

HOW TO WRITE BETTER TESTS

A Handbook for Improving Test Construction Skills

Introduction

This handbook is designed to help instructors write better tests—better in that they more closely assess instructional objectives and assess them more accurately. A number of problems keep classroom tests from being accurate measures of students' achievement. Some of these problems are:

1. Tests include too many questions measuring only knowledge of facts. One of the most common complaints from students is that the test content did not reflect the material discussed in class or what the professor seemed to indicate was most important. This may happen because knowledge questions are the easiest to write.
2. Too little feedback is provided. If a test is to be a learning experience, students must be provided with prompt feedback about which of their answers were correct and which were incorrect.
3. The questions are often ambiguous and unclear. According to Milton (1978), ambiguous questions constitute the major weakness in college tests. Ambiguous questions often result when instructors put off writing test questions until the last minute. Careful editing and an independent review of the test items can help to minimize this problem.
4. The tests are too short to provide an adequate sample of the body of content to be covered. Short tests introduce undue error and are not fair to students.
5. The number of exams is insufficient to provide a good sample to students' attainment of the knowledge and skills the course is trying to develop. The more samples of student achievement obtained, the more confidence instructors have in the accuracy of their course grades.

PLANNING THE TEST

A taxonomy of teaching objectives (Bloom, 1956) lists several cognitive outcomes typically sought in college instruction. These outcomes are listed hierarchically in Table 1 and include Knowledge, Comprehension, Application, Analysis, Synthesis, and Evaluation. If these are desired outcomes of instruction, then classroom tests must include assessment of these objectives.

Table 1. Examples of Bloom's Cognitive Levels

| Bloom's Cognitive Level | Student Activity | Words to Use in Item Stems |
|-------------------------|--|--|
| Knowledge | Remembering facts, terms, concepts, definitions, principles | Define, list, state, identify, label, name, who? when? where? what? |
| Comprehension | Explaining/interpreting the meaning of material | Explain, predict, interpret, infer, summarize, convert, translate, give example, account for, paraphrase |
| Application | Using a concept or principle to solve a problem | Apply, solve, show, make use of, modify, demonstrate, compute |
| Synthesis | Producing something new or original from component parts | Design, construct, develop, formulate, imagine, create, change, write a poem or short story |
| Evaluation | Making a judgment based on a pre-established set of criteria | Appraise, evaluate, justify, judge, critique, recommend, which would be better? |

The easiest way to ensure a representative sample of content and cognitive objectives on the test is to prepare a table of specifications. This table is simply a two-way chart listing the content topics on one dimension and the cognitive skills on the other. We want to include content and skills in the same proportion as they were stressed during instruction. Table 2 shows a simple table of specifications; it is intended to be illustrative, not comprehensive.

Table 2. Table of Specifications for a Chemistry Unit Test on Oxygen

| Content (%) | Knowledge | Comprehension | Application | Total (%) |
|---------------------|-----------|---------------|-------------|-----------|
| Physical Properties | 8 | 6 | 6 | 20 |
| Chemical Properties | 12 | 9 | 9 | 30 |
| Preparation | 4 | 3 | 3 | 10 |
| Uses | 16 | 12 | 12 | 40 |
| Total | 40 | 30 | 30 | 100 |

This table indicates the content topics and the objectives to be covered and the proportion of the test that will be devoted to each. Evidently, more class time was spent on the uses of oxygen because 40 percent of the test questions deal with uses compared with only 10 percent on preparation. The column totals indicate that 40% of the items will be written at the knowledge level with the remaining divided equally between comprehension and application. Using the percentages assigned to each cell, one writes the appropriate number of items. For example, because 20% of the test is to cover physical properties and 30% is to be application, then 6% of the total test would measure the ability to apply knowledge about oxygen's physical properties to new situations.

Coordinating test content with instruction content ensures content validity of the test. Using a table of specifications also helps an instructor avoid one of the most common mistakes in classroom tests, namely writing all the items at the knowledge level.

THE TEST FORMAT

After planning the content and cognitive objectives for the test, instructors must decide on the best way to measure them; that is, they decide on the test format. The format refers to whether the test will be objective (multiple choice, true false, matching, etc.) or essay. What factors do faculty consider when deciding on the format of the test?

1. What is to be Measured?

We should choose the format that is most appropriate for measuring the cognitive objectives on the test. If instructors want students to contrast A and B, take a position on an issue and defend it, create a plan, and perform other similar tasks, then they would most likely use an essay format. For example, if an instructor wants students to explain the role of the press in the coming of the Civil War, he/she would probably choose an essay item. But if the objective is to identify the authors of selected writings about the coming of the War, then the instructor could use an objective type format.

Many times instructors have a choice. Objective-type items can be used quite effectively to measure high level cognitive objectives. A common myth depicts objective items as measuring simple factual recall and essays as evaluating higher-order thinking. But multiple choice items, for example, can be written to measure reasoning, comprehension, application, analysis, and other complex thinking processes. What other factors might influence the decision about format?

2. The Size of the Class

Class size is often an important factor influencing the decision about test format. It is very difficult to give essay tests when there are 400 students in the class because the scoring time is prohibitive. A survey of 1100 professors from across the country (Cross, 1990) showed that class size is the factor that professors consider most important when they decide what test format to use. Two-thirds of the faculty surveyed said they preferred the essay format but could not use it because of the size of their classes. They used essay tests only in small classes.

3. Time Available to Prepare and Score Test

It takes a long time to score an essay test. By contrast, it takes a long time to construct a multiple-choice test. Instructors must consider whether they will have more time available when preparing or when scoring the test. If instructors are short of time when a test must be prepared, then they might choose an essay test, if class size permits. We are not implying that good essay questions are easy to write; essay tests are easier to prepare only because fewer questions have to be written.

ESSAY ITEMS

Let us look at the relative strengths and weaknesses of the essay format.

Strengths of Essay Items

1. Essay items are an effective way to measure higher-level cognitive objectives. They are unique in measuring students' ability to select content, organize and integrate it, and present it in logical prose.
2. They are less time-consuming to construct.
3. They have a good effect on students' learning. Students do not memorize facts, but try to get a broad understanding of complex ideas, to see relationships, etc.
4. They present a more realistic task to the student. In real life, questions will not be presented in a multiple-choice format, but will require students to organize and communicate their thoughts.

Limitations of Essay Items

1. Because of the time required to answer each question, essay items sample less of the content.
2. They require a long time to read and score.
3. They are difficult to score objectively and reliably. Research shows that a number of factors can bias the scoring:

A) Different scores may be assigned by different readers or by the same reader at different times.

- B) A context effect may operate; an essay preceded by a top quality essay receives lower marks than when preceded by a poor quality essay.
- C) The higher the essay is in the stack of papers, the higher the score assigned.
- D) Papers that have strong answers to items appearing early in the test and weaker answers later will fare better than papers with the weaker answers appearing first.
- E) Scores are influenced by the expectations that the reader has for the student's performance. If the reader has high expectations, a higher score is assigned than if the reader has low expectations. If we have a good impression of the student, we tend to give him/her the benefit of the doubt.
- F) Scores are influenced by quality of handwriting, neatness, spelling, grammar, vocabulary, etc.

Writing Good Essay Items

1. Formulate the question so that the task is clearly defined for the student. Use words that "aim" the student to the approach you want them to take. Words like discuss and explain can be ambiguous. If you use "discuss", then give specific instructions as to what points should be discussed.

Poor: Discuss Karl Marx's philosophy.

Better: Compare Marx and Nietzsche in their analysis of the underlying problems of their day in 19th century European society.

Clearly stated questions not only make essay tests easier for students to answer, but Also make them easier for instructors to score.

2. In order to obtain a broader sampling of course content, use a relatively large number of questions requiring shorter answers (one-half page) rather than just a few questions involving long answers (2-3 pages).
3. Avoid the use of optional questions on an essay test. When students answer different questions, they are actually taking different tests. If there are five essay questions and students are told to answer any three of them, then there are ten different tests possible. It makes it difficult to discriminate between the student who could respond correctly to all five, and the student who could answer only three. Use of optional questions also affects the reliability of the scoring. If we are going to compare students for scoring purposes, then all students should perform the same tasks. Another problem is that students may not study all the course material if they know they will have a choice among the questions.
4. Indicate for each question the number of points to be earned for a correct response. If time is running short, students may have to choose which questions to answer. They will want to work on the questions that are worth the most points.
5. Avoid writing essay items that only require students to demonstrate certain factual knowledge. Factual knowledge can be measured more efficiently with objective-type items.

Writing Essay Items at Different Levels of Bloom's Taxonomy

The goal is to write essay items that measure higher cognitive processes. The question should represent a problem situation that tests the student's ability to use knowledge in order to analyze, justify, explain, contrast, evaluate, and so on. Try to use verbs that elicit the kind of thinking you want them to demonstrate. Instructors often have to use their best judgment about what cognitive skill each question is measuring. You might ask a colleague to read your questions and classify them according to Bloom's taxonomy.

Another point that should be emphasized when writing items that measure higher cognitive processes is that these processes build on and thus include the lower levels of knowledge and comprehension. Before a student can write an essay requiring analysis, for example, he/she must have knowledge and a basic understanding of the problem. If the lower level processes are deficient, then the higher-level ones won't operate at the maximum level. The following are examples of essay items that appear to measure at different levels:

Knowledge: Identify the "wage fund doctrine".

Comprehension: Explain the following: Aquinas was to Aristotle as Marx was to Ricardo.

Application: Use the "wage fund doctrine" to explain wage rate in the writing of J.S. Mill.

Analysis: Compare and contrast the attitudes toward male and female sex roles in the work of Ibsen and Huysmans.

Synthesis: Write an essay contrasting Nietzsche's approach to the question of "truth" with that of Comte's development.

Evaluation: Using the five criteria discussed in class, critically evaluate Adam Smith's theory of economic

Scoring Essay Tests

The major task in scoring essay tests is to maintain consistency, to make sure that answers of equal quality are given the same number of points. There are two approaches to scoring essay items: (1) analytic or point method and (2) holistic or rating method.

1. Analytic: Before scoring, one prepares an ideal answer in which the major components are defined and assigned point values. One reads and compares the student's answer with the model answer. If all the necessary elements are present, the student receives the maximum number of points. Partial credit is given based on the elements included in the answer. In order to arrive at the overall exam score, the instructor adds the points earned on the separate questions.
2. Holistic: This method involves considering the student's answer as a whole and judging the total quality of the answer relative to other student responses or the total quality of the answer based on certain criteria that you develop.

As an instructor reads the answers to a particular question, he/she sorts the papers into stacks based on the overall quality. The best answers go into the first stack, the average go into the second stack, and the poorest into the third stack. After further examination of the answers in each stack, one may want to divide some of these stacks to make additional ones. Then points are written on each paper appropriate to the stack it is in.

Suggestions for Scoring Essays

1. Grade the papers anonymously. This will help control the influence of our expectations about the student on the evaluation of the answer.
2. Read and score the answers to one question before going on to the next question. In other words, score all the students' responses to Question 1 before looking at Question 2. This helps to keep one frame of reference and one set of criteria in mind through all the papers, which results in more consistent grading. It also prevents an impression that we form in reading one question from carrying over to our reading of the student's next answer. If a student has not done a good job on say the first question; we could let this impression influence our evaluation of the student's second answer. But if other students' papers come in between, we are less likely to be influenced by the original impression.
3. If possible, also try to grade all the answers to one particular question without interruption. Our standards might vary from morning to night, or one day to the next.
4. Shuffle the papers after each item is cored throughout all the papers. Changing the order reduces the context effect and the possibility that a student's score is the result of the **location** of the paper in relationship to other papers. If Mary's B work always followed John's A work, then it might look more like C work and her grade would be lower than if her paper were somewhere else in the stack.
5. Decide in advance how you are going to handle extraneous factors and be consistent in applying the rule. Students should be informed about how you treat such things as misspelled words, neatness, handwriting, grammar, and so on.
6. Be on the alert for bluffing. Some students who do not know the answer may write a well-organized coherent essay but one containing material irrelevant to the question. Decide how to treat irrelevant or inaccurate information contained in students' answers. We should not give credit for irrelevant material. It is not fair to other students who may also have preferred to write on another topic, but instead wrote on the required question.
7. Write comments on the students' answers. Teacher comments make essay tests a good learning experience for students. They also serve to refresh your memory of your evaluation should the student question the grade.

Preparing Students to Take Essay Exams

Essay tests are valid measures of student achievement only if students know how to take them. Many college freshmen do not know how to take an essay exam, because they haven't been required to learn this skill in high school. You may need to take some class time to tell students how to prepare for and how to take an essay exam. You might use some of your old exam questions, and let students see what an A answer looks like and how it differs from a C answer.

MULTIPLE-CHOICE ITEMS

Many users regard the multiple-choice item as the most flexible and probably the most effective of the objective item types. A multiple-choice item consists of two parts: (1) the stem, which presents a specific problem to the test taker and (2) a list of possible solutions or answers called distractors. The stem may be written either as a question or as an incomplete statement. There should be only one correct or best answer while the other three or four options serve as distractors.

Strengths of Multiple-Choice Items

1. Versatility in measuring all levels of cognitive skills.
2. Permit a wide sampling of content and objectives.
3. Provide highly reliable test scores.
4. Can be machine-scored quickly and accurately.
5. Reduced guessing factor compared with true-false items.

Limitations of Multiple-Choice Items

1. Difficult and time-consuming to construct.
2. Depend on student's reading skills and instructor's writing ability.
3. Ease of writing low-level knowledge items leads instructors to neglect writing items to test higher-level thinking.
4. May encourage guessing (but less than true-false).

Writing Multiple-Choice Items

The challenge is to write questions that test a significant concept, that are unambiguous, and that don't give test-wise students an advantage.

1. The stem should fully state the problem and all qualifications. To make sure that the stem presents a problem, always include a verb in the statement.
2. Concentrate on writing items that measure students' ability to comprehend, apply, analyze, and evaluate as well as recall.
3. Include words in the stem that would otherwise be repeated in each option. Following this guideline not only saves time for the typist but also saves reading time for the student.

- Poor: Sociobiology can be defined as
- a. the scientific study of humans and their relationships within the environment.
 - b. the scientific study of animal societies and communication.
 - c. the scientific study of plants and their reproductive processes.
 - d. the scientific study of the number of species in existence.

- Better: Sociobiology can be defined as the scientific study of
- a. humans and their relationships within the environment.
 - b. animal societies and communication.
 - c. plants and their reproductive processes.
 - d. the number of species in existence.

4. Eliminate excessive wording and irrelevant information in the stem.
5. Make sure there is only one correct or best response.

Poor: The function of the hypothesis in a research study is to provide

- a. tentative explanation of phenomena.
- b. proven explanation of phenomena.
- c. framework for interpretation of the findings.
- d. direction for the research.

There is no single or best answer, options a, c, and d are correct. The options need to be reworded so that only one is clearly best or correct. Or one could change the stem to read: According to the lecture (or the text), the **most important** function of the hypothesis is...

6. Provide a minimum of three, but not more than five, plausible and attractive options for each item. A good procedure is to think of errors that students are likely to make and use these as distractors.

Poor: The recent (1989) research suggesting that controlled nuclear-fusion could be effected in a laboratory experiment at room temperature was conducted by

- a. Watson and Crick.
- b. Pons and Fleischmann.
- c. Koch and Jenner.
- d. Fermi and Bohr.

While the first two options are plausible, the last two are not. The latter should be replaced by the names of contemporary scientists.

7. Make all the options for an item approximately homogeneous in content, form, and grammatical structure. Increasing the homogeneity of the content among the options can increase the difficulty of an item. (Difficulty of a test should not be based on inclusion of obscure content.)
8. Avoid the use of the all-of-the-above and none-of-the-above options. The problem with “all of the above” as an option is that it makes the item too easy. If students can recognize at least one incorrect option, they can eliminate “all of the above” as a viable option. On the other hand, if they can recognize at least two correct options, then they know that “all of the above” is the correct answer. Furthermore, research shows that when “all of the above” is used as a distractor, it is too often the correct response. Students are quick to pick up on this clue.

“None of the above” should be used only when absolute standards of correctness can be applied, such as in math, grammar, spelling, geography, historical dates, and so on. Otherwise, students can often argue about the correctness of one of the other options.

9. Avoid verbal associations between the stem and the correct option, e.g., the same reference word should not appear in the stem and an option. Also make sure that the options are grammatically consistent with the stem.

Poor: The correlation coefficient found by correlating students' scores on a classroom math test with their scores on a standardized math test is called a

- a. validity coefficient.
- b. index of reliability.
- c. equivalence coefficient.
- d. internal consistency coefficient.

Option (a) is the only one that is grammatically consistent with the stem. It could be correctly selected without knowing anything about the content. One should change the "a" in the stem to "a(n)".

10. Avoid making the correct answer markedly longer or shorter than the other options.
11. If there is a logical sequence in which the alternatives can be arranged (alphabetical if a single word, in order of magnitude if numerals, in temporal sequence, or by length of response), use that sequence.
12. Use negatively stated stems sparingly. When used, call attention to the negative word by underlining and/or capitalizing.
13. Randomly distribute the correct response among the alternative positions throughout the test. That is, have approximately the same proportion of A's, B's, C's, D's, and E's as the correct response.
14. Watch for specific determiners such as "all", "always", "never" which are more likely to be in incorrect options. Others like "usually" and "sometimes" are more likely to be in the keyed response.
15. Multiple-choice items should be independent. That is, an answer to one question should not depend on the answer to another question.
16. Avoid the use of language that your students won't understand. For example (unless it's a French test), use "cause" instead of "raison d'etre" in the question.
17. State items so there can be only one interpretation of their meaning.

Poor: Which one of the following is the best source of heat for home use?

- a. Gas
- b. Electricity
- c. Oil
- d. Geo-thermal

The answer would depend on how one interprets the question. Are we talking about the best source economically, in terms of cleanness, in terms of efficiency, or just what? Also the correct answer might depend on what part of the world we're asking about.

Better: The most economical source of heat in the Midwestern U.S. is

- a. gas
- b. electricity
- c. oil
- d. geo-thermal

Suggestions for Writing Multiple-Choice Items Which Measure Higher Objectives

It is difficult and time-consuming to write multiple-choice items that measure the higher thinking skills. The item writer has to be creative in order to develop challenging questions. The following suggestions may provide some ideas for writing these kinds of questions.

1. Present practical or real-world situations to the students. These problems may use short paragraphs describing a problem in a practical situation. Items can be written which call for the application of principles to the solution of these practical problems, or the evaluation of several alternative procedures.
2. Present the student with a diagram of equipment and ask for application, analysis, or evaluations, e.g., “What happens at point A if ...?,” “How is A related to B?”
3. Present actual quotations taken from newspapers or other published sources or contrived quotations that could have come from such sources. Ask for the interpretation or evaluation of these quotations.
4. Use pictorial materials that require students to apply principles and concepts.
5. Use charts, tables or figures that require interpretation.

Table 3 shows multiple-choice items that measure at different levels.

Table 3. Multiple-Choice Items That Measure at Various Levels.

1. Knowledge
Which of the following are the raw materials for photosynthesis?
 - a. Water, heat, sunlight
 - b. Carbon dioxide, sunlight, oxygen
 - c. Water, carbon dioxide, sunlight
 - d. Sunlight, oxygen, carbohydrates
 - e. Water, carbon dioxide, carbohydrates

2. Comprehension
If living cells similar to those found on earth were found on another planet where there was no molecular oxygen, which cell part would most likely be absent?
 - a. cell membrane
 - b. nucleus
 - c. mitochondria
 - d. ribosome
 - e. chromosomes

3. Application
Phenylketonuria (PKU) is an autosomal recessive condition. About one in every fifty individuals is heterozygous for the gene but shows no symptoms of the disorder. If you select a symptom-free male and a symptom-free female at random, what is the probability that they could have a child afflicted with PKU?
 - a. $(.02)(.02)(.25) = 0.0001 = 0.01\%$, or about 1/10,000
 - b. $(.02)(.02) = 0.0004 = 0.04\%$, or about 1 /2,500
 - c. $(1)(50)(2) = 100\% = \text{all}$
 - d. $(1)(50)(0) = 0 = \text{none}$
 - e. $1/50 = 2\%$, or 2/100

4. Analysis
Mitochondria are called the powerhouses of the cell because they make energy available for cellular metabolism. Which of the following observations is *most* cogent in supporting this concept of mitochondrial function?
 - a. ATP occurs in the mitochondria.
 - b. Mitochondria have a double membrane.
 - c. The enzymes of the Krebs cycle, and molecules required for terminal respiration, are found in mitochondria.
 - d. Mitochondria are found in almost all kinds of plant and animal cells.
 - e. Mitochondria abound in muscle tissue.

5. Evaluation
Disregarding the relative feasibility of the following procedures, which of these lines of research is likely to provide us with the most valid and direct evidence as to evolutionary relations among different species?
 - a. Analysis of the chemistry of stored food in female gametes.
 - b. Analysis of the enzymes of the Krebs cycle.
 - c. Observations of the form and arrangement of the endoplasmic reticulum.
 - d. Comparison of details of the molecular structure of DNA.
 - e. Determination of the total percent protein in the cells.

Note: The writers are indebted to Dr. Michael Tansey of the Biology Department of Indiana University, Bloomington, for these items.

TRUE-FALSE ITEMS

The true-false item typically present a declarative statement that the student must mark as either true or false. Instructors generally use true- false items to measure the recall of factual knowledge such as names, events, dates, definitions, etc. But this format has the potential to measure higher levels of cognitive ability, such as comprehension of significant ideas and their application in solving problems.

T F 1. Jupiter is the largest planet in the solar system.

T F 2. If Triangle ABC is isosceles and angle A measures 100 degrees, then angle B is 100 degrees.

T F 3. If a distribution of scores has a few extremely low scores, then the median will be numerically larger than the mean.

T F 4. The larger the number of scores in a distribution, the larger the standard deviation of the score must be.

The first example above measures recall of a specific fact. The other examples, however, show how a true-false item can be written to measure comprehension and application.

Strengths of True -False Items

1. They are relatively easy to write and can be answered quickly by students. Students can answer 50 true- false items in the time it takes to answer 30 multiple-choice items.
2. They provide the widest sampling of content per unit of time.

Limitations of True -False Items

1. The problem of guessing is the major weakness. Students have a fifty-percent chance of correctly answering an item without any knowledge of the content.
2. Items are often ambiguous because of the difficulty of writing statements that are unequivocally true or false.

Writing True -False Items

1. Test significant content and avoid trivial statements.
2. Write items that can be classified unequivocally as either true or false.
3. Avoid taking statements verbatim from textbooks.

T F Poor: The square of the hypotenuse of a right triangle equals the sum of the squares of the other two sides.

T F Better: If the hypotenuse of an isosceles right triangle is 7 inches, each of the two equal sides must be more than 5 inches.

4. Include only a single major point in each item.
5. Avoid trick questions.

T F Poor: "The Raven" was written by Edgar Allen Poe.

T F Better: "The Raven" was written by Edgar Allan Poe.

The intent of the question should be to determine if students know that Poe write "The Raven", not to see if they notice the incorrect spelling of his middle name.

6. Try to avoid using words like “always,” “all”, or “never which tend to make the statement false; words like “usually,” “often,” “many” usually make the statement true.
7. Avoid using negatively worded statements.

T F Poor: *Silas Marner* was not written by Thomas Hardy.

T F Better: *Silas Marner* was written by Thomas Hardy.

8. Put the items in a random order so as to avoid response patterns that could serve as clues (such as T,T,F,T,T,F)
9. Try to avoid long drawn-out statements or complex sentences with many qualifiers.
10. Avoid making items that are true consistently longer than those that are false.
11. Use slightly more false items than true items. False items tend to discriminate more highly among students than do true items. Research shows that when students guess they are more inclined to respond with a true than with a false. We can compensate for this “acquiescent response set” by having a few more false statements than true.

Variations of the T-F Format

Changing false statements to make them true:

The student indicates whether the statement is true or false; if false, he/she must change an underlined word to make the statement true.

T F electrons 1. Subatomic particles of negatively charged electricity are called protons.

T F _____ 2. The green coloring matter in plants is called chlorophyll.

Items measuring ability to recognize cause-and-effect:

The item has two parts, both of which are true; the student must decide if the second part explains why the first part is true.

Yes No 1. Leaves are essential *because* they shade the tree trunk.

Yes No 2. Iron rusts *because* oxidation occurs.

MATCHING

A matching exercise typically consists of a list of questions or problems to be answered along with a list of responses. The examinee is required to make an association between each question and a response.

Example:

| I | II |
|---|-----------------|
| 1. a substance of low solubility | A. distillation |
| 2. two liquids that do not dissolve in each other | B. miscible |
| 3. a substance that does the dissolving | C. immiscible |
| 4. a method of purifying a substance | D. precipitate |
| 5. the substance being dissolved | E. soluble |
| | F. solute |
| | G. solvent |

The problems can be in various forms. The most common is to use verbal statements, but other types of material can be used. For example, the problems might be locations on a map, geographic features on a contour map, parts of a diagram of the body or biological specimens or math problems. Similarly, the responses don't have to be terms or labels, they might be functions of various parts of the body, or methods, principles, or solutions.

Example:

| I | II |
|--------------------|----------|
| 1. $\frac{3}{4}$ | A. 0.060 |
| 2. $\frac{3}{5}$ | B. 0.500 |
| 3. $\frac{5}{8}$ | C. 0.600 |
| 4. $\frac{3}{50}$ | D. 0.625 |
| 5. $\frac{14}{28}$ | E. 0.750 |
| | F. 0.875 |

Previously, it was difficult to use machine scoring for the matching format. However, a ten-choice, machine-scannable answer sheet that makes it possible to use matching exercises with up to 10 possible responses per question can be purchased from BEST.

Because matching items permit one to cover a lot of content in one exercise, they are an efficient way to measure. It is difficult, however, to write matching items that require more than simple recall of factual knowledge.

Guidelines for Constructing Matching Items

1. Use homogeneous material in each list of a matching exercise. Mixing events and dates with events and names of persons, for example, makes the exercise two separate sets of questions and gives students a better chance to guess the correct response. For example, if one stem were "president of U.S. during World War II", the student could ignore all the responses other than names. Using homogeneous materials requires students to distinguish or discriminate among things which makes for a more challenging task.
2. Include directions that clearly state the basis for the matching. Inform students whether or not a response can be used more than once and where answers are to be written.

- Put the problems or the stems (typically longer than the responses) in a numbered column at the left, and the response choices in a lettered column at the right. Because the student must scan the list of responses for each problem, one should keep the responses brief. This saves reading time for the student.
- Always include more responses than questions. If the lists are the same length, the last choice may be determined by elimination rather than knowledge.
- Arrange the list of responses in alphabetical or numerical order if possible in order to save reading time.
- All the response choices must be plausible, but make sure that there is only one correct choice for each stem or numbered question.

COMPLETION ITEMS

The completion format requires the student to answer a question or to finish an incomplete statement by filling in a blank with the correct word or phrase. The advantages of completion items are (1) they provide a wide sampling of content; and (2) they minimize guessing compared with multiple-choice and true-false. The limitations are they (1) rarely can be written to measure more than simple recall of information; (2) are more time-consuming to score than other objective types; (3) are difficult to write so there is only one correct answer and no irrelevant clues.

Guidelines for Writing Completion Items

- Omit only significant words from the statement, but do not omit so many words that the statement becomes ambiguous.

Poor: The Constitutional Convention met in _____ in _____ .
 Better: The Constitutional Convention met in the city of _____ n 1787.

- Write completion items that have a single correct answer, if possible.

Poor: Abraham Lincoln was born in _____ .
 There are several legitimate answers: Kentucky, 1809, February, a log cabin, etc.

Better: Abraham Lincoln was born in the state of _____ .

- Use blanks of the same length throughout the test so that the length is not a clue
- Avoid grammatical clues to the correct response. For example, if the indefinite article is required before a blank, use a(n) so that the student doesn't know if the correct answer begins with a vowel or a consonant.

Poor: A subatomic particle with a negative electric charge is called an _____ .
 The student could eliminate proton, neutron, and meson as possible responses.

Better: A subatomic particle with a negative electric charge is called a(n) _____ .

5. If possible, put the blank at the end of a statement rather than at the beginning. Asking for a response before the student understands the intent of the statement can be confusing and may require more reading time.

Poor _____ is the measure of central tendency that is most affected by extremely high or low scores.
Better: The measure of central tendency that is most affected by extremely high or low scores is the ____ .

6. Avoid taking statements directly from the text.

Scoring

Scoring completion items is less objective than multiple-choice or true-false because the student supplies his/her own response. It is difficult to write completion items so that there is only one correct answer. When preparing a key, one should list the correct answer and any other acceptable alternatives. Be consistent in using the key; it would not be fair to accept an answer as right on one paper and not accept it on others.

Adapted from:

Jacobs, L. C. and Chase, C. I. (1992) *Developing and Using Tests Effectively*. San Francisco: Jossey-Bass.