Aviation Instructor's Handbook (FAA-H-8083-9)
Appendix B: Developing a Test Item Bank

Developing a bank of test questions is difficult and time consuming. This task demands a mastery of the subject, an ability to write clearly, and an ability to visualize realistic situations for use in developing problems. Because it is so difficult to develop good test items, many instructors rely on commercially-prepared test materials. However, there are a few instructors who write their own questions and maintain them in a database or a system of folders. *[Figure B-1]*
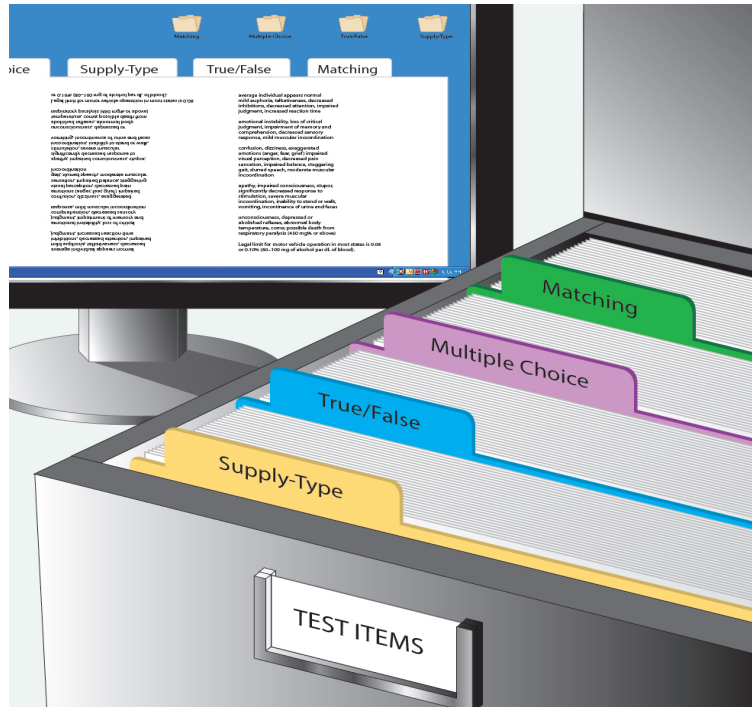

Figure B-1. *A bank of test items makes it easier to construct new tests.*

Written Test Items
Supply Type

Supply type test items ask the learner to furnish a response in the form of a word, sentence, or paragraph. The supply-type item allows the learner to organize knowledge. It shows an ability to express ideas and measures the learner's generalized understanding of a subject. For example, the supply type item on a pre-solo knowledge test can be very helpful in determining whether the pilot-in-training has adequate knowledge of procedures.

There are several disadvantages of supply-type items. First, they cannot be graded with reliability. The same test graded by different instructors could be assigned different scores. Even the same test graded by the same instructor on consecutive days might be assigned a different score. Second, supply-type items require more time for the learner to complete and more time for the instructor to grade.

Selection Type

Selection type test items require the learner to select from two or more alternatives. There is a single correct response for each item. It assumes all learners should learn the same thing, and relies on rote memorization of facts. Written tests made up of selection type items are highly reliable, meaning that the results would be graded the same regardless of the learner taking the test or the person grading it. In fact, this type of test item lends itself very well to machine scoring.

Also, selection type items make it possible to directly compare learner accomplishment. For example, it is possible to compare the performance of learners within one class to learners in a different class, or learners under one instructor with those under another instructor. By using selection type items, the instructor can test on many more areas of knowledge in a given time than could be done by requiring the learner to supply written responses. This increase in comprehensiveness can be expected to increase validity and discrimination. Selection type tests are well adapted to statistical item analysis.

True-False

The true-false test item consists of a statement followed by an opportunity for the learner to choose whether the statement is true or false. This item type has a wide range of usage. It is well adapted for testing knowledge of facts and details, especially when there are only two possible answers.

The chief disadvantage is that true-false questions create the greatest probability of guessing. Also, true-false questions are more likely to utilize rote memory than knowledge of the subject. In general, therefore, true-false questions are not considered valid (i.e., they do not measure what they are intended to measure).

To use true-false questions, consider the following guidelines for effective test items:

- Include only one idea in each statement.

- Use original statements rather than verbatim text.

- Make the statement entirely true or entirely false.

- Avoid the unnecessary use of negatives, which tend to confuse the reader.

- Underline or otherwise emphasize the negative word(s) if they must be used.

- Keep wording and sentence structure as simple as possible.

- Make statements both definite and clear.

- Avoid the use of ambiguous words and terms (some, any, generally, most times, etc.)

- Use terms which mean the same thing to all learners whenever possible.

- Avoid absolutes (all, every, only, no, never, etc.). These words are known as determiners because they provide clues to the correct answer.

- Avoid patterns in the sequence of correct responses because learners can often identify the patterns.

- .Make statements brief and approximately same length.

- State the source of a statement if it is controversial (sources have differing information).

Multiple Choice

A multiple choice test item consists of two parts: the stem, which includes the question, statement, or problem; and a list of possible responses. Incorrect answers are called distractors. When properly devised and constructed, multiple choice items offer several advantages that make this type more widely used and versatile than either the matching or the true-false items. *[Figure B-2]*

Multiple choice test questions can help determine learner achievement, ranging from acquisition of facts to understanding, reasoning, and ability to apply what has been learned. It is appropriate to use multiple choice when the question, statement, or problem has the following characteristics:

- Built-in and unique solution, such as a specific application of laws or principles.

- Wording of the item is clearly limiting, so that the learner must choose the best of several offered solutions rather than a universal solution.

- Several options that are plausible or even scientifically accurate.

- Several pertinent solutions, with the learner asked to identify the most appropriate solution.
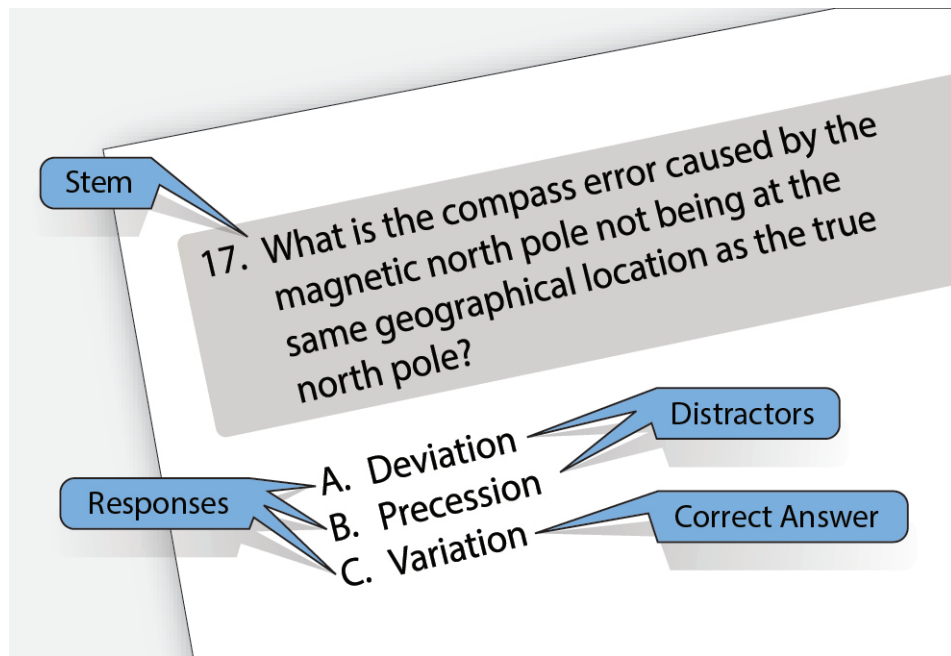
Figure B-2. *Sample multiple choice test item.*

Three major challenges are common in the construction of multiple choice test items. One is the development of a question or an item stem that must be expressed clearly and without ambiguity. A second is that the statement of an answer or correct response cannot be refuted. Finally, the distractors must be written in such a way that they are attractive to those learners who do not possess the knowledge or understanding necessary to recognize the keyed response.

A multiple choice item stem may take one of several basic forms:

- A direct question followed by several possible answers.

- An incomplete sentence followed by several possible phrases that complete the sentence.

- A stated problem based on an accompanying graph, diagram, or other artwork followed by the correct response and the distractors.

The learner may be asked to select the one correct choice or completion, the one choice that is an incorrect answer or completion, or the one choice that is the best answer option presented in the test item.

Beginning test writers find it easier to write items in the question form. In general, the form with the options as answers to a question is preferable to the form that uses an incomplete statement as the stem. It is more easily phrased and is more natural for the learner to read. Less likely to contain ambiguities, it usually results in more similarity between the options and gives fewer clues to the correct response.

When multiple choice questions are used, three or four alternatives are generally provided. It is usually difficult to construct more than four convincing responses; that is, responses which appear to be correct to a person who has not mastered the subject matter. Learners are not supposed to guess the correct option; they should select an alternative only if they know it is correct. An effective means of diverting the learner from the correct response is to use common learner errors as distractors. For example, if writing a question on the conversion of degrees Celsius to degrees Fahrenheit, providing alternatives derived by using incorrect formulas would be logical, since using the wrong formula is a common learner error.

Items intended to measure the rote level of learning should have only one correct alternative; all other alternatives should be clearly incorrect. When items are to measure achievement at a higher level of learning, some or all of the alternatives should be acceptable responses—but one should be clearly better than the others. In either case, the instructions given should direct the learner to select the best alternative.

To use multiple choice questions, consider the following guidelines for construction of effective test items:

- Make each item independent of every other item in the test. Do not permit one question to reveal, or depend on, the correct answer to another question.

- Design questions that call for essential knowledge rather than for abstract background knowledge or unimportant facts.

- State each question in language appropriate to the learners.

- Include sketches, diagrams, or pictures when they can present a situation more vividly than words. They generally speed the testing process, add interest, and help to avoid reading difficulties and technical language.

- When a negative is used, emphasize the negative word or phrase by underlining, bold facing, italicizing, or printing in a different color.

- Avoid questions containing double negatives, which invariably cause confusion.

- Avoid trick questions, unimportant details, ambiguities, and leading questions that confuse and antagonize the learner. If attention to detail is an objective, detailed construction of alternatives is preferable to trick questions.

Stems

When developing the stem of a multiple choice item, the following general principles should be utilized. *[Figure B-3]*

- The stem should clearly present the central problem or idea. The function of the stem is to set the stage for the alternatives that follow.

- The stem should contain only material relevant to its solution, unless the selection of what is relevant is part of the problem.

- The stem should be worded in such a way that it does not give away the correct response. Avoid the use of determiners, such as clue words or phrases.

- Put everything that pertains to all alternatives in the stem of the item. This helps to avoid repetitious alternatives and saves time.

- Generally avoid using "a" or "an" at the end of the stem. They may give away the correct choice. Every alternative should grammatically fit with the stem of the item.
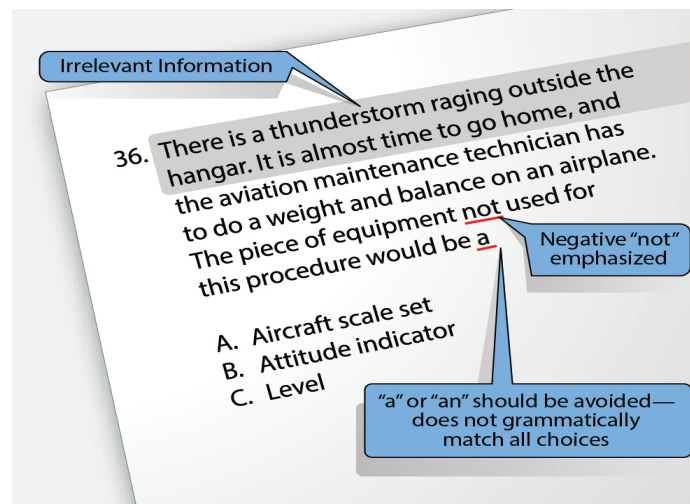

Figure B-3. *This is an example of a multiple choice question with a poorly written stem.*

Alternatives

The alternatives in a multiple choice test item are as important as the stem. They should be formulated with care; simply being incorrect should not be the only criterion for the distracting alternatives.

Popular distractors are:

- An incorrect response related to the situation and which sounds convincing.

- A common misconception.

- A statement which is true, but which does not satisfy the requirements of the problem.

- A statement that is either too broad or too narrow for the requirements of the problem.

Research of instructor-made tests reveals that, in general, correct alternatives are longer than incorrect ones. When alternatives are numbers, they should generally be listed in ascending or descending order of magnitude or length.

Matching

A matching test item consists of two lists, which may include a combination of words, terms, illustrations, phrases, or sentences. The learner matches alternatives in one list with related alternatives in a second list.

In reality, a matching exercise is a collection of related multiple choice items. In a given period of time, more samples of a learner's knowledge usually can be measured with matching rather than multiple choice items. The matching item is particularly good for measuring a learner's ability to recognize relationships and to make associations between terms, parts, words, phrases, clauses, or symbols listed in one column with related items in another column. Matching reduces the probability of guessing correct responses, especially if alternatives may be used more than once. The testing time can also be used more efficiently.

The following guidelines help in the construction of effective matching test items:

- Give specific and complete instructions. Do not make the learner guess what is required.

- Test only essential information; never test unimportant details.

- Use closely related materials throughout an item. If learners can divide the alternatives into distinct groups, the item is reduced to several multiple choice items with few alternatives, and the possibility of guessing is distinctly increased.

- Make all alternatives credible responses to each element in the first column, wherever possible, to minimize guessing by elimination.

- Use language the learner can understand. By reducing language barriers, both the validity and reliability of the test is improved.

- Arrange the alternatives in some sensible order. An alphabetical arrangement is common.

Matching-type test items are either equal column or unequal column. An equal column test item has the same number of alternatives in each column. When using this form, always provide for some items in the response column to be used more than once, or not at all, to preclude guessing by elimination. Unequal column type test items have more alternatives in the second column than in the first and are generally preferable to equal columns.