter from Column B which identifies the time period during which the event occurred. Each date in Column B may be used once, more than once, or not at all.

	Historical Event	Time Period
B	1. Boston Tea Party	A. 1765-1769
A	2. Repeal of the Stamp Act	B. 1770-1774
E	3. Enactment of the Northwest Ordinance	C. 1775-1779
C	4. Battle of Lexington	D. 1780-1784
A	5. Enactment of Townshend Acts	E. 1785-1789
B	6. First Continental Congress	
E	7. United States Constitution drawn up	

10. Place all premises and responses for one matching exercise on the same page.

Objective Supply Test Questions

Fill in the Blanks or Completion

A fill in the blanks or completion question is to be utilized when recall of factual material is being measured. One word responses such as names, dates, and places are expected. A direct question or an incomplete statement can be used. The distinction between the fill in the blanks and completion question involves the placement of the blank when the statement is phrased as a declarative statement. In a fill in the blanks question, the blank is located within the question; in a completion question, the blank is at the end of the question. Several authors indicated an interrogative question followed by a question mark is preferable to a declarative statement with a blank at the end because of the specificity which can be obtained via a question format.

For example, the following question can be posed in the form of an interrogative question or a declarative statement:

_____ 1. What is the capital of Ohio?

OR

1. What is the capital of Ohio? _____

OR

1. The capital of Ohio is _____.

Several authors suggested writing the question in the form of an interrogative is more precise than writing the question in a declarative or fill in the blanks format. Several authors suggested placing the blank in front of the question on the left; other authors suggested the blank be placed at the end of the question aligned with the right margin. Both formats aid scoring. Gronlund (1981) suggested placing the blank on the left facilitates the use of a strip scoring key. The benefit to placing the blank on the right at the end of the line, however, is that that the student does not have to return to the beginning of the question to respond. For example,

1. What are warmblooded animals that are born alive and suckle their young called? (mammals)

Regardless of the format selected, the following guidelines are suggested for writing fill in the blanks or completion questions:

1. Provide directions for each set of fill in the blanks or completion questions. For example,

   Directions:

   Read each question. Place the single word answer to the question in the blank to the right of the question.

2. Provide enough information in the question to enable the student to determine the information being requested. For example, in the following example, the question does not contain sufficient information for the student to determine the specific answer being requested:

   John Glenn made his first orbital flight around the earth in _____.

   A better phrasing of the question would be as follows:

   John Glenn made his first orbital flight around the earth in the year _____.

   Phrasing this example in the form of an interrogative question, however, achieves greater specificity and conciseness than its companion declarative form:

   In what year did John Glenn make his first orbital flight around the earth? (1962)

3. Make each blank of adequate and equal length, placed preferably at the end of the question.

4. Omit only key words. Do not test for trivia.

5. Have only one blank per statement or question.

6. Request only one word responses. A phrase or a sentence is appropriate for a short response question, not a completion or a fill in the blanks question.

7. When the response is to be expressed in numerical units, specify the desired units. For example,

   If oranges weigh 5 2/3 oz. each, how much will a dozen oranges weigh? Answer (4) lbs. (4) oz.

   If remainders are involved, indicate the degree of precision expected in the answers--for example, carried out to 2 decimal places, rounded to the nearest tenth, etc.

8. Do not provide a list of words from which the students may select an answer. Providing a word list constitutes a matching exercise and changes the cognitive skill required from recall to recognition. If recognition is desired, write the question in a multiple choice or matching format.

9. Do not use questions verbatim from the student's textbook or classroom instruction. Use one's own wording.

10. Assure there is only one correct answer, but anticipate possible synonyms or acceptable variants of the desired response.

Objective Supply Test Questions

Short Answer or Short Response

Short answer or short response test questions require recall of specific information which can be relayed in a few words, a phrase, or a sentence or two, but not a paragraph. They should be phased as concise, simple interrogative questions. Short response questions should query understanding and interpretation; questions requiring only names, dates, places, and events should be designed as fill in the blanks or completion questions.

For example,

Why did Tom Sawyer become angry with the raft after the storm?

This question can be answered in a few words.

There are also several words which signal that a short answer response is desired. These words are as follows:

1. Name
2. List
3. Identify
4. Give
5. Mention
6. State
7. Give the principle of

The guidelines for constructing short answer/short response questions with the above words are as follows:

1. For the words name, list, mention, and give, the student is being asked to list the information requested. No sentences are expected. For example,

   List the 3 things the Dawes Act gave the Indians.
   1. the right to own property
   2. schools where they could learn farming and obtain an education
   3. the promise that they would become full citizens of the United States

   The word "list" could have been "name," "identify," "give," or "mention."

2. For the word "state," the question is asking the student to describe, define, or point out the requested information. No discussion is desired. A single sentence or a brief list is to be judged as adequate.

   For example,

   State the event which started World War I.

   Answer: World War I began when Archduke Francis Ferdinand of Bosnia and his wife Duchess Sophie were assassinated by a man from Serbia in Sarajevo, Yugoslavia.

3. For the words "give the principle of" the students are expected to provide the law, rule, or principle being requested. The student may add an example to support his/her response.

   For example,

   Give the principle of flotation.

   Answer: Materials lighter than water float; materials heavier than water sink. For example, a tennis ball floats; a rock sinks.

Subjective Test Questions

Essay

Essay questions assess the student's ability to recall, select, organize, and integrate ideas and record them in written form. Essay questions are the preferred format for tapping the higher order cogntive skills of application, analysis, synthesis, and evaluation. Essay questions should not be used to measure factual data and should not elicit single word responses or a list of items. A minimum of two paragraphs should be expected.

Essay questions vary along a continuum of freedom to respond or of restrictiveness, and hence have been categorized as restricted response or extended response (Gronlund, 1981). The more restricted an essay question, the more objectivity enters into its scoring and the higher its reliability; conversely, the less restricted a question, or the more freedom allowed in a student's response, the more subjectivity is involved in evaluating a student's response and the less reliability can be expected.

Gronlund (1981) offers the following example of an essay question which varies along this continuum of restrictiveness or freedom to respond:

> Highly restricted: Outline the events which, according to the text, led to the Depression of the thirties.
>
> Somewhat restricted: What events led to the Depression of the thirties? What part did each event play in causing the Depression?
>
> Some freedom: Discuss the cause and effect of the Depression of the thirties. Include in your answer documented evidence of your position.
>
> A great deal of freedom: Write 4 or 5 pages about the Depression of the thirties.

In constructing and scoring essay questions, the following guidelines should be followed:

1. Write clear, unambiguous questions. Do not ask broad, general questions. For example, the following question is too broad to be meaningful:

   Discuss mathematics.

2. Assure only higher order cognitive skills are being tested; do not ask questions which are primarily asking for factual data. For example, the following is a completion, not an essay, question:

   What is the formula for finding the area of a parallelogram?

3. Assure each essay question corresponds to the learning outcome(s) specified in the table of specifications.

4. In the directions, specify the following information:

    a. the content desired in the student's response
    b. the length of the desired response
    c. the amount of time allowed for responding
    d. the sources to be used
    e. the style of the response--e.g., discuss, compare and contrast, interpret, evaluate. (See Appendix B for a list of words commonly used in essay questions and the expected response.)
    f. a reminder that each part of the question must be answered
    g. the total point value of the question
    h. the procedure for handling unrelated information

    For example, the following question contains a clear statement of the problem and a precise description of the desired response:

    Directions:

    Read the following question carefully. Respond to and label each part of the question. Confine your response to the space provided. Use your text and class notes for supporting documentation and use the entire class period to respond. The question is worth 15 points. Points will be deducted for including irrelevant data in your response.

    Question:

    Should governments maintain social welfare programs?

    Answer "yes " or "no" and then defend your position in 1-3 pages. Include in your response a discussion of at least 3 alternative types of programs and describe the effects each type is likely to have on the recipients of the program.

5. Ask several brief, restricted response questions rather than one or two questions with a high degree of freedom. Use essay questions to supplement objective items, and do not permit essay questions to outweigh the student's performance on the objective sections of the test. This practice not only increases the reliability of the test, but also provides a more accurate assessment of a student's competencies, particularly for the student whose underdeveloped writing skills may have decreased his/her overall test performance.

6. Do not provide optional questions; all students should answer the same question(s). Otherwise, students are, in effect, taking different tests and a common basis for evaluating their achievement does not exist.

7. Construct a model answer (in outline form if desired), and specify the scoring criteria prior to administering the question. Decide whether analytic or holistic scoring procedures will be used.

   Analytic scoring identifies the essential points of the correct response and scores each student's response accordingly. For analytic scoring, the teacher must develop a scoring key using the following guidelines:

   a. specify each major and minor point and determine an associated point value
   b. determine the amount of credit to allot to other characteristics of the answer including the following:

      1. organization
      2. relevance of ideas
      3. building a logical argument
      4. citing of appropriate examples
      5. placing events in proper sequence
      6. comprehensiveness
      7. sentence structure
      8. spelling
      9. punctuation
      10. handwriting
      11. neatness

   c. determine the procedure for handling irrelevant information contained in a student's response--e.g., applying a penalty

   Holistic scoring, or global quality scoring, is based upon the teacher's general impression of the overall adequacy and quality of the student's response. The student's answer is scored as a whole rather than based on its component parts. Gronlund (1981) suggested using a rating procedure for holistic scoring which involves assigning each paper to one of a number of categories based upon its overall quality. If, for example, 10 points are to be awarded for the question, the paper should be assigned to one of 11 categories, ranging in value from 0 to 10 points.

   Gronlund (1981) and other authors suggest analytic scoring may be best suited for restricted response essay questions, while holistic scoring may be more appropriate for extended response questions. Analytic scoring may prove too complex, time consuming, and cumbersome for extended response questions which involve a high degree of freedom to respond.

8. When grading papers, Gronlund (1981) suggests the following guidelines:

   a. Read a small sample of responses--e.g., 5 or 6, to gain a general impression of the quality of responses that may be expected.
   b. Read the test papers anonymously--e.g., use a number system or put the student's names on the back of the final test page.
   c. Score all the responses to one question before preceding to score the next question. This practice enables the scorer to concentrate on one item at a time and enhances the consistency of scoring.
   d. Reorder the papers in a random fashion after scoring each question in order that a given student's paper is not consistently in the same relative position. This process counteracts a rater's stiff initial standards and fatigue.
   e. Reevaluate the first 5 or 6 papers after scoring all the test papers to assure the scoring criteria has remained constant.
   f. Have 2 independent raters if possible.

   The procedures listed above are designed to counteract the following empirical results:

   a. Test scores are affected by the quality of the papers scored previously. Research has shown the following:

      1. Essays of average quality are rated more highly when preceded by poor quality essays than when preceded by good quality essays.
      2. On a given test with 2 or more essay items, if the response to 1 item is scored high, there will be a tendency to score the response to the next item high as well.

   b. Teachers become more lenient as they progress through a certain set of responses. Research has shown that tests scored first are scored more critically than later exams.

   c. Teachers do not ignore errors in language mechanics--i.e., errors in spelling, punctuation, and capitalization, and concentrate on content.

   d. There is a tendency to give a higher score to a longer response than to a shorter response even when the shorter response includes the essential content.

   e. Research has shown the presence of a "halo effect"--i.e., a tendency to give high scores to students who are known to be "good" and vice versa.

Item Analysis

After administering the test, it is important to conduct an item analysis to determine if the question is well-constructed. Measures for assessing the difficulty of an item and the item's sensitivity to instruction follow:

1. Item Difficulty

Item difficulty is the proportion of students who answer a test question correctly. A difficulty index ranging from 0.00 to 1.00 can be computed using the following formula:

$$\text{Item difficulty} = \frac{\text{\# of students responding correctly}}{\text{\# of students in class}}$$

A difficulty index of 0.00 indicates that no students responded correctly; a difficulty index of 1.00 indicates all students responded correctly.

For example, for the following multiple choice question, the response profile may be the following:

Question:

_____ If the odds in favor of an event occurring are 6 to 1, the probability of the event occurring is
a. 1/7
b. 1/6
c. 1/13
*d. 6/7

Response Profile:

a	b	c	d	Alternative
4	6	0	15	Number of student responses

Given that alternative "d" is the keyed response, the item difficulty may be computed as follows:

$$\frac{15}{25} = .60$$

Using the following guidelines, the item would be rated as moderate in difficulty:

Low difficulty = .70 or greater
Moderate difficulty = .30 - .70
High difficulty = .30 or less

Gronlund (1981) suggests an item difficulty should not be less than .30; otherwise, either the item is faulty or the instruction needs improvement.

The fact that no students chose alternative "d" in the above example provides important information to the teacher. The alternative should be rewritten to have it serve as a more plausible distractor. Similarly, based on item analyses, items which are frequently answered incorrectly should be reviewed. The item may be ambiguous, keyed incorrectly, or be identifying content topics which have not be taught, or learned, as thoroughly as intended (Gronlund, 1981; Nimmer, 1984).

2. Sensitivity to Instruction Index

The Sensitivity to Instruction Index requires both pre- and post-testing and measures both the effectiveness of instruction and the appropriateness of a given item for assessing the instruction. The formula for computing the Sensitivity to Instruction Index is as follows:

$$S = \frac{RA}{T} - \frac{RB}{T}$$

where

S = sensitivity to instruction

RA = # of pupils who answered the item correctly after instruction

RB = # of pupils who answered the item correctly before instruction

T = total # of pupils who tried the item both times

The index ranges from -1.00 to 1.00 with the ideal range falling between .70 - 1.00. The following examples of various situations involving the Sensitivity to Instruction Index illustrate its use:

$$\text{Item 1} = S = \frac{0}{6} - \frac{6}{6} = -1.00$$

This item is either defective or too easy.

Item 2 = S = $\frac{6-6}{6}$ = .00

This item is too easy to measure the effects of instruction.

Item 3 = S = $\frac{0}{6} - \frac{0}{6}$ = .00

This item is either too difficult to measure the effects of instruction or the instruction was inappropriate.

Item 4 = S = $\frac{4}{6} - \frac{2}{6}$ = .50

This item is effective since some pupils responded correctly before instruction, but more pupils responded correctly after instrcution.

Item 5 = S = $\frac{6-0}{6}$ = 1.00

This is an ideal item since all students answered correctly after instruction, but none did so before the instruction.

Checklist for Writing Quality Teacher-Made Tests

	Yes	No
1. Does the table of specifications contain both the content and the instructional objectives?	_____	_____
2. Does the table of specifications specify the relative emphasis for each content area and instructional objective?	_____	_____
3. Does the format of each item correspond to the specified learning outcome?	_____	_____
4. Are there at least 10 objective test items for each learning outcome?	_____	_____
5. Are directions provided for each section of the test?	_____	_____
6. Are the directions clear and concise and at a reading level commensurate with the students' ability?	_____	_____
7. Do the directions specify the task, the procedure for answering, and the time allowed for responding?	_____	_____
8. Are sample items provided for each set of directions?	_____	_____
9. Does each item present a clear and definite task to be performed?	_____	_____
10. Is each item free from grammatical clues, specific determiners, and verbal associations?	_____	_____
11. Is each item independent from all others?	_____	_____
12. Does each item contain vocabulary at the appropriate reading level?	_____	_____
13. Does each objective item have only one correct answer?	_____	_____
14. Are the alternatives for multiple choice test items aligned and on the same page as the stem?	_____	_____
15. Is each blank of equal length and on the right (or left) aligned with the margin?	_____	_____

Checklist for Writing Quality Teacher Made Tests--Continued

	Yes	No
16. Is there only one blank per completion item?	_____	_____
17. Is adequate space provided for short response and essay questions?	_____	_____
18. Are test items of the same type grouped together with the test?	_____	_____
19. Are the test items arranged from easy to more difficult within sections of the test and within the test as a whole?	_____	_____
20. Have you included "spiraling" items-- i.e., items which build upon material previously taught?	_____	_____
21. Is the test clear and free of spelling and typographical errors?	_____	_____
22. Are the items numbered consecutively?	_____	_____
23. Are the margins adequate?	_____	_____
24. Is the test as a whole representative of the content taught?	_____	_____
25. Is the test long enough to sample the content adequately, but not so long that it is a test of speed, not power?	_____	_____
26. Have you taught test-taking skills?	_____	_____
27. Have you informed the students of the test content and format to assure appropriate studying?	_____	_____
28. Have you tested frequently to decrease test anxiety?	_____	_____
29. Have you tested at the beginning of the class period so tests can be graded, returned, and reviewed?	_____	_____
30. Have you provided a calm testing environment?	_____	_____
31. Have you continuously revised your test items to assure the items parallel course content?	_____	_____

Appendix A

Verbs used in Teacher-Made Tests for Bloom's Taxonomy of Educational Objectives--Cognitive Domain

Cognitive Level	Verbs	
Knowledge	choose	locate
	complete	match
	define	name
	describe	recall
	identify	recognize
	indicate	select
	label	state
	list	
Comprehension	add	express
	balance	factor
	calculate	interpret
	classify	measure
	compare the importance of	multiply
	compute	put in order
	convert	subtract
	divide	suggest
	expand	summarize
	explain	trace
Application	apply	determine
	choose	develop
	compare	differentiate
	compute	discuss
	construct	draw
	defend	explain
	demonstrate	experiment
	design	express in a discussion
	find	predict
	make	prepare
	organize	prove
	outline	relate
	paint	select
	participate	sketch
	perform	solve
	plan	test

Verbs Used in Teacher-Made Tests for Bloom's Taxonomy of Educational Objectives--Cognitive Domain

Cognitive Level	Verbs	
Analysis	analyze	dissect
	categorize	differentiate
	compare	distinguish
	compare/contrast	
	conclude	draw conclusions
	critique	explain
	debate	form generalizations
	describe	identify
	detect	interpret
	deduce	organize
	determine	relate
	diagram	separate
Synthesis	add to	infer
	assemble	imagine
	combine	invent
	compose	organize
	conduct	predict
	construct	produce
	create	recreate
	describe	suppose
	design	what if
	develop	write (an original composition
	formulate	
	hypothesize	
Evaluation	appraise	examine
	compare and contrast	evaluate
	criticize	judge
	critique	recommend
	decide	solve
	debate	weigh
	determine	

(Bloom, 1976)

Appendix B

Understanding Words Used in Essay Questions

Listed below are the 12 words or phrases most often used in essay questions with the expected responses.

	Word/Phrase	Expected Response
1.	Outline	Arrange information in outline form.
2.	Trace	Give events in the order in which they occurred. Present major points and cause and effect relationships.
3.	Summarize; give the significance of	Present both major points and generally accepted conclusions or outcomes. Apply principles and concepts stressed in class.
4.	Give examples of; illustrate	Give instances of, or sample occurrences; usually a list is accepted as part of the answer.
5.	Put in your own words	Translate technical, literary, or other special language into own words.
6.	Identify, explain, show, describe, prove, define	Give the pertinent characteristics of events, classes, principles, or groups. Distinguish a particular event, class, item, from some other.
7.	Compare	Give and itemize both similarities and differenes.
8.	Contrast, distinguish	Show differences between two events, theories, entities.
9.	Interpret	Give own meaning and conclusions about the meaning of a quotation, event, theory, etc.; relate cause and effect.

Understanding Words Used in Essay Questions--Continued

10.	Discuss	Tell all pertinent data regarding the topic.
11.	Comment	State own reaction to the topic, supported with facts and illustrations.
12.	Criticize, evaluate	Gives evidence on both sides of an issue, draw conclusions, and make a judgement as to the relative worth, quality, or value of the topic.

(Helping Students Do Better on Tests, 1975)

Appendix C

Sample Item Stems for Higher Order Cognitive Questions

1. Comparing

   Describe the similarities and differences between ...
   Compare the following two methods for ...

2. Summarizing

   State the main points included in ...
   Briefly summarize the contents of ...
   Which of the following best summarizes ...

3. Classifying

   Group the following items according to ...
   What do the following items have in common?
   Which of these is an example of ...
   What is the relationship between ...

4. Applying

   Using the principle of ... as a guide, describe how you would solve the following problem/situation.
   Describe a situation that illustrates the principle of ...

5. Generalizing

   Formulate several valid generalizations from the following data.
   State the set of principles that can explain the following events.

6. Relating Cause and Effect

   What are the major causes of ...
   What would be the most likely effects of ...
   What is the reason for ...

7. Inferring

   In light of the facts presented, what is most likely to happen when ...
   How would (Senator X) be likely to react to the following issue?

Sample Item Stems for Higher Order Cognitive Questions

8. Justifying

Which of the following alternatives would you favor and why?
Explain why you agree or disagree with the following statement.

9. Creating

List as many ways as you can think of for ...
Make up a story describing what would happen if ...

10. Analyzing

Describe the reasoning errors in the following paragraph.
List and describe the main characteristics of ...

11. Synthesizing

Describe a plan for proving that ...
Write a well-organized report that shows ...

12. Evaluating

Describe the strengths and weaknesses of the following ...
Using the criteria developed in class, write a critical evaluation of ...

(Gronlund, 1981)

References

Anderman, R. C., & Williams, J. M. Teaching test-taking and notetaking skills to learning disabled high school students. (ERIC Document Reproduction Service No. 268 772)

Barnes, S. A study of classroom pupil evaluation: The missing link in teacher education. Journal of Teacher Education, 1985, 36(4), 46-49.

Brannon, D. R., Day, G. F., & Maley, D. How to evaluate students in a CBVE program. Creasaptown, Maryland: The Maryland Vocational Curriculum Production Project, 1978.

Bloom, B. S. Human characteristics and school learning. New York: McGraw Hill, 1976.

Carlson, S. A. The ethical appropriateness of subject-matter tutoring for learning disabled adolescents. Learning Disabilities Quarterly, 1985, 8(4), 310-314.

Carter, K. Do teachers understand principles for writing tests? Journal of Teacher Education, 1984, 35(6), 57-60.

Carter, K. Test-wiseness for teachers and students. Educational Measurement and Practice, 1986, 5(4), 20-23.

ERIC Clearinghouse on Handicapped and Gifted Children. Curriculum-Based Assessment, Brief 2. Reston, Virginia: The Council for Exceptional Children, December, 1988.

Ediger, M. Appraising student achievement in mathematics. (ERIC Documentation Reproduction Service No. ED 269 425).

Fraenkel, J. R. Helping students think and value: Strategies for teaching the social studies (2nd ed.). Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1980.

Gage, N. L., & Berliner, D. C. Educational psychology (4th ed.). Boston: Houghton Mifflin Company, 1988.

Gilman, D. A. Improving schools through inservice test construction: The Rossville Model. (ERIC Documentation Reproduction Service No. ED 304 445)

Griswold, P. A. Teachers' attitudes toward testing: Recommendations for district policy and staff development: An occasional report to the superintendent. Portland, Oregon: Parkrose Public Schools, 1988. (ERIC Documentation Reproduction Service No. ED 305 363).

Gronlund, N. E. Constructing achievement tests (2nd ed.). Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1977.

Gronlund, N. E. Measurement and evaluation in teaching (4th ed.). New York: Macmillan Publishing Company, Inc., 1981.

Gullickson, A. R. Teacher education and teacher-perceived needs in educational measurement and evaluation. Journal of Educational Measurement, 1986, 23(4), 347-354.

Gullickson, A. R. Teacher perspectives of their instructional use of tests. Journal of Educational Research, 1984, 77(4), 244-248.

Gullickson, A. R., & Ellwein, M. C. Post-hoc analysis of teacher-made tests: The goodness-of-fit between prescription and practice. Educational Measurement: Issues and Practice, 1985, 40(1), 15-18.

Guskey, T. R. Implementing mastery learning. Belmont, California: Wadsworth Publishing Company, 1985.

Halpern, A. S., & Benz, M. R. A statewide examination of secondary special education for students with mild disabilities: Implications for the high school curriculum. Exceptional Children, 1987, 54(2), 122-129.

Helping students do better on tests, Section II. Rockville, Maryland: Montgomery County Public Schools, 1975.

Herman, J. A. A practical approach to local test development. Resource Paper No. 6. Research into practice project. Los Angeles, California: Center for the Study of Evaluation, 1984. (ERIC Documentation Reproduction Service No. ED 252 563)

Herman, J. A., & Winters, L. Test design manual: Guidelines for developing diagnostic tests. Los Angeles, California: Center for the Study of Evaluation, 1985. (ERIC Documentation Reproduction Service No. ED 266 159)

Howell, K. W., & Morehead, M. K. Curriculum-based evaluation in remedial and special education. Columbus, Ohio: Charles E. Merrill, 1987.

Idol, L., Nevin, A., & Paolucci-Whitcomb, P. Models of curriculum-based assessment. Rockville, Maryland: Aspen Publishers, Inc., 1986.

Issues in the delivery of educational services to individuals with learning disabililities: A position paper of the National Joint Committee on Learning Disabilities, February 21, 1982. Journal of Learning Disabilities, 1987, 20(5), 286-8.

Kaufman, J. M. Arguable assumptions underlying the regular education initiative. Journal of Learning Disabilities, 1988, 21(1), 6-11.

Kirby, P. C., & Oescher, J. Testing for critical thinking: Improving test development and evaluation skills of the classsroom teacher. (ERIC Documentation Reproduction Service No. 291 757).

Marso, R. N. Testing practices and test item preferences of classroom teachers. (ERIC Documentation Reproduction Service No. ED 268 145)

Marso, R. N., & Pigge, F. L. The status of classroom teachers' test construction proficiencies assessments by teachers, principals, and supervisors validated by analyses of actual teacher-made tests. (ERIC Documentation Reproduction Service No. ED 306 283)

Marston, D., & Magnusson, D. Implementing curriculum-based measurement in special and regular education settings. Exceptional Children, 1985, 52(3), 266-276.

McKenzie, R. G. Content area instruction delivered by secondary learning disabilities teachers: A national survey. Learning Disabilities Quarterly, 1991, 14(2), 115-122.

Miller, P. W., & Erickson, H. E. Teacher-written student tests: A guide for planning, creating, administering, and assessing. Washington, D.C.: National Education Association, 1985. (ERIC Documentation Reproduction Service No. 266 182)

Miller, P. W., & Erickson, H. E. How to write tests for students. Washington, D.C.: National Education Association, 1990.

Office of Educational Research and Improvement. Help your child improve in test-taking. Washington, D.C. Author, 1987. (ERIC Document Reproduction Service No. 280 899)

Pigge, F. L., & Marso, R. N. Supervisors agenda: Identifying and alleviating teachers' test construction errors. (ERIC Document Reproduction Service No. 280 899)

Regular education/special education initiative: A position paper by the ACLD. Exceptional Parent, 1986, 16(5), 59-64.

Reynolds, M. C. A reaction to the JLD special series on the regular education initiative. Journal of Learning Disabilities, 1988, 21(6), 352-356.

Salvia, J., & Ysseldyke, J. E. Assessment in special and remedial education (2nd ed.). Boston: Houghton Mifflin Company, 1981.

Schumaker, J. B., & Deschler, D.D. Implementing the regular education initiative in secondary schools: A different ball game. Journal of Learning Disabilities, 1988, 21(1), 36-42.

Schumaker, J. B., Deshler, D. D., & Ellis, E. S. Intervention issues related to the education of LD adolescents. In J. K. Torgeson & B. Y. L. Wong (Eds.), Psychological and educational perspectives on learning disabilities. New York: Academic Press, 1986.

Seidenberg, P. L., & Koenigsberg, E. A survey of regular and special education high school teachers and college faculty: Implications for program development for secondary learning disabled students. Learning Disabilities Research, 1990, 5(2), 110-117.

Smith, G. Facilitating mainstreaming through a school-wide study skills program. Learning Disabilities Focus, 1987, 3(1), 53-54.

Stiggins, R. J. Improving assessment where it means the most: In the classroom. Educational Leadership, 1985, 43(2), 69-74.

Stiggins, R. J. Make sure your teachers understand student assessment. The Executive Educator, 1988, 10(8), 24-26,30.

Stiggins, R. J., & Bridgeford, N. J. The ecology of classroom assessment. Journal of Educational Measurement, 1985, 22(4), 271-286.

TenBrink, T. B. Evaluation: A practical guide for teachers. New York: McGraw-Hill Book Company, 1974.

U.S. Department of Education. "To assure the free appropriate public education of all handicapped children": Twelfth annual report to Congress on the implementation of the Education of the Handicapped Act. Washington, D.C.: U.S. Government Printing Office, 1990.

U.S. Department of Education. "To assure the free appropriate public education of all handicapped children": Thirteenth annual report to Congress on the implementation of The Individuals with Disabilities Education Act. Washington, D.C.: U. S. Government Printing Office, 1991.

Wagner, M. The school programs and school performances of secondary students classified as learning disabled: Findings from the national longitudinal transition study of special education students. Menlo Park, California: SRI International, 1990.

Wang, M. C., & Birch, J. W. Comparison of a full-time mainstreaming program and a resource room approach. Exceptional Children, 1984, 51(1), 33-40.

Wiersma, W., & Jurs, S. G. Educational measurement and testing. Boston: Allyn and Bacon, Inc., 1985.

Will, M. Educating students with learning problems--A shared responsibility: A report to the secretary. Washington, D.C.: U.S. Department of Education, 1986.

Williams, J. M. Tips for being a successful student: A handbook for students and teachers. (ERIC Document Reproduction Service No. 306 750)

Wood, J. W., Miederhoff, J. W., & Bishop, J. C. Stress the knowledge, not the student: Testing the mildly handicapped. The Science Teacher, 1988, 55(8), 22-25.

Written test construction. Vocational instructor teaching skills project. Gresham, Oregon: Mount Hood Community College. (ERIC Document Reproduction Service No. 262 232)