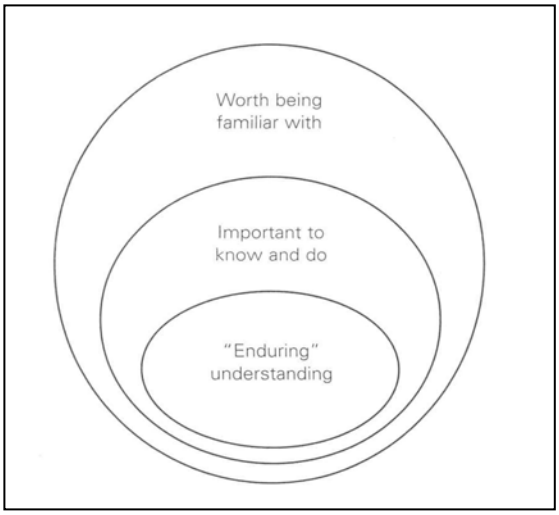
Writing Better Objective Tests
Bill Cerbin
UW-La Crosse, Center for Advancing Teaching & Learning

Prioritize the subject matter
What subject matter—topics, ideas, concepts are

1. Essential
2. Important
3. Worth being familiar with



Use items that measure the course/unit learning objectives. What are the major objectives of the unit or course that should be measured by the test?

Create a Test Plan (*Table of Specifications*) that classifies each item according to what topic or concept it tests AND what objective it addresses. The table can help you write a test that has *content validity*—there is a match between what was taught and what is tested. The table helps insure that you

1. emphasize the same content you emphasized in day-to-day instruction (e.g., more items about topic X and fewer about topic Y because you consider X to be more important and you spent more time on X)
2. align test items with learning objectives (e.g., important topics might include items that test interpretation, application, prediction, and unimportant topics might be tested only with simpler recognition items)
3. do not overlook or underemphasize an area of content

Test Plan (Table of Specifications) Format

Subject Matter concepts, topics, ideas	Learning Objectives				
	Knowledge (recall)	Analyze	Apply	Interpret	Total
Topic A					
Topic B					
Topic C					
Topic D					
Etc.					

Write matching pairs of items after each class. While the class period is still fresh write 2-3 items that focus on the most important material from the class. Ideally, write matching pairs of items, i.e., two items that test the same thing. Use one for formative evaluation (e.g., non-credit quiz or review) and the other one for summative evaluation, i.e., the test).

The question stem

1. There are two acceptable forms of stems a) one that poses a complete, direct question and b) one that is a statement that needs to be completed.
   - Direct question: What color results from the mixture of equal parts of yellow and blue paint?
   - Completion: The actual election of the U.S. president to office is done by

2. Put all relevant information in the stem and make it sufficiently specific to stand on its own without qualification.
3. Also try to make the stem short and sharply focused.
4. For questions that ask for recall of information do not introduce information into the stem that has not been part of the material covered in class.
5. Avoid negative wording because it can confuse students.
   - Bad: Which of the following is not a characteristic of Brutalism?
   - Better: Which of the following best distinguishes Brutalism from other architectural movement?

Correct/Best answer

1. Write the best answer first, then response alternatives
2. Be sure there is only one "best" answer

Response alternatives (AKA response options or distractors)

1. Should be brief and grammatically parallel. Make response alternative grammatically consistent with the stem. An example of a grammatically inconsistent item:

   The functions of the Federal Reserve are to provide the nation with an elastic money supply and to
   A. help stabilize the economy
   B. correction of national income statistics
   C. correction of tax laws
   D. help levy property taxes

2. Make responses roughly equal in length. FYI—Instructors tend to make the best answer longer.

3. Eliminate repetitive information from response alternatives

Bad: Between 1950 and 1965
A. Interest rates increased
B. Interest rates decreased
C. Interest rates fluctuated greatly
D. Interest rates did not change

Better: What was the trend in interest rates between 1950 and 1965?
A. Increased only
B. Decreased only
C. Increased, then decreased
D. Remained unchanged

4. Take special care or avoid using no-exception words such as "never," "all," "none," and "always." These words tend to give clues about the correct answer. Few statements are absolute or universally true.

5. Avoid using "all of the above" as an alternative. If the student can identify one alternative that is not true then it eliminates that alternative plus "all of the above," making it easier to guess.

6. Vary the position of the best answer. Instructors tend to put the best answer in the B or C position.

7. Vary the number of response options as appropriate.
    A. It is acceptable practice to vary the number of response alternatives on a test.
    B. There is no single best number of response alternatives. Research indicates that 3 alternatives is just as effective as four.
    C. If appropriate, vary the number of response alternatives within the same test.

8. Avoid using "all of the above" or none of the above" just to fill space
   An alternative to the *all of* format is to ask students <u>to find all correct answers</u> to an item.
   For example, "Acceptable practices for response alternatives on multiple choice tests include"
        A. There is no single best number or response alternatives for multiple choice items
        B. Avoid using *all of the above* and *none of the above* as response alternatives
        C. Response alternatives should be plausible answers
        D. Response alternatives for a test item should be roughly equal in length
        <u>Caveat</u>: Scantron can't handle multiple correct responses

9. If you have difficulty creating enough plausible incorrect answers for an item, put it on the test as a fill-in item and students will produce a range of incorrect responses that you can use next time.

Overall test construction

1. Test Instructions.
    - Instruct students to select the "best answer" and acknowledge that sometimes the response options have elements of accuracy, but there is a best answer for each item. Using the "correct answer" invites arguments from students that their answers are correct as well.
    - State whether there are rewards or penalties for *guessing.*
        - Types of advantages: might get additional credit; instructor allows students to justify their guesses and gives partial credit.
        - Type of penalty: Instructor deducts additional points for incorrect answers.
        - Consider using: "I don't know" as a response alternative.
2. Include a few easy items at the beginning of the test. You can help to reduce test anxiety by including a few easy items to start the test.
3. Group test items related to the same topics. Items related to a specific topic should be grouped together on the test. This allows students to think about each topic and section of the material rather than jump from topic to topic. Consider using labels or headings to indicate topics and then group related items under the headings.
4. Test length. If you want to test what students know, as opposed to how fast they can read and respond, be sure to give sufficient time to complete the test. Otherwise, you penalize poor readers, slow readers, deliberate and reflective responding. Students can complete 1-2 multiple choice items per minute.
5. Time limits. Students with certain forms of learning disabilities need longer to complete a test. Students whose first language is not English will also need longer. While you can likely accommodate such students individually, consider carefully what educational goal would lead you to impose a time limit on a test at all. Another class following into your room is not a good

educational reason to refuse to allow a student to finish a test. Numerous short tests or quizzes work better for most students than big, high-stakes tests.

6. Test-taking accommodations: Students with disabilities must register with Disability Resource Services (DRS) to qualify for accommodations. DRS will let you know what kinds of accommodations a student needs for test-taking. In addition to more time, another common need is for a separate, *quiet* room. Hallways are poor choices because of distractions.

Improving items using feedback and item analysis

1. Feedback from fellow instructors
2. Feedback from students—Review test results with students; opportunity to give feedback to students and get feedback from them about test items
3. Item Analysis. If you have tests scored by IT you can request an item analysis that provides item difficulty and item discrimination. IT has a handout that helps you interpret the analysis. Ask for *Revised Test Scoring Display and Item Analysis*.
   A. item difficulty-- Percentage of students who correctly answer item.
   Desirable difficulty levels for different types of items

Question Type	Item Difficulty % correct
5 response options	60%
4 response options	62%
3 response options	66%
True/False	75%

Review items at the extremes—very easy, >90% correct, or difficult, <20% correct.

B. item discrimination—Relationship between how well students do on item and how well they do overall on the test. High discrimination is good—it means that students who do well on the test tend to get the item correct and those who do poorly on the test tend to get it incorrect

How Well Item Discriminates	Discrimination Value
Very good	>.40
Good	.30-.39
Fairly good	.20-29
Poor	<.20

High discrimination (>.40): Students with high test scores responded correctly and students with low scores responded incorrectly.

Very low discrimination (<.19): Students with high test scores do poorly on the item, and low scoring students may do better.

Examples of Different Types of Multiple Choice Items

Multiple choice questions ask students to discriminate among different plausible options and select the best answer. Typically, these items are written at a basic level and students can answer the question based on recognition or recall memory. There is nothing wrong with items that test for basic knowledge or familiarity. But, what if you want to test more complex objectives such as understanding, problem solving, or analytical reasoning?

To test complex learning objectives you can either 1) use different forms of items such as essay, short answer, performance tasks) or 2) develop multiple choice items that address the objectives.

Multiple choice questions that measure complex learning objectives. It is challenging to write multiple choice questions that test complex learning. Typically these items present the student with a scenario, passage, or graphic representation and then pose several questions related to the material. These can be used to assess complex thinking abilities such as prediction, interpretation, application, or evaluation. Here are some alternative types of multiple choice items (Davis, 2009, pp. 391-393).

Prediction question. In the stem of the question present a problem or situation and ask students to predict an outcome.

2. A large city is investigating the elimination of rent controls on housing at a time when the vacancy rate is extremely low—only 1 percent. Which of the following is most likely to occur if rent controls are eliminated?

A. An increase in the demand for housing, followed by a decrease in the supply of housing.

B. An increase in rents, followed by an increase in the supply of housing.

C. A decrease in rents and a decrease in the supply of housing.

D. No change in rents because price controls are usually set where supply and demand intersect.

Identify a principle or theory question. In the stem of the question present an example of some phenomenon and ask students to identify a principle or theory that it illustrates.

3. Because of rapidly rising national defense expenditures, the country of Parador will experience price inflation unless measures are taken to restrict the growth of aggregate private demand. If Parador wishes to minimize the adverse effects of anti-inflationary policies on economic growth, it should implement

A. A tight monetary policy because that would restrict consumption expenditures more than investment.

B. A tight monetary policy because that would restrict consumption expenditures.

C. An increase in personal income taxes because that would restrict consumption expenditures more than investment.

D. Either a tight monetary policy or an increase in personal income taxes because both depress investment equally.

Assertion-reason questions (ARQ). The stem of these questions includes "two statements—an assertion and a reason—linked by *because*" (Davis, 2009, p. 393)

(*Assertion*) In a small open economy, if the prevailing world price of a good is lower than the domestic price, the quantity supplied by the domestic producer will be greater than the domestic quantity demanded, increasing domestic producer surplus.

Because

(*Reason*) In a small open economy any surplus in the domestic market will be absorbed by the rest of the world. This increases domestic consumer surplus.

A. The assertion and reason are both correct, and the reason is valid.

B. The assertion and reason are both correct, but the reason is invalid.

C. The assertion is correct, but the reason is incorrect.

D. The assertion is incorrect, but the reason is correct.

E. Both the assertion and the reason are incorrect.

You-are-the-teacher-question. The stem of the question asks the student to assume the role of teacher and then evaluate short written passages (i.e., flawed short test answers).

Pretend you are a science teacher who is correcting the following answer on a quiz. How many scientific errors does the answer contain? *Note:* There is a maximum of one error per sentence.

During the depolarization phase of an action potential, sodium gates are open and sodium diffuses from the extra-cellular fluid to the intra-cellular fluid. At the end of the depolarization phase, sodium gates close and potassium gates open. Repolarization begins when potassium moves by active transport from the intra-cellular fluid to the extra-cellular fluid of the cell. After the action potential passes, ion gradients are maintained by the sodium/potassium pump.

A. 0 errors

B. 1 error

C. 2 errors

D. 3 errors

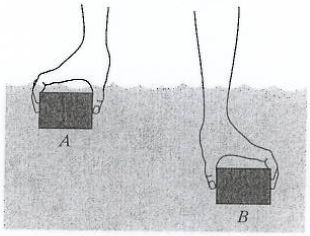
E. 4 errors

Multiple choice + short written justification. If you want to know the reasoning behind their answers ask students to write a short justification of why they selected their specific choice.

ConcepTests. Harvard physicist, Eric Mazur, uses ConcepTests to evaluate students' conceptual understanding. Typically, the stem of the item includes a diagram, picture or description of a situation that illustrates a physics concept, and asks the student to predict an outcome. Below are some examples from Mazur, E. (1997). *Peer instruction: A user's manual*. Prentice Hall, Upper Saddle River, NJ.

BUOYANCY

Imagine holding two identical bricks under water. Brick *A* is just beneath the surface of the water, while brick *B* is at a greater depth. The force needed to hold brick *B* in place is

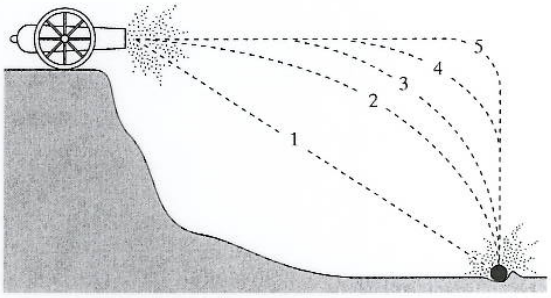


1. larger than
2. the same as
3. smaller than

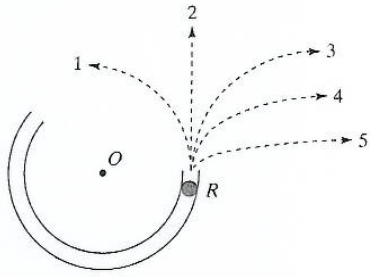
the force required to hold brick *A* in place.

Figure 2.1 *ConcepTest* question on Archimedes' principle. For an incompressible fluid such as water, the second choice is correct.

12. A ball is fired by a cannon from the top of a cliff as shown below. Which of the paths 1–5 would the cannon ball most closely follow?



6. Which of the paths 1–5 below would the ball most closely follow after it exits the channel at *R* and moves across the frictionless table top?



Handout References

Slavin, R. E. (2000). *Educational psychology: Theory and practice, 6th edition*. Allyn and Bacon: Boston, MA.

Davis, B. G. (2009). *Tools for* teaching, 2nd edition. Jossey Bass Publisher: San Francisco.

Stiggins, R. J. (1994). *Student-Centered Classroom Assessment*. Macmillan Publisher: NY.

Wiggins, G. (1998). Understanding by design. Association fo Supervision and Curriculum Development: Alexandria, VA.

Available from IT, *Revised Test Scoring Display and Item Analysis*.

Online Resource related to writing objective tests

http://testing.byu.edu/info/handbooks/betteritems.pdf
Steven J. Burton, et al. "How to Prepare Better Multiple-Choice Test Items: Guidelines for University Faculty," Brigham Young University Testing Services and The Department of Instructional Science, 1991. All you need in one place. Explains the advantages and limitations of multiple-choice tests, how to decide whether you should use such tests or not, how to write questions that measure more complex types of learning than recall of facts, and provides a variety of formats for multiple-choice questions. Even has a checklist at the end.