

Test Construction: Some Practical Ideas

Marilla D. Svinicki, CTE

General steps in test construction (elaborations for some of these steps are presented on the following pages):

1. Outline either a) the unit learning objectives or b) the unit content major concepts to be covered by the test.
2. Produce a test blueprint, plotting the outline from step 1 against some hierarchy representing levels of cognitive difficulty or depth of processing.
3. For each check on the blueprint, match the question level indicated with a question type appropriate to that level.
4. For each check on the blueprint, jot down on a 3x5 card three or four alternative question ideas and item types which will get at the same objective.
5. Put all the cards with the same item type together and write the first draft of the items following the guidelines for the item types on the accompanying pages. Write these on the item cards.
6. Put all the cards with the same topic together to cross check questions so that no question gives the answer to another question.
7. Put the cards aside for one or two days.
8. Reread the items from the standpoint of a student, checking for construction errors.
9. Order the selected questions logically:
 - a) Place some simpler items at the beginning to ease students into the exam;
 - b) group item types together under common instructions to save reading time;
 - c) if desirable, order the questions logically from a content standpoint (e.g. chronologically, in conceptual groups, etc.).
10. Put the questions away for one or two days before rereading them or have someone else review them for clarity.
11. Time yourself in actually taking the test and then multiply that by four to six depending on the level of the students. Remember, there is a certain absolute

minimum amount of time required to simply physically record an answer, aside from the thinking time.

12. Once the test is given and graded, analyze the items and student responses for clues about well written and poorly written items as well as problems in understanding of instruction.

More specific ideas about:

• the blueprint (step 2)

Don't make it overly detailed. It's best to identify major ideas and skills rather than specific details.

Use a cognitive taxonomy that is most appropriate to your discipline, including non-specific skills like communication skills or graphic skills or computational skills if such are important to your evaluation of the answer.

Weigh the appropriateness of the distribution of checks against the students' level, the importance of the test, the amount of time available. Obviously one can have more low level questions in a given time period, for example.

A blueprint looks something like this:

CON-CEPT	COGNITIVE LEVEL			
	BASIC FACTS	APPLI-CATION	SYNTH-ESIS	ANALYSIS/EVALUATION
Steps in test design	x	x		x
item types	x	x		
errors in items		x		x
item analysis		x		x

• the alternative item types (step 3)

The following array shows the most common question types used at various cognitive levels. It is possible to test almost all levels with all types, but some are simply more efficient than others.

Factual Knowledge	Application	Analysis and Evaluation
Multiple Choice True/False Matching Completion Short Answer	Multiple Choice Short Answer Problems Essay	Multiple Choice Essay

• the item cards (step 4)

These can be helpful in making up alternative forms of a test since each item idea on a given card should test the same content at the same level, but in potentially different ways.

Once the test is over, these same cards can be used to keep a record of which ideas were used and how the students responded. This practice aids in producing more effective future exams.

• writing different item types (step 5)

The next few sections describe some suggestions to keep in mind in the construction of various item types.

• *essay questions*

1. Be sure the task is clearly defined and you give the students some idea of the scope and direction you intended for the answer to take. It helps to start the question with a description of the required behavior (e.g. the verb such as "compare" or "analyze") to put them in the correct mindframe to think about the rest of the question.
2. Write the question at a linguistic level appropriate to the students.
3. Construct questions that require a student to demonstrate command of background information, but are not a simple repeat of that information.
4. If you're going to ask questions which call for opinions or attitudes from the students, be sure that

the emphasis is not on the opinion but on the way it is presented and argued.

5. Use a larger number of shorter, more specific questions rather than one or two longer questions so that more information can be sampled.
6. The use of optional questions results in different exams for different students and makes comparative grading more difficult. They do, however, often reduce student anxiety. Both effects must be weighed when considering their use.

Some additional suggestions aside from constructing the item itself:

1. Give the students a "key word" list containing the verbs you most commonly use in your essay questions and a short description of what you look for in questions using that verb.
2. Give the students a pair of sample answers to a question of the type you will give. Indicate why one is good, one bad; be specific.
3. In grading, sketch out a grading scheme for each question before reading the papers OR randomly select a few papers, read the answers and make up the grading scheme based on the range represented by those answers.
4. Detach identifying information from a paper and use code numbers instead to avoid letting personality factors influence you.
5. Grade one question at a time across all papers before moving to the next question.
6. After grading all the papers on one item, reread the first few papers over to be sure that you have maintained consistent standards.
7. Scramble the order of the papers between grading successive questions.
8. Be clear with yourself and the students the extent to which factors other than content (such as grammar, handwriting, etc.) will influence the grade.

• *multiple choice questions*

1. Write the item stem first.
 - a. Be sure the stem asks a clear question.
 - b. Stems phrased as questions are usually easier to write.

- c. Stems should not contain a lot of irrelevant information.
 - d. Reading level should be kept appropriate to the students.
 - e. Be sure the stem is grammatically correct.
 - f. Try to avoid negatively stated stems.
2. Write the correct response next.
 - a. Be sure it is correct.
 - b. Keep the reading level of the stem and the response consistent.
 - c. Avoid too many qualifiers.
 - d. Avoid grammatical clues like articles and tense.
 - e. Avoid using stereotyped responses as the correct response (e.g. catch phrases or very familiar terms).
 - f. Assign the response to a random position in the answer sequence right away.
 3. Read the stem and the correct response together to be sure they sound right.
 4. Assign the correct response to a random position in the response list.
 5. Generate the distractors.
 - a. Create a "truth table" to assist in identifying plausible distractors. For example, some tables for different items might look like this:

For a cause and effect item -			
Relationship is:			
		<i>true</i>	<i>false</i>
<i>true</i>	correct answer		effect is true but unrelated to cause
<u>effect is</u>	<i>false</i>	effect is false, but related	effect is a false description and unrelated
<u>Example:</u> What is the effect of releasing a ball in positive gravity?			
	a) it will fall "down" (correct)		
	b) it will retain its mass (true but unrelated)		
	c) it will rise (false but related)		
	d) its shape will change (false and unrelated)		

- b. Parallel the construction of the correct response.

- c. Keep the statements simple.
 - d. Do use stereotyped responses for distractors.
 - e. Use common misconceptions.
 - f. Make the distractors different enough from one another to be discriminable.
 - g. Avoid obscure words or highly implausible distractors.
 - h. Be cautious in the use of "all" or "none" of the above.
 - i. Put the distractors in logical order.
 - j. Put words which occur in all the responses in the stem instead.
6. Read the stem and the distractors together to be sure they sound right together.

• matching question

Think of this type of question as a more efficient multiple choice item.

1. Only homogeneous premises (list on the left to be answered) and homogeneous responses (list on the right consisting of possible answers) should be grouped in one item. For example, in a question on common and scientific names of plants, don't include animal names in the lists. Doing so in effect cuts the list of choices by however many non-plant names there are.
2. Relatively short lists of responses should be used.
3. Premises should be arranged for maximum clarity and convenience.
4. Response options should be arranged alphabetically or numerically.
5. Directions should clearly indicate the basis for matching.
6. Position of matches should be varied.
7. All of the choices of each matching set should be on one page.
8. More responses than premises should be used in a set or a single response should be used to answer more than one premise.

• completion question

1. Only significant words should be omitted in incomplete statement items.

2. When omitting words, enough clues should be left so that the student who knows the answer can supply the correct response.
3. Grammatical clues to the correct answer should be avoided.
4. Blanks should occur at the end of the statement, if possible.
5. Limit the length of the responses to single words or short phrases.
6. Verbatim quotes from the text should be avoided.
7. Be clear to yourself and the students the level of specificity required in this type of question.

• true/false question

1. Each statement should be clearly true or clearly false.
2. Trivial details should not make a statement false.
3. The statement should be concise without more elaboration than necessary.
4. Exact statements should not be quoted from the text.
5. Use quantitative terms as opposed to qualitative terms if possible.
6. Specific determiners (always, never, etc.) which give a clue to the answer should be avoided.
7. Negative statements are often confusing and should be minimized.
8. When a controversial statement is used, authority should be quoted.
9. A pattern of answers should be avoided.

doing an item analyses (step 2)

The purpose of the item analysis is to help you identify questions that can differentiate between students who "know" the material or are "good" based on some other criterion measure and those who don't "know" the material or aren't "good." It also helps identify poorly written items which mislead students.

The basic theory requires that the class be divided into identifiable groups representing the best and worst performers on some measure. One possible measure is the overall score on the test itself. Theoretically students

who get high scores "know" the material while students who get low scores don't. Therefore you can divide the class into thirds or quarters on the basis of overall test score and evaluate the performance of the two extreme groups on each question.

You can use the same idea, but some other criterion on which to split the class if there is some other level of performance you're interested in predicting. For example, you should use lab grades or subsequent test grades to divide the class into extremes. The results of the item analysis would then tell you how well each item on the test predicted or correlated with performance on the lab or the subsequent unit.

The three pieces of information which are used to evaluate the item are:

- a) the difficulty level - what overall percentage of the students in the analysis groups answered correctly? It varies from 0.0 to +1.0 with numbers approaching +1.0 indicating more students answering correctly.
- b) the discrimination level - How well does this item discriminate between the top students and the bottom students? It varies from -1.0 to +1.0. The closer the number is to either end, the better that item is in differentiating between those who know the material and those who don't. The only difference is that when the number is positive, the items will be answered correctly by the good students and incorrectly by the poor students; when it is negative, the reverse is true. The formula is:

$\frac{H - L}{N}$	H = number of top students correct
	L = number of bottom correct
	N = number of students in either top or bottom (should be equal)
- c) the breakdown - How do the students in each group who answered incorrectly actually answer the question? This is done by tallying the choices made by those who miss the question and looking for particularly popular incorrect answers, especially among the top students. If one distractor is favored, the instructor should make sure it is indeed incorrect.