Dental Teaching Institute

Assessment: Designing Effective Multiple-choice and Short Answer Questions

Debra Dawson & Michael Atkinson

December, 2006
Assessment is an iterative feedback process for continual program improvement, based on the model shown below.

*Step one* is to define intended program learning objectives: specifically, what do we want our graduates to know and actually to be able to do?

*Step two* is to define measurable outcomes that will serve as evidence of how well each objective has been met, and then actually to measure them. Because this step requires explicit articulation of program success criteria, it often has the added benefit of clarifying faulty assumptions.

*Step three* is to compare actual observed outcomes to intended program objectives: how well did we meet our objectives in general, and our student learning objectives in particular?

Finally, in *step four*, based on how well or how poorly achieved outcomes compare to intended outcomes, elements of the program (including assessment elements) are redesigned as appropriate, and a new assessment cycle begins.

The assessment cycle is an integral part of student-centered education. It provides an ongoing mechanism for challenging tacit assumptions about program effectiveness, identifying conflicting program elements, and assuring that student learning objectives are met. It also allows for evolution of program goals over time. Although it is by no means an easy task to define learning objectives and measurable outcomes for an
educational program, faculty engaged in the process inevitably and uniformly are rewarded by identifying with heightened clarity what it is they are trying to accomplish and how they can better go about it.

Good assessment practice is based on a number of assumptions:

1. The first precept of good assessment practice is to assess what is most important;
2. Anything that can be taught or learned can be assessed;
3. Assessment should be applied at course, program, and institutional levels;
4. Every program and every course should be organized around clearly articulated learning goals and objectives, explicit assessment methods, and measurable outcomes;
5. An assessment process should be logistically feasible and practically manageable to insure that it is regular and ongoing.

What is important varies by program and program level, and program competencies are often sequentially developed; general education competencies tend to be first steps in the development of a range of integrative abilities across the curriculum. In Washington State, four such integrative abilities are being explored in considerable depth to find consistent methods of assessing how well graduates of Washington schools have mastered them: writing, information technology literacy, quantitative reasoning, and critical thinking.

Major Categories in the Taxonomy of Educational Objectives (Bloom 1956)

The following information came from: (http://faculty.washington.edu/krumme/guides/bloom.html)

Categories in the Cognitive Domain: (with Outcome-Illustrating Verbs)

1. Knowledge of terminology; specific facts; ways and means of dealing with specifics (conventions, trends and sequences, classifications and categories, criteria, methodology); universals and abstractions in a field (principles and generalizations, theories and structures):
   Knowledge is (here) defined as the remembering (recalling) of appropriate, previously learned information.
   - defines; describes; enumerates; identifies; labels; lists; matches; names; reads; records; reproduces; selects; states; views.
2. Comprehension: Grasping (understanding) the meaning of informational materials.
   - Classifies; cites; converts; describes; discusses; estimates; explains; generalizes; gives examples; makes sense out of; paraphrases; restates (in own words); summarizes; traces; understands.
3. Application: The use of previously learned information in new and concrete situations to solve problems that have single or best answers.
Acts; administers; articulates; assesses; charts; collects; computes; constructs; contributes; controls; determines; develops; discovers; establishes; extends; implements; includes; informs; instructs; operationalizes; participates; predicts; prepares; preserves; produces; projects; provides; relates; reports; shows; solves; teaches; transfers; uses; utilizes.

4. Analysis: The breaking down of informational materials into their component parts, examining (and trying to understand the organizational structure of) such information to develop divergent conclusions by identifying motives or causes, making inferences, and/or finding evidence to support generalizations.

Breaks down; correlates; diagrams; differentiates; discriminates; distinguishes; focuses; illustrates; infers; limits; outlines; points out; prioritizes; recognizes; separates; subdivides.

5. Synthesis: Creatively or divergently applying prior knowledge and skills to produce a new or original whole.

Adapts; anticipates; categorizes; collaborates; combines; communicates; compares; compiles; composes; contrasts; creates; designs; devises; expresses; facilitates; formulates; generates; incorporates; individualizes; initiates; integrates; intervenes; models; modifies; negotiates; plans; progresses; rearranges; reconstructs; reinforces; reorganizes; revises; structures; substitutes; validates.

6. Evaluation: Judging the value of material based on personal values/opinions, resulting in an end product, with a given purpose, without real right or wrong answers.

Appraises; compares & contrasts; concludes; criticizes; critiques; decides; defends; interprets; judges; justifies; reframes; supports.

---

Other Domains for Educational Objectives:

- Affective Domain (emphasizing feeling and emotion)
- Psychomotor Domain (concerned with motor skills)

---

Internet Resources:

- Learning Domains or Bloom's Taxonomy: The Three Types of Learning [www.nwlink.com/~donclark/hrd/bloom.html]
- Bloom's Taxonomy [www.corrosion-doctors.org/Training/Bloom.htm]

"Following the 1948 Convention of the American Psychological Association, B.S. Bloom took a lead in formulating a classification of "the goals of the educational process". Three "domains" of educational activities were identified...."

- Bloom's Taxonomy [DLRN Technology Resource Guide, Chapter 4; http://rtecexchange.edgateway.net/cs/rtecp/view/rtec_sub/130]
In 1956, Benjamin Bloom headed a group of educational psychologists who developed a classification of levels of intellectual behavior important in learning. This became a taxonomy including three overlapping domains; the cognitive, psychomotor, and affective. Each of the domains can be utilized through the interaction of media.

- **Bloom's Taxonomy** [www.coun.uvic.ca/learn/program/hndouts/bloom.html]

  Benjamin Bloom created this taxonomy for categorizing level of abstraction of questions that commonly occur in educational settings. The taxonomy provides a useful structure in which to categorize test questions…

- **Skills for each of the six levels of Bloom's Taxonomy** [Counselling Services - University of Victoria] [www.coun.uvic.ca/learn/program/hndouts/bloom.html]

- **Bloom et al.'s Taxonomy of the Cognitive Domain** (Bill Huitt) [http://chiron.valdosta.edu/whuitt/col/cogsys/bloom.html]
  - Writing Behavioral Objectives [http://chiron.valdosta.edu/whuitt/col/plan/behobj.html]
  - Critical Thinking [http://chiron.valdosta.edu/whuitt/col/cogsys/critthnk.html]

- **Kizlik, Robert** [http://www.adprima.com/kizlikbio.htm]
  - Examples of Behavioral Verbs and Student Activities [http://www.adprima.com/examples.htm]
  - Definitions of Behavioral Verbs [http://www.adprima.com/verbs.htm]
  - Writing Behavioral Objectives [http://www.ed.psu.edu/pawin/resources/linkssub/Objectives.asp]

- **Overview** [www.ceap.wcu.edu/Houghton/Learner/think/bloomsTaxonomy.html]

- **MCQs and Bloom's Taxonomy** [Cape Town, SA: UCT's page on Designing and Managing Multiple Choice Questions; [www.uct.ac.za/projects/cbe/mcqman/mcqappc.html]

- **Curriculum Development & Evaluation (under construction)** [http://faculty.washington.edu/krumme/guides/taba.html]

- **Other Sites:**
  - Explorations in Learning & Instruction: The Theory Into Practice Database (Taxonomy) [http://tip.psychology.org/taxonomy.html]
  - Bloom Resources [http://www.ulm.edu/%7Ecasey/401intcon.html]
  - http://its.foxvalley.tec.wi.us/iss/curric-assessment/COLUMN.html
  - Guide to Theories of Learning [http://commhum.mccneb.edu/PHILOS/learntheo.htm]
  - Internet Resources for Higher Education Outcomes Assessment [http://www2.acs.ncsu.edu/UPA/assmt/resource.htm]
Literature:

Extensive Online Bloom Bibliography


Francis P Hunkins (UW), Teaching Thinking Through Effective Questioning (1989), and others. "How to write learning outcomes", by Alan Jenkins (Oxford Brookes University) & Dave Unwin (Birkbeck College London) http://ncgia.ucsb.edu/giscc/units/format/outcomes.html

The information below came from :
http://www.park.edu/cetl/quicktips/writinglearningobj.html

Generally, instructors will want to design learning objectives to target a range of levels of student understanding. The phrasing of learning objectives will help guide both instructional activities and assessment, thus instructors should carefully select the emphasis of learning and the relevant verb.
<table>
<thead>
<tr>
<th>Domain</th>
<th>Emphasis</th>
<th>Relevant Verbs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cognitive</td>
<td>Knowledge</td>
<td>Recall, identify, recognize, acquire, distinguish, state, define, name, list, label, reproduce, order</td>
</tr>
<tr>
<td>Cognitive</td>
<td>Comprehension</td>
<td>Translate, extrapolate, convert, interpret, abstract, transform, select, indicate, illustrate, represent, formulate, explain, classify, comprehend</td>
</tr>
<tr>
<td>Cognitive</td>
<td>Application</td>
<td>Apply, sequence, carry out, solve, prepare, operate, generalize, plan, repair, explain, predict, demonstrate, instruct, compute, use, perform, implement, employ, solve</td>
</tr>
<tr>
<td>Cognitive</td>
<td>Analysis</td>
<td>Analyze, estimate, compare, observe, detect, classify, discover, discriminate, explore, distinguish, catalog, investigate, breakdown, order, determine, differentiate, dissect, contrast, examine, interpret</td>
</tr>
<tr>
<td>Cognitive</td>
<td>Synthesis</td>
<td>Write, plan, integrate, formulate, propose, specify, produce, organize, theorize, design, build, systematize, combine, summarize, restate, argue, discuss, derive, relate, generalize, conclude, produce</td>
</tr>
<tr>
<td>Cognitive</td>
<td>Evaluation</td>
<td>Evaluate, verify, assess, test, judge, rank, measure, appraise, select, check, judge, justify, evaluate, determine, support, defend, criticize, weigh, assess</td>
</tr>
<tr>
<td>Affective</td>
<td></td>
<td>Agree, avoid, support, participate, cooperate, praise, help, offer, join</td>
</tr>
<tr>
<td>Psychomotor</td>
<td></td>
<td>Adjust, repair, taste, bend, measure, perform, operate, use, move</td>
</tr>
</tbody>
</table>

**Learning Objectives**

**Review Checklist:**

- [ ] Does the learning objective stem from a course goal?
- [ ] Is the learning objective measurable?
- [ ] Does the learning objective target one specific aspect of expected performance?
- [ ] Is the learning objective student-centered?
- [ ] Does the learning objective utilize an effective, action verb that targets the desired level of performance?
- [ ] Do learning objectives measure a range of educational outcomes?
- [ ] Does the learning objective match instructional activities and assessments?
- [ ] Does the learning objective specify appropriate conditions for performance?
- [ ] Is the learning objective written in terms of observable, behavioral outcomes?

From: [http://www.park.edu/cetl/quicktips/writinglearningobj.html](http://www.park.edu/cetl/quicktips/writinglearningobj.html)
Design Evaluation

From: This next section is adapted from a Design Evaluation Workshop created by Dr. Marilyn Robinson (1992).

Preamble

Once you have decided what you expect the students to be able to do as a result of learning, the next step is to decide how you will assure yourself, and them, that they have reached those objectives. Some think that when evaluation is designed before instruction, teachers will simply teach to the test. However, since test items should directly reflect the objectives of instruction, teaching to the test is justified. The evaluation system ultimately ‘drives’ the curriculum, the course, the specific instruction, and the students' learning. Designing an evaluation that really measures the learning outcomes before you plan instruction, makes it easier to design an appropriate learning strategy. If learning objectives are written in terms of concrete, observable behaviours, the design of evaluation is a relatively easy task.

Overview

We will review some of the fundamentals and techniques of evaluation. In particular, the principles and methods of evaluation of students' performance: how to select and devise measurement tools, administer tests, and interpret the results of evaluation for purposes of guiding your own teaching and the students' further learning, and for certification of students' achievement. Whether you are evaluating knowledge or clinical skills, the principles are the same - the tools of evaluation are different.

C. Learning Objectives

1. You will be able to distinguish between formative and summative evaluation and describe the purposes and characteristics of each.

2. You will be able to discriminate between test reliability and test validity and identify ways of determining each.

3. You will be able to describe methods of evaluation of learning and choose appropriate assessment methods for the learning objectives of your course.

4. You will be able to design the evaluation system for your course.

D. Evaluation

You will develop a test blueprint for your course and final examination (or examples of test items).
Introduction

Because evaluation ‘drives’ the instructional system, design of evaluation is one of the most important processes involved in formal instruction. Nevertheless, it is usually the most poorly done. This is particularly true when instructors do not have explicitly stated learning objectives because under these circumstances the task is very difficult and time consuming. In addition to evaluating students, it is important to evaluate the effectiveness of the teaching/learning strategy including the reliability and validity of the test measurements themselves. It is much more comfortable to evaluate students if we know that the evaluation system is accurate and that it measures knowledge, skills, attitudes that are valuable according to the real world assessment. It is also more comfortable if we know that our instructional strategies are sound and that they really help students to achieve the learning objectives.

Evaluation

Evaluation is a systematic procedure for measuring a sample of a student's behaviour in order to measure that behaviour against standards and norms. Good evaluations reflect the intent of the instruction - they reflect precisely the objectives of the course and, as such, they reflect as closely as possible how the students will perform in the real world.

Formative Evaluation

Formative evaluation is the process of assessing, during a course, how well the students and the course are doing. The objective of this type of evaluation is to allow
students and instructors to make changes that will improve the learning experience and ensure that students are able to meet the learning objectives.

**Summative Evaluation**
- Is the process of assessing the degree to which students have achieved the learning objectives.
- Of the course assesses the extent to which the instruction was successful in ensuring that students were able to meet the learning objectives.
- Of students is always comparative - student performance is compared either to other students' performance (*norm-referenced*) or to some pre-determined absolute standard (*criterion-referenced*).

**Norm-Referenced vs Criterion-Referenced Tests**

All summative evaluation measures student performance in a relative sense. *Norm-referenced tests* are those that use the performance of other people on the same test as the basis for interpreting an individual student's relative test performance. *Criterion-referenced tests* are those that measure an individual's ability with respect to some criterion, or standard which has been determined in advance by knowledgeable people in the field of study. In norm-referenced grading, the students set the standards. In criterion-referenced grading, the instructors set the standards.

**Norm-Referenced Tests**

A norm-referenced measure allows us to compare one individual with others. Norm-referenced grading was designed to measure students' performance relative to a very large population. In the classic method, the test scores are fit to a standard curve. However, experienced instructors can plan their test to yield the same outcome without using the statistical method of curve fit. They simply include questions which they ‘know’ only the top 10%, 15%, 20%, etc. can answer. If the purpose of including these ‘difficult’ questions is to lower the test score average or to differentiate between students, then the test is norm-referenced. Norm-referenced test scores can tell us that one student knows more about the topic than another student but unfortunately, the scores by themselves tell us nothing about what or how much the two students know in an absolute sense. If the purpose of the test is to choose the best students for a limited number of places in a program, then norm-referenced tests are the best choice. **They are not, however, appropriate for grading within a professional school.**

**Applications**
- classifying students
- selecting students for fixed quota requirements
- making decisions as to how much (more or less) a student has learned in comparison to others
**Criticisms**

- The final grade conceals the student's misunderstandings, and possible limitations. For any interpretation to have meaning, each student's score needs to be related to the content of the test.
- Any given mark does not signify an actual amount of knowledge.
- Students are competitive rather than cooperative since their score is improved if others do more poorly.
- Norm-referenced tests do not take into account the possibility that a given group of students may all be competent (or incompetent).
- Students may suffer diminished motivation. Even though they master everything they are supposed to, they may never get an ‘A’ unless they are in the very top group of students.
- Norm-referenced tests conceal that fact some courses are very good while others are very bad or that different teachers set different standards.

**Criterion-Referenced Tests**

If the purpose of the test is to be sure that every competent student is able to practice dentistry but no incompetent student is allowed to practice, then a *criterion-referenced test* is best. The criterion-referenced test is deliberately constructed to give information that is directly interpretable in terms of absolute criterion of performance. The absolute criterion is directly related to the essential performance level for each essential learning objective. The teacher or curriculum committee uses their knowledge and experience with the subject area to determine the essential learning objectives and required performance level. Usually criterion-referenced tests are graded with a pass-fail system. Teachers use their experience with past student test performance and ability to perform in the *real world* (or in future courses that are dependent on these abilities) to determine the test score for a passing grade. The passing score is often much higher than in norm-referenced tests since fewer questions are asked that are not directly related to essential knowledge, skills or attitudes. Air-line pilots take criterion-referenced tests for licensure.

**Applications**

- Evaluating individualized learning programs.
- Diagnosing student difficulties.
- Estimating a student's ability in a particular area.
- Measuring what a student has learned.
- For certification of competence.
- To control entry to successive units of instruction.
- Whenever mastery of a subject or skill is of prime concern.
- Whenever quota-free selection is used.
- To encourage student cooperation rather than competition.
**Criticisms**

- Does not provide information about a student's ability in comparison to others in the same pool.
- Requires the instructor to identify the essential learning objectives and describe them in measurable terms.
- Requires the instructor to determine the level of test performance which constitutes a pass.
- The responsibility for accurate determination of the criterion lies with the instructor or curriculum committee.

**Purposes of Evaluation**

Evaluation may serve many purposes. Some purposes are best served by *formative evaluation* and others served only by *summative evaluation*. Some purposes are best served by *norm-referenced tests* and other by *criterion-referenced tests*. Examine the list below and consider which purposes are best served by: (1) formative and/or summative evaluations, and (2) norm-referenced and/or criterion-referenced tests.

1. Selection of students for entry into a program.
2. To determine students' entry level in a course.
3. To assess students' progress; to diagnose learning difficulties; to guide further teaching/learning needs.
4. To measure and certify students' achievement of a recognized standard.
5. To motivate students to learn.
6. To encourage students to value the regular evaluation of their own performance as part of their ‘lifestyle’.
7. To determine the effectiveness of a course so that the instructor may make modifications during the course and/or assess the value of a course when it has been completed.
8. To licence students to practice a profession.
9. To communicate to students what material is important.
10. To identify areas of deficiency in need of remedial work.

**Test Reliability and Validity**

Tests are absolutely accurate at measuring the students' performance on the test. As instructors or curriculum administrators, we want to use the students' performance on the test as a measure of their ability in the content area and ultimately as a measure of their ability to perform well in the real world. In order to be able to do this the test must be both *reliable* and *valid*.

**Reliability**

Reliability refers to the extent to which a test score is precise, consistent, or stable. If a test is reliable, the student will get approximately the same score on a similar
test that is taken within a short period of time. In the case of norm-referenced tests, students' scores should be in approximately the same rank order on the two similar tests. The correlation coefficient between the scores of the two tests should be .80 or better in order to consider that the test is reliable. In the university setting, most tests are remarkably unreliable - the error of measurement is often as large as the standard deviation of the students' scores. The most common causes of low reliability are subjective measurement and too small a sample size (not enough questions, simulations, standardized patients, etc.).

**Validity**

Validity refers to the degree to which a test, and our interpretation of it, measures what it was intended to measure. Validity is the single most important issue to consider when evaluating a test. If a test asks questions that could be answered on the basis of general intelligence, test wiseness, etc., then the test may be measuring something besides the student's ability with respect to the learning objectives.

Tests need to have both content validity and construct validity. Content validity refers to the degree to which a test measures what the students were intended to learn. Construct validity refers to the degree to which a test measures the cognitive or affective level at which the students were expected to operate - eg. synthesis, evaluation, commitment to new values, etc.. If a test is not reliable, it cannot be valid. Excellent reliability, however, does not ensure validity. Every conscientious instructor must do battle with the problems of reliability and validity whenever they are designing a test or an entire evaluation system for a course.

**Reliability and Validity of Methods of Assessment**

**Oral Examinations**

*Description:* Examiners use hypothetical case materials with specific lines of questioning and marking criteria; discuss cases with examinees.

*Reliability:* Low. Requires many cases and raters to yield reproducible scores.

*Validity:* Often poor because of inadequate case sampling and poor inter-rater agreement. Unclear what skills are measured.

*Utility:* Examiner time is expensive, difficult to give equivalent examination to all examinees. Popular as a rite of passage.

**Practical Examinations (Standardized Patients, etc.)**

*Description:* Trained patients present the same cases to all examinees; exam usually focused on history, physical examination, diagnosis, communication skills.

*Reliability:* Requires many cases to yield reproducible scores. If patients are ‘real’, then examinees have different examinations.
Validity: Probably high for assessing hands-on skills with patients. Potential problems with response-style artifacts, practice effects, performance rate, and rater bias.
Utility: Expensive and logistically difficult to administer if standardized patients are used. Best available option for assessing hands-on skills and professional attitudes with patients.

Application-Oriented Short Answer/Essay Question Examinations

Description: Description of a problem/clinical situation with probes for solutions of problems.
Reliability: Excellent if many questions are used, marking criteria are carefully defined, and multiple scorers are involved.
Validity: Excellent for assessment of cognitive skills without cuing if questions are explicit.
Utility: Good. Costs are fairly low relative to benefits. Large amount of marker time can be required.

Written and/or Computer-Based Clinical Simulation Examinations

Description: Many different formats but most involve a series of ‘scenes’, in which examinees gather clinical data and initiate treatment. Computer simulations of clinical environment and patient findings superior to written.
Reliability: Low inter-case reliability results in the need for long tests. Developing an answer/scoring key is difficult.
Validity: Generally not good because of response style artifacts and scoring problems.
Utility: Low for summative evaluation purposes, but may be useful for formative feedback. Expensive to develop and require extensive testing time.

Higher-Order Multiple-Choice Question Examinations

Description: Description of a problem/clinical situation; probes for solutions.
Reliability: Excellent. Many situations can be sampled in a short amount of testing time. Objective scoring.
Validity: Excellent for assessment of cognitive skills (can test application, analysis, synthesis and evaluation). Not useful for assessing hands-on clinical skills or professional attitudes.
Utility: Very high. Costs are very low relative to benefits. Instructor time is taken in developing questions but computer scored.

Recall Multiple-Choice Question Examinations

Description: Tests knowledge of isolated facts.
Reliability: Excellent. Many topic areas can be sampled in a short amount of testing time.
Validity: Excellent for assessment of facts, but there is little reason to measure knowledge of facts in isolation. Focus should be on use of knowledge.

Utility: Costs are very low, but higher-order MCQs have the same advantage and are more appropriate. Could be useful for formative feedback.

Designing an Examination

What Should be Tested?

· exam content and construct (cognitive, affective or psychomotor level) should match the instructional objectives
· important topics should be weighted more heavily than less important topics
  · the number of items devoted to each topic should reflect their relative importance
  · the sample of items should be representative of the instructional content and construct and should be as large as practicable

What Assessment Method Should You Use?

This decision will depend on the instructional objectives (what test situation will allow you to observe the measurable behaviour?) and on the purpose of the test. It will also depend on the resources available and the abilities of the instructor.

· oral exams
· practical exams (standardized patients)
· observational reports/in-training assessments
· Essay exams
· course work, assignments, projects
· small group work, participation
· laboratory products
· Short answer, fill-in-the-blanks exams
· Written or computer-based simulations
· True/false, matching exams
· Objective structured clinical evaluations
· Knowledge/comprehension multiple-choice question exams
· Application/analysis/synthesis/evaluation multiple-choice question exams

Length, Difficulty, Clarity

· You need a sufficiently large sample size to provide reliability.
· Do not make the exam so long (3 hours as a maximum may be too long for many students) that you are only testing endurance.
· Do not have too many questions per hour - you should be testing what they have learned rather than ability to write quickly. If students leave the exam early it does not necessarily mean that the test was too easy.
If you can include questions of various difficulty on the same topic, the examination will have more diagnostic potential.

Have a colleague or senior student examine your test items for ambiguity and after the test ask the students to identify questions that they though were unclear.

**Additional Key Issues in Designing Evaluation**

- Match between instructional objectives and test content.
- Emphasis on recall or higher-order cognitive abilities.
- Importance of affective and psychomotor abilities.
- Use of separate or integrated testing of related material (related material from more than one course tested together?)
- Resources available for development, administration, scoring.
- Decisions re timing; frequency of testing; pre-distribution of sample test items; in-class vs take-home exams; open vs closed book exams, etc.
- Policies regarding security and availability of old exams - the greatest value of an examination for the student is to get it back and see what remedial action is needed.
- Value of course work, assignments vs examinations for overall grade.
- Value of process (participation, etc.) vs product for overall grade.
- Reporting of grades: confidentiality; immediacy (the sooner the better); percentage, letter grade, pass/fail.
TEST BLUEPRINTS

The sample Test Blueprint below is for a midterm examination. The test covers five different content areas, each subdivided into several main topics. The central area of the blueprint reflects the number of questions for each topic and indicates the type of knowledge tested according to Bloom’s Taxonomy (Factual, Application, Analysis/Synthesis, or Integrative). This section of the blueprint can also be used to as a reminder for the type of question to be constructed (for example, Type 6). The final column gives an overall weighting (number of questions and percentage of the test) devoted to each content area and subdivided category. Each section is given a particular weighting that should reflect the emphasis that it has received in the classroom. This test blueprint “helps ensure that all required course topics are covered at an appropriate level of understanding.” (Marini 2000)

<table>
<thead>
<tr>
<th>Bloom’s Taxonomy</th>
<th>Content Area</th>
<th>Factual</th>
<th>Application</th>
<th>Analysis /Synthesis</th>
<th>Integrative</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Psychology</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a) As Science</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>3 (4%)</td>
</tr>
<tr>
<td>b) Methods</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>4 (5%)</td>
</tr>
<tr>
<td>c) History</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>3 (4%)</td>
</tr>
<tr>
<td></td>
<td>10 (13%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Biological Basis</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a) Human Evolution</td>
<td>2</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>5 (7%)</td>
</tr>
<tr>
<td>b) Genetic</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>5 (7%)</td>
</tr>
<tr>
<td>c) Genetics and Behavior</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>5 (7%)</td>
</tr>
<tr>
<td></td>
<td>15 (20%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>The Brain</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a) Stem</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>2 (3%)</td>
</tr>
<tr>
<td>b) Limbic System</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>2 (3%)</td>
</tr>
<tr>
<td>c) Cerebrum</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>5 (7%)</td>
</tr>
<tr>
<td>d) Electrical</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>3 (4%)</td>
</tr>
<tr>
<td></td>
<td>12 (16%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Perception</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a) Hearing</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>4 (5%)</td>
</tr>
<tr>
<td>b) Vision</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>7 (9%)</td>
</tr>
<tr>
<td>c) Taste and Smell</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>4 (5%)</td>
</tr>
<tr>
<td>d) Somatic Senses</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>5 (7%)</td>
</tr>
<tr>
<td></td>
<td>20 (27%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Learning</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a) Classical</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>3 (4%)</td>
</tr>
<tr>
<td>b) Operant</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>3 (4%)</td>
</tr>
<tr>
<td>c) Social</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>6 (8%)</td>
</tr>
<tr>
<td>d) Complex</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>6 (8%)</td>
</tr>
<tr>
<td></td>
<td>18 (24%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>TOTAL</td>
<td>23 (31%)</td>
<td>31 (41%)</td>
<td>9 (12%)</td>
<td>12 (16%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>75 (100%)</td>
</tr>
</tbody>
</table>
Test Blueprint

http://www.edtech.vt.edu/edtech/id/assess/blueprint.html

Another example of a test blueprint can be found on the website listed above.

The test blueprint should list your goals and objectives in the left-hand columns, and the outcome behaviors you are using across the top row. In the example, goal one and its two objectives represent "knowledge" behaviors or outcomes. The task, then, is to identify assessment items that help us measure those outcomes. We have selected supplied response, matching, and true/false questions. Each of these item types is useful for measuring lower-order student "knowledge." Since goal one seeks lower-order outcomes, we weight these questions lower than goals three through four which represent higher-order outcomes. Continuing with the example, goal four represents an "evaluation" behavior or outcome. We select the interpretive exercise as an appropriate assessment type to measure these higher-order student abilities.

Designing Multiple-choice Examinations

Some of the information contained in here is also from the National Board of Medical Examiners, Constructing Written Test Questions for the Basic and Clinical Sciences, 3rd Edition: http://www.nbme.org/PDF/ItemWriting_2003/2003IWGwhole.pdf
When info is from the NBME this will be indicated in the text.

Introduction

Multiple-choice examinations have been much maligned by both faculty and students. Faculty complain that MCQ only test the recognition of information at the knowledge level and students complain that MCQ are picky and ambiguous. Multiple-choice examinations do share some of the limitations of other objective tests. They deny the demonstration of learning beyond that which is included in the alternatives provided in the question and they encourage guessing. Although, scoring is a rapid procedure, constructing good objective questions is difficult and very time consuming particularly if the questions are going to measure levels of thinking above simple factual knowledge.

Multiple-choice examinations also share some of the advantages objective tests have over essay style tests. Since objective questions require less student time to answer than essay questions it is possible to test the objectives of the course more comprehensively and to test the student through a wider range of difficulty levels. Objective tests can be scored much more quickly and accurately (often by machine) than essay tests which is often an important consideration for large classes.

The multiple-choice format is most frequently used to test knowledge and comprehension but instructors with expertise in writing MCQ can devise questions to evaluate higher order learning that includes, Application, Analysis, Synthesis and Evaluation. MCQs can also be used to assess the ability of students to integrate information from several sources. However, since MCQ which test higher-order learning require significant intellectual effort both
in reading and answering, students often become anxious when confronted with questions of this type.

Some of the ‘faults’ of multiple-choice examinations are not intrinsic to the MCQ themselves but rather result from a lack of skill in constructing clear, unambiguous, challenging questions which measure the level of cognitive performance and higher-order thinking which best match the objectives of the course. the questions and MCQ formats to be used.

Bibliography


What is a Multiple-choice Question?

A MCQ consists of a problem and a list of alternative solutions. The problem may be stated in the form of a direct question or incomplete statement called the *Stem* of the question. The student is requested to read the stem and alternative solutions and choose the one that is correct or best. This alternative is called the *Answer*. Depending on the format of the MCQ, there may be more than one answer. The other (incorrect) alternatives are called *Distracters*.  

19
Example

All of the following are involved in the synthesis of epinephrine EXCEPT

(A) Tyrosine hydroxylase
(B) Monoamine oxidase
(C) Dopa
(D) Norepinephrine
(E) Dopamine

The Answer is (B), while (A), (C), (D) and (E) are the Distracters.

The Stem

Characteristics of good stems include:

1. The best stems state the problem clearly and completely using a direct question in the positive form.

2. The stem should include as much of the item as possible (i.e., words repeated in each of the alternatives are better included in the stem).

3. Stems should not include irrelevant material. When testing higher-order learning, however, it may be relevant to include information which is usually available, although not required to solve the problem. This allows you to test whether the student knows which information is relevant to the problem.

4. Some problems, particularly those which test higher-order learning, are not amenable to direct questions and incomplete statements may be better in these cases. The stem should still direct the student to the problem being tested.

5. Whenever questions or incomplete statements must be made in the negative form it is crucial to emphasize this by using bold, italics, or CAPITALS, or better yet a COMBINATION of these, to alert the student.

The Answer and Distracters

Some recommendations for answers and distracters:

1. The answer(s) should be correct; not just within the context of the course material, but generally.

2. The distracters should be plausible. They are intended to distract students with insufficient understanding from the correct answer. They should represent common conceptual errors. If the question has numerical answers, calculate what the answers would be when students make specific errors and include these values as distracters. If a distracter is not selected by any student, it makes no contribution to the function of the item and it should be revised.
3. All of the alternatives should be *grammatically consistent* with the stem. If the format allows for more than one correct answer then be sure to use *is/are* etc. in the stem.

4. Avoid the use of *always* or *never* in distracters. Since few phenomenon or situations occur always or never, this qualifier is a cue that the distracter is incorrect.

5. Make the answer and distracters *equally long and qualified*. In a effort to satisfy the **generally correct** criterion of a good answer, instructors often make the answer longer and more carefully qualified than the distracters. It is also best to use similar phrases and/or sentence structures for all of the alternatives.

6. Beware of distracters *none of the above* or *all of the above* (usually used out of desperation in trying to find a fourth or fifth distracter). When the distracter is *all of the above* it is better to use a different format of MCQ which allows for this possibility automatically. If you use *none of the above*, then be sure that for some questions, none of the above, is the correct response.
Bloom's Taxonomy of Educational Objectives

It is a simple task to write MCQ which test at the knowledge and comprehension levels but well designed MCQ can test the higher levels of learning as well.

**Bloom's Taxonomy of Educational Objectives - Cognitive Domain**

- **Knowledge**
  - list, name, state, define, identify,
  - match, recall

- **Comprehension**
  - discuss, paraphrase, compute, extrapolate,
  - describe, explain, distinguish

- **Application**
  - choose, classify, use, interpret, calculate,
  - relate, demonstrate

- **Analysis**
  - separate, recognize, test, differentiate, solve

- **Synthesis**
  - design, order, develop, create, summarize,
  - combine, propose

- **Evaluation**
  - evaluate, justify, critique, appraise
Unless the course has emphasized the concept of conclusions as more precise than the way If the objectives of your course require students to be able to apply knowledge to novel situations or to integrate information to determine outcomes etc. then it is essential that you design questions which measure these skills. If you test only at the knowledge level then students will ignore the stated objectives and learn the material at the level of the examination questions. Likewise, it is unfair to have objectives or teach the material only at the knowledge/comprehension level but ask questions (whether MCQ or Essay Style) which require higher-order learning.

**Multiple-choice Question Formats**

There are many different formats of MCQs. Some are especially well suited for certain types of content. Some are particularly good for testing higher-order learning. Some are inherently ‘easier’ or ‘more difficult' than others. Four basic types will be described here with notes about their specific uses and advantages.

**A. One-Choice Completions - Best Answer**

This is the most commonly used MCQ format. It is simply a short-answer question with a number of alternatives to choose from. Some evaluation experts believe that this is the only type of MCQ that should be used because other formats are more susceptible to ambiguity and test wiseness.

**Example**

**Directions:** Each of the questions or incomplete statements below is followed by five suggested answers or completions. Select the ONE that is BEST in each case and blacken the appropriate space on the answer card.

1. Which of the following is the LEAST safe when driving on icy roads?

   (A) Accelerating slowly  
   (B) Applying the brakes rapidly and forcefully  
   (C) Holding the wheel firmly  
   (D) Slowing down and making turns gradually  
   (E) Increasing the distance to the car in front of you

   One-choice completions are often ‘easy' and frequently test at the knowledge level. The above is one such example. Nevertheless, it is a well formed question and much better than the following version.
1.1 Which of the following is not a safe driving practice on icy roads?

(A) Accelerating slowly
(B) Jamming on the brakes
(C) Holding the wheel firmly
(D) Slowing down gradually
(E) None of the above

The stem could be a problem for students (if the question was more difficult). The negative in the stem might be missed altogether or confuse students with the alternative 'none of the above'. In this version, (B) is a dead give-away because one should never jam on the brakes. Changing this alternative to "applying the brakes rapidly and forcefully" gives it the same soft phrasing as the other alternatives but now its longer and more qualified than the others; hence the reason for changing (D) to "slowing down and making turns gradually". (E) is just a wasted alternative, no one will think this is the correct response but "increasing the distance to the car in front of you" has the same weight as the other distracters.

Higher-order learning also can be tested using the one-choice completion format, and item difficulty can be very high.

Example - comprehension/synthesis

2. Normally, in the body, excitation of motoneurons in the spinal cord follows which of the following principles?

(A) Large neurons are excited more readily than small neurons
(B) Motoneurons to extensor muscles are excited more readily than those to flexor muscles
(C) Large motor units are recruited before small motor units
(D) Graded increases in muscle force are accomplished by exciting small motoneurons first
(E) Abrupt increases in muscle force are accomplished by exciting small motoneurons first

To answer this question the student must clearly understand the principles of motoneuron excitability and because the alternatives address these principles from different aspects of the content, the student must synthesize information from various sources. Note that the alternatives with similar phrasing and hence which address similar aspects of the content, are clumped together (A and B; D and E). This serves two purposes. It helps the student decide between pairs or groups of alternatives about similar content as s/he reads through the question and since the material is simply presented in a systematic way, the ordering itself offers no 'cue' to the exam-wise student.
**Example - analysis/evaluation**

4. If the Bank of Canada were to lower the lending rate to banks from 9% to 7.5%, there would likely be

(A) increased foreign investment in Canada  
(B) a rise in the Canadian dollar relative to the American dollar  
(C) an increase in savings deposits  
(D) an increase in housing starts  
(E) no change in mortgage interest rates

**Example - analysis/synthesis/application**

5. Which of the following would be produced by a 20 second occlusion of both carotid arteries between the heart and the carotid sinuses?

(A) increased blood pressure as measured in the femoral artery  
(B) decreased heart rate  
(C) increased activity in the afferent nerves from the carotid sinuses  
(D) decreased sympathetic outflow from the vasomotor center  
(E) none of the above

As shown on the next page, we could have avoided using "none of the above" in this question by simply using a different MCQ format. However, the correct answer, (A), is not obvious and the student must do a great deal of analysis and synthesis to realize that this is the correct response. In this circumstance, "none of the above" is the most attractive distracter.

**B. Multiple-Choice Completions (Also called Extended Matching or R types)**

This MCQ format allows for more than one correct answer and hence avoids the need for "all of the above" or "more than one of the above is correct", etc. It also makes the question inherently more difficult since the student is not just looking for one correct response among four incorrect responses. However, the intent of this format is not to test four separate points but rather to set up an interpretive exercise. Usually the instructor is still just testing one concept. When it is the instructor's intent to set a more difficult question on a topic, it is preferable to use this format rather than to increase the level of difficulty of a one-choice completion by including very obscure points (which are often not important points).
Example

Directions: For each of the questions or incomplete statements below, **ONE** or **MORE** of the alternatives given is correct. Answer:

(A) if only 1, 2 and 3 are correct
(B) if only 1 and 3 are correct
(C) if only 2 and 4 are correct
(D) if only 4 is correct
(E) if all are correct

6. **INCREASED** activity of parasympathetic nerves may

1. decrease heart rate
2. constrict pupils
3. increase gastric motility
4. vasodilate coronary arteries

Although this format appears confusing initially, students adapt quickly to it and rarely have difficulties because of the format itself. Nevertheless, it is helpful to summarize the directions at the top of each page of questions of this type. For example:

**Directions Summarized**

- A-  -B-  -C-  -D-  -E-
  1, 2 & 3          1 & 3           2 & 4              4              all are
  only     only  only  only         correct

The above question, like many one-choice completions, is testing at the **knowledge/comprehension** level. It is more difficult, however, than a one-choice completion on the same material because the student must know more information to get the question right. This fact leads to the most common complaint students make about this particular MCQ format, namely that they can know quite a lot about the particular content being tested but if they don't know all about it, they get the question wrong. This claim by the students is sometimes valid and we'll come back to discuss it when we consider overall construction of the multiple-choice examination. However, in this particular question, the instructor's intent is simply to test whether students appreciate that alternative #4 is correct; an exception to the usual rule that sympathetic nerves innervate blood vessels.

The multiple-choice completion format allows the testing of several points around a specific concept or topic. It is ideal for testing higher-order learning objectives. Although this format allows for more than one correct alternative, there are many combinations of alternatives which are not possible answers. For example:

1 and 2 alone cannot be correct
2 and 3 alone cannot be correct
3 and 4 alone cannot be correct
1 and 4 alone cannot be correct
4 is the only alternative which can stand alone as correct
there is no "none of the above" possibility
Other combinations besides the conventional ones which we have chosen are just as valid 
**BUT** be sure they are adequately explained. Whatever, multiple-choice completion format you 
choose, however, should be used consistently so that both students and faculty are familiar with 
them.

Although the possible and impossible combinations must be kept in mind while designing 
each question, this format has a number of advantages. For example, if there are two relatively 
easy points about a concept or topic and one point which is more difficult (or just more 
important), then the easy ones should be put in position (1) and (3) and the difficult or important 
one should be put in position (2). Position (2) can be thought of as the "tested" position. The (1) 
or (3) position, on the other hand can be used to "teach" information **during** the exam since if the 
student knows (1) is correct, for instance, then (3) must be correct.

Alternatives which are mutually contradictory should not be used in this format because 
they reduce its effectiveness.

**Example**

7. Which of the following symptoms are indicators of hyperthyroidism?

1. elevated heart rate  
2. lowered heart rate  
3. increased heat production  
4. decreased heat production

**Example - analysis/synthesis/application**

8. If a nerve/muscle preparation were bathed with a solution containing tetrodotoxin

1. sodium channels in the nerve axon would be blocked  
2. application of acetylcholine at the neuromuscular junction would fail to cause an end plate 
   potential  
3. application of acetylcholine at the neuromuscular junction would fail to cause a muscle 
   contraction  
4. direct electrical stimulation of the muscle will cause a normal contraction

This question illustrates the specific value of MCQs as diagnostic tools. When third year 
physiology students are asked to discuss or explain how tetrodotoxin would affect the 
excitation/contraction coupling mechanism, almost all of them can do so quite adequately. 
Nevertheless, experience tells me that many of them will not answer the above MCQ correctly. 
The reason students fail to get this question correct is not because the question is "tricky", it is 
because they fail at the **analysis, synthesis or application steps** which are required to decide that 
(2) and (4) are not correct responses. They know that tetrodotoxin acts at voltage dependent 
sodium channels. They know acetylcholine acts at ligand dependent channels. They know that 
the muscle contraction involves voltage dependent sodium channels. In other words they 
know each piece of information and can recite them either when asked as individual questions (as could 
be done with very simplistic MCQs) or in a paragraph of their own (or more likely, your own) 
words. **BUT** when they need to analyze the situation in the question, synthesize the information 
they have learned in the course and apply it, many of the students do not do it. The multiple-
choice completion format MCQ demonstrates this flaw in the way they "understand" the content.
When you give students opportunities to practice the problem-solving strategies needed for this kind of cognitive exercise (in laboratories, in-class problem solving sessions, practice with this sort of question) they respond by "making sense of the information and concepts" rather than on "learning what the instructor said about the topic".

Example - analysis/synthesis/application

9.

<table>
<thead>
<tr>
<th></th>
<th>&quot;A&quot;</th>
<th>&quot;B&quot;</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1 M NaCl</td>
<td>1 M KCl</td>
</tr>
</tbody>
</table>

Membrane

Compartment "A" contains 1 molar sodium chloride, compartment "B" contains an equal volume of 1 molar potassium chloride. These compartments are separated by a membrane which is permeable only to sodium and water. **At equilibrium**

1. the concentration of chloride in compartment "A" will be greater than in compartment "B"
2. an electrical potential will exist across the membrane with compartment "A" negative with respect to compartment "B"
3. the concentration of potassium in compartment "B" will be less than it was before equilibrium
4. there will be no **net** movement of water or sodium across the membrane

The preceding question is an **interpretive exercise**, and is an excellent way of testing higher-order learning objectives by MCQs. The illustration and stem set up a situation which must be **analyzed**; the student must be able to picture what happens and what the consequences are. Simple recall or recognition of correct statements are insufficient to answer this question. This is a particularly difficult version of this question because there are very few 'cues' for students about strategies for problem solving.

All of the alternatives are correct. (4) is the definition of equilibrium and is, therefore, a **knowledge** item. It is in position (4) because it can be selected by itself in this position, presumably by students who know the definition of equilibrium but cannot analyze the situation. Many students will appreciate that sodium and, therefore, water will move from compartment "A" to compartment "B", and that the volumes in the two compartments will change. It is a higher cognitive task, however, to appreciate that the concentrations of the non-diffusible substances (potassium and chloride) will also change. The question is cognitively less difficult, if one of the alternatives 'cues' them to this concept.

eg. compartment "A" will have a smaller volume than compartment "B"

The decisions about how to design the questions require a thorough understanding of the content, the concepts, the common conceptual difficulties that students have; i.e., all the same requirements for effective teaching are needed for the effective design of MCQs.
Many variations of the same interpretive exercise can be used (for example on subsequent tests), simply by changing some aspect of the initial situation or stem (eg. changing the permeability of the membrane) or by altering some the alternatives to make them wrong instead of correct or by using different answers and distracters. Some others, for example include:

- the concentration of sodium will be identical in both compartments **wrong**
- the concentration of chloride will be the same in both compartments **wrong**

**Example - evaluation**

10. Data:

<table>
<thead>
<tr>
<th>Substance</th>
<th>Concentration</th>
<th>Flow Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Substance &quot;X&quot; in plasma</td>
<td>0.05 mg/mL</td>
<td>-</td>
</tr>
<tr>
<td>Substance &quot;X&quot; in urine</td>
<td>7.5 mg/mL</td>
<td>-</td>
</tr>
<tr>
<td>Glomerular Filtration Rate</td>
<td>120.0 mL/min</td>
<td>-</td>
</tr>
<tr>
<td>Urine Flow Rate</td>
<td>1.0 mL/min</td>
<td>-</td>
</tr>
</tbody>
</table>

From the above data, it can be **CONCLUDED** that substance "X" is

1. freely filtered at the glomerulus
2. reabsorbed by the renal tubules
3. filtered at a rate of 6.0 mg/min
4. secreted by the renal tubules

The conceptual task for this question is to realize that although it is possible for (1), (2) and (3) to be true, the only conclusion that can be made is (4). It is important to pair alternatives (1) and (3) in this way, since if students think that (1) is correct, then (3) is the correct calculation. People use the word everyday, then it is unfair to expect this kind of precision of language on an exam. On the other hand, if a concept is an important one, it must be taught and tested.

**Multiple-choice Completions Based on Cases (Vignettes)**

In some disciplines it is possible to write several MCQs based on a single case or vignette. Cases test students in a situation more like the real world and make the exam seem more relevant to the students. Authors of these types of questions need to be aware, however, that the ability to answer one question may depend on the correct response to the previous question. Furthermore, students often answer the questions to a particular case, either, all correctly or all-incorrectly.

**Example - analysis/synthesis**

A 62 year old obese woman had a stormy postoperative course following abdominal surgery which involved resection of 35 cm of ileum. On the evening of the third postoperative day she had an attack of retching, during which her abdominal wound separated. She was returned to surgery and the wound was closed except for the insertion of a drainage tube. The patient was then started on an intravenous drip of 1 liter of 10% glucose in distilled water. At that time her blood pressure was 134/85 (normal is 120/90).
At midnight the nurse called the house physician because the IV had clotted. The physician spent 30 minutes attempting to reestablish the intravenous infusion, but all available superficial veins appeared to have thrombosed as a result of previous venipunctures. He reasoned that the intestinal cavity contained a large surface area for fluid absorption and, therefore, instilled 1 liter of 10% glucose in water through the abdominal drainage tube.

Two hours later the nurse found the patient unresponsive, with a blood pressure of 55/?.

A surgeon was called, who did an emergency cutdown to establish a route for intravenous therapy, and a transfusion of plasma was administered. The patient's blood pressure was promptly restored and she went on to a full recovery.

11. The bolus injection of 10% glucose into the abdominal cavity caused a precipitous drop in arterial pressure because it

1. was a hypertonic solution
2. caused a shift of water, but not solute, from vasculature into abdominal cavity
3. altered the Starling forces to increase filtration at the abdominal cavity
4. caused shrinkage of the osmoreceptors in the hypothalamus

12. At the time when the patient was found unresponsive, with a blood pressure of 55/?, you would expect that

1. the permeability of the collecting duct to water would be greater than normal
2. the permeability of the luminal membrane of the distal tubule to sodium chloride would be greater than normal
3. urine excretion would be less than normal
4. GFR would be greater than normal

13. At the time when the patient was found unresponsive, with a blood pressure of 55/?, you would expect that

1. the plasma concentration of Angiotensin II would be greater than normal
2. there would be an increase in the firing rate of the arterial baroreceptors
3. the Starling forces at capillary beds in skeletal muscle would favor reabsorption
4. the plasma concentration of sodium would be increased

Quantitative and Functional Relationships

Two other types of MCQ formats are useful for content which deals with quantitative and/or functional relationships. They are generally best for knowledge/comprehension but can sometimes be used to test higher-order learning objectives.

C. Quantitative Relationships

Example - knowledge
Directions: The following paired statements describe two entities which are to be compared in a quantitative sense. Answer:

(A) if (a) is greater than (b)
(B) if (b) is greater than (a)
(C) if the two are equal or very nearly equal

14. (a) The gross national product of Canada
(b) The gross national product of Sweden

Example - synthesis

15. (a) The amount of sodium reabsorbed by the **proximal** convoluted tubule in the presence of aldosterone
(b) The amount of sodium reabsorbed by the **distal** convoluted tubule in the presence of aldosterone
D. Functional Relationships

Example - comprehension

Directions: Each of the following pairs of phrases describes conditions or quantities which may or may not be related. Answer:

(A) if increase in the first is accompanied by increase in the second, or if decrease in the first is accompanied by decrease in the second.
(B) if increase in the first is accompanied by decrease in the second, or if decrease in the first is accompanied by increase in the second.
(C) if changes in the first are not necessarily accompanied by changes in the second.

16. (1) Plasma protein concentration
(2) interstitial fluid volume

Example - analysis/synthesis

17. (1) concentration of angiotensin II in plasma
(2) arterial pressure

18. (1) arterial pressure
(2) concentration of angiotensin II in plasma

The above two questions illustrate that the functional relationship format is basically asking the question "what happens to (2) when (1) is increased or decreased?" This format sets up a more specific task than simply asking whether the two entities are directly related, inversely related, or not related.
Assembling the Multiple-choice Examination

A few recommendations about assembling the examination:

1. MCQs must be checked much more carefully than essay examination questions because it is so easy for students to misinterpret them. Make sure the distracters as well as the answer are grammatically consistent with the stem. It is a very good idea to have a colleague or two check your questions. Ambiguities, answers which fail the 'generally true' rule, etc. are often more obvious to someone who is not as close to the content.

2. Group the questions according to MCQ format. I put one-choice completions first followed by the relationships formats and put multiple-choice completions last. I do this because the 'easier' format types warm the students up for the more difficult questions. Some students choose to do the more difficult ones first, however, so clearly there is no consensus on this arrangement.

3. Within each format, I arrange the questions according to topic as much as possible (analysis, synthesis and evaluation questions often cross topics) and usually arrange them chronologically from the beginning of the year. I do this primarily because students say it helps them focus on the questions but there is no evidence to suggest that it alters their performance.

Post-examination analysis

Multiple-choice examinations are typically marked by computer (example: MARKEX Program). This means that other types of computed data about your examination are also readily available (see attached example). Two important pieces of information are 'item difficulty' and 'point biserial'.

Item Difficulty

This is simply a question by question listing of the percent of students who got each question correct. All exams should have a range of item difficulties. Some questions should be answered correctly by 90% of the students (or better). Others should be answered correctly by only 35%. If all questions are answered correctly by 70% of the students the average for the examination will be 70% but very likely the same group of students will have answered each of the questions correctly, and the other 30% will be batting zero; probably not an accurate measure of their knowledge.
If students have done more poorly than you anticipated or much better than other years, the item difficulty data can often assist you in determining whether the reason lies with the examination itself or a difference in students. If similar questions have been used other years you can compare their respective item difficulties. The item difficulty printout also tells you how many students selected each of the alternatives. If a distracter is never selected you may want to revise it or replace it with one which is more attractive. If a distracter is chosen almost as often (or more often) than the answer you should look at the question critically to see if it is misleading (or you may want to look at your lecture notes to see whether there is a more effective way of teaching this material). Knowing the item difficulty for a question, can often help when analyzing an individual student's examination because it indicates how far off the average s/he is.

**Point Biserial**

This number represents a comparison between how often the top portion of the class answered a given question correctly and how often the bottom portion of the class answered it correctly. The higher the point biserial number, the better the question discriminates between the top and bottom of the class. When 80% or more of the students have answered a question correctly, the point biserial is often low because most of both the top and bottom of the class have got the question correct. The point biserial is not very important for these questions. For more difficult questions, however, it becomes very important. If the question is very difficult a low point biserial (i.e., the top half of the class does no better than the bottom half) is an indication that something may be wrong with the question. If all alternatives are chosen more or less equally, it probably means that the whole class was "just guessing". A review of the question may reveal ambiguous wording, emphasis on a very "picky" point, contradictions with other course material, or you didn't teach the material very well, or that they thought a crucial piece of information was "picky". When the item is difficult, the point biserial low and a distracter is favored over the answer there are very likely serious problems with the question or the teaching of the concept. Ambiguously worded questions, or content which conflicts with other information usually are to blame in these cases.
Example - Post-examination Analysis

DIRECTIONS SUMMARIZED

<table>
<thead>
<tr>
<th></th>
<th>-A-</th>
<th>-B-</th>
<th>-C-</th>
<th>-D-</th>
<th>-E-</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1, 2 &amp; 3</td>
<td>1 &amp; 3</td>
<td>2 &amp; 4</td>
<td>4</td>
<td>all are</td>
</tr>
<tr>
<td>2</td>
<td>only</td>
<td>only</td>
<td>only</td>
<td>only</td>
<td>correct</td>
</tr>
</tbody>
</table>

19. **Increased** activity of parasympathetic nerves may

1. decrease heart rate
2. constrict pupils
   3. increase gastric motility
   4. vasodilate coronary arteries

I.D. = .36
P.B. = -.02

Correct Answer is (E) All of the above

Number of Students who Answered

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>48</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>29</td>
</tr>
<tr>
<td>Low</td>
<td>44</td>
<td>7</td>
<td>0</td>
<td>0</td>
<td>30</td>
</tr>
</tbody>
</table>

This was a difficult question (only 36% of students got the correct answer). The question failed to discriminate between the top and bottom portions of the class. The students, however, are not just guessing. Virtually all of them know that 1, 2 and 3 are correct but only some of them realize that 4 is also correct. The **general** rule is that the sympathetic nervous system innervates blood vessels and the degree of constriction is related to the level of sympathetic activity. However, the exception is the coronary circulation which is innervated by the parasympathetic nervous system. The instructor was really only testing this one piece of information with this question and, therefore, is not concerned about (B), (C) and (D) not being effective distracters. S/he might be concerned, however, about how well this point was taught in class.
20. Which of the following statements pertaining to the electrocardiogram (ECG) are true?

1. PR interval represents the time from the beginning of atrial muscle depolarization to the beginning of ventricular depolarization
2. PR segment represents the delay in transmission of the impulse at the atrioventricular node
3. ST segment represents the interval of time between completions of depolarization and beginning of repolarization of the ventricular muscle
4. ST interval is measured from the end of the QRS complex to the beginning of the T wave

I.D. = .19
P.B. = .11

Correct Answer is (A) 1, 2 and 3 are correct

This question was answered very poorly by all students. Only 19% of them got the correct answer and there was very little difference between the top and bottom portions of the class. The distribution of the answers indicates that many students were simply guessing (probably even the 19% who got the correct answer). An examination of the question reveals a likely problem. The students may know a great deal about the ECG but are confused about which are the "segments" and which are the "intervals". As it turns out, this question only addressed a fairly "picky" point; the naming of the parts of an ECG. The intent of the instructor was to ask about the electrical activity of the heart during the ECG cycle. A better question would simply have given a picture of an ECG tracing and labelled it with letters and asked about what was happening between X and Y, Y and Z, etc.
Use of Multiple-choice Completions to Analyze Student Problems

Analysis of a student's multiple-choice examinations can reveal some helpful indicators of the nature of their difficulties. When students do very much better on the "easier" format questions (one-choice completions, quantitative and functional relationships) than on the multiple-choice completions, this is often an indication that they are studying the material as facts without doing any analysis or synthesis. They know what you said but they haven't made sense of it for themselves. An analysis of the types of errors which are made on the multiple-choice completions format itself can also reveal the nature of the student's problems. For example if the student most frequently gets this type of MCQ wrong because s/he includes distracters as answers (eg. says 1, 2 and 3 are correct when only 1 and 3 are correct, or says 2 and 4 are correct when only 4 is correct, etc.) then the basic problem may be more one of confidence than lack of understanding. They may simply be willing to believe that they don't know all the information and hence willing to believe an incorrect statement is right. When they fail to recognize correct answers (eg. they say 1 and 3 are correct when 1, 2 and 3 are correct, or they say only 1, 2 and 3 are correct when all are correct, etc.), however, the problem is one of lack of depth of knowledge. The student who says 1 and 3 are correct when 1, 2 and 3 are correct is, however, in a much better position than a student who says 1 and 3 are correct when actually 2 and 4 are correct. Teaching students how to analyze their own exams is well worth the effort. It will save you time and the act of examination-analysis is a very effective teaching tool and they will learn a great deal from it.

Acknowledgements

This Self-instructional Package on Designing Multiple-choice Examinations is adapted and extended from one which was produced by Dr. Bruce P. Squires when he was director of the Office of Health Sciences Educational Development at The University of Western Ontario. Dr. Squires inspired Dr. Robinson’s interest in teaching and enlightened me on the merits of multiple-choice examinations.

For a quick review of these materials there is a short course online:

http://teachingacademy.wisc.edu/Assistance/course/questions.htm

Multiple Choice Tests

The University of Tennessee at Chattanooga

http://www.ute.edu/Administration/WalkerTeachingResourceCenter/FacultyDevelopment/Assessment/test-questions.html#multiple%20choice
Good for:

- Application, synthesis, analysis, and evaluation levels

Types:

- Question/Right answer
- Incomplete statement
- Best answer

Advantages:

- Very effective
- Versatile at all levels
- Minimum of writing for student
- Guessing reduced
- Can cover broad range of content

Disadvantages:

- Difficult to construct good test items.
- Difficult to come up with plausible distractors/alternative responses.

Tips for Writing Good Multiple Choice items:

- Stem should present single, clearly formulated problem.
- Stem should be in simple, understood language; delete extraneous words.
- Avoid "all of the above"--can answer based on partial knowledge (if one is incorrect or two are correct, but unsure of the third...).
- Avoid "none of the above."
- Make all distractors plausible/homogenous.
- Don't overlap response alternatives (decreases discrimination between students who know the material and those who don't).
- Don't use double negatives.
- Present alternatives in logical or numerical order.
- Place correct answer at random (A answer is most often).
- Make each item independent of others on test.
- Way to judge a good stem: student's who know the content should be able to answer before reading the alternatives
- List alternatives on separate lines, indent, separate by blank line, use letters vs. numbers for alternative answers.
- Need more than 3 alternatives, 4 is best.

Another interesting website on multiple choice can be found at the University of North Carolina. They provide a lot of interesting about the construction of multiple choice items:

Addressing the following topics: Validity and Reliability, Cognitive Levels, The test matrix, Guidelines for writing questions, Analyzing the responses, Developing complex application questions.

PDF available at: http://ctl.unc.edu/FYC8.pdf#search=%22multiple%20choice%20matrix%22

Short Answer Questions

The University of Tennessee at Chattanooga

http://www.utc.edu/Administration/WalkerTeachingResourceCenter/FacultyDevelopment/Assessment/test-questions.html

Good for:

- Application, synthesis, analysis, and evaluation levels

Advantages:

- Easy to construct
- Good for "who," what," where," "when" content
- Minimizes guessing
- Encourages more intensive study-student must know the answer vs. recognizing the answer.

Disadvantages:

- May overemphasize memorization of facts
- Take care - questions may have more than one correct answer
- Scoring is laborious

Tips for Writing Good Short Answer Items:

- When using with definitions: supply term, not the definition-for a better judge of student knowledge.
- For numbers, indicate the degree of precision/units expected.
- Use direct questions, not an incomplete statement.
- If you do use incomplete statements, don't use more than 2 blanks within an item.
- Arrange blanks to make scoring easy.
- Try to phrase question so there is only one answer possible.

Alabama Professional Development Modules

http://web.utk.edu/~mccay/apdm/short/short_b.htm

Developing Short Answer and Completion Questions

Short answer and completion items are both forms of "supply" items in which students have to provide the response, rather than selecting a response from among several provided in the test.
Supply items are frequently used for recall of information and for problem solving in math and science (where the student is asked to supply the answer to a calculation or the result of a formula).

Educational objectives frequently state that the student will "know" certain information, procedures, etc. When "knowing" is defined as "recall," "remember," "label," "list," "state," "define," "describe," or "name," those objectives can be measured with short answer or completion items.

These types of questions have some advantages. Like true-false questions, short-answer and completion items can be written fairly easily. Students can complete a large number of items in a fairly short time (unless they involve working complex math problems), thus sampling a lot of content. Since the student has to generate the answers, the possibility of guessing the correct answers to these questions is greatly reduced when compared with true-false questions. While these items can be easy to score, poor student handwriting poses a potential problem.

Completion items are those in which a statement is written with blanks substituted for one or more words which the student is to supply. Example 1 shows the format for a typical completion item. Students sometimes write their answers directly on the blanks embedded in the sentences, but scoring can be facilitated by providing answer blanks in a column along either the right or left side of the paper (Example 1b).

1a There are ______ inches in a foot.
1b There are ______ inches in a foot.

Short answer questions are similar to completion items except that a question is written in its entirety, with the student supplying a correct response of one word or a short phrase. The use of a short answer question (Example 2) may be preferable to the completion item in Example 1 if it makes the question more specific and leads to the one answer that you are seeking. Also, it may be easier for younger children to respond to a question than to fill in a blank completing a sentence.

2 How many inches are there in a foot? __________ inches

A potential problem with both Short Answer and Completion items is that unless the items are well written, students may give an answer that is not the one you wanted when you wrote the item, but one which is also correct. It takes careful attention to write the item with enough specificity that the answer you are seeking is the only correct one. In Example 3a, the teacher wanted to know if the students remembered the year in which Columbus landed. Unfortunately other answers could be supplied relating to the season of the year, month, weather conditions, etc. Examples 3b and 3c show modifications that retain the completion format, while Example 3d has changed to the short-answer type of question. Note that both 3c and 3d provide a place for the student to write the answers down the right side of the paper for ease of grading.

3a Columbus landed in this country in __________.
3b Columbus landed in this country in the year __________.
3c Columbus landed in this country in the year __________. __________.
3d In what year did Columbus land in this country? __________.
The definition of a “short answer” question varies among the textbook authors. Sometimes "short-answer" refers to responses that may be a sentence or more in length (while not as long as an "essay" question) (Example 4). At other times it denotes a word (Example 5) or series of related terms (Example 6), phrase, list (Example 7), number (Example 8), or symbol (Example 9) to complete a statement or respond to a question.

4 Define the vocabulary words listed below:
Vegetarian ________________________________
Vegan ________________________________
Carnivore ________________________________
Herbivore ________________________________

5 What scientific law is illustrated when a leaf falls from a tree limb to the ground?

6 In the paragraph below, name the part of speech that is underlined. Write the name of the part of speech in the blank to the right corresponding to the letter labeling the word.

James (a) was anxious to begin the trip. (a)___________
He happily climbed (b) into the back of the car. (b)___________
James and his parents were going to (c) the zoo. (c)___________

7 Name the three primary colors ____________
________________________
________________________

8 What is the diameter of a circle with a circumference of 12 inches?
   Round your answer to two decimal places. ________ Inches

9 What are the symbols on the periodic table for the following elements:
   Silver ____________
   Nitrogen ____________
   Calcium ____________
   Potassium ____________
   Silver ____________

General Guidelines for Writing Short Answer and Completion Questions

Most of the following guidelines apply to both short answer and completion items. Guidelines that apply ONLY to one type of item are listed separately.

The requested answer should be brief and specific.

Statements should not be quoted directly from the text.

Answer blanks for short-answer items should be in the same place on the page (along the left or right margin) to aid in scoring, unless it creates confusion for the students. Separate answer blanks can also be provided in this same way for completion items, as shown in Example 10.
10 The last two states to be admitted to the United States are (a) in the year (b) and (c) in the year (d).

Directions for responding (including spelling requirements and use of sentences, if appropriate) should be provided.

There should be only one blank in an item unless the terms are part of a series.

The wording/grammar in the statement should not provide clues to the answer ("a/an" "is/are").

If the answer is a number, indicate the unit of measurement (pounds, cents, dollars, etc.) and the degree of specificity (three decimal places) that you require.

If you are trying to determine whether students know a process, use an example that focuses on the process rather than computational skills. Use numbers that are easy to compute. Example 11a below has a greater potential for mathematical errors than example 11b.

11a What is the average of the following numbers:
59 1/2, 83 1/2, and 44

11b What is the average of the following numbers:
63, 86, and 40

The questions should measure accomplishment of objectives appropriately.

The questions should be at an appropriate reading level for the students.

Blanks should be the same length to avoid giving clues about the respective answers.

**Completion Questions Only**

Only key or important words should be replaced by blanks in completion items.

Embedded blanks should be the same length.

Embedded blanks should be near the end of the sentence rather than at the beginning so that the student has an opportunity to formulate a framework before encountering the missing word or phrase.

**Issues in Scoring**

One issue that may arise is how to score answers that are correct but not the ones the teacher was looking for.

Another issue is whether to award partial credit for questions in which the student correctly supplies the word to fill one blank while producing an incorrect response for a second blank in the same sentence.

Since these items involve writing, the teacher should decide in advance (and inform the students in advance) if errors in spelling will be penalized.

If the answer is to be in sentence form, a similar decision must be made about the grammatical correctness of the sentence and what to do in case the response is a fragment or phrase rather
than a sentence.

**Short-Answer Questions**

*Short-answer questions are “constructed-response,” or open-ended questions that require students to create an answer.* Short-answer items typically require responses of one word to a few sentences. “Fill in the blank” and “completion” questions are examples of short-answer question types.

**Advantages:**
Short-answer questions assess unassisted recall of information, rather than recognition. Compared to essay questions, they are relatively easy to write.

**Disadvantages:**
Short-answer items are only suitable for questions that can be answered with short responses. Additionally, because students are free to answer any way they choose, short-answer questions can lead to difficulties in scoring if the question is not worded carefully. It’s important when writing short-answer questions that the desired student response is clear.

The above information was taken from the following website - http://teachingacademy.wisc.edu/Assistance/course/sanswer.htm

---

**Advantages:**
- Easy to construct
- Good for "who," what," where," "when" content
- Minimizes guessing
- Encourages more intensive study-student must know the answer vs. recognizing the answer.

**Disadvantages:**
- May overemphasize memorization of facts
- Take care - questions may have more than one correct answer
- Scoring is laborious

**Most Appropriate For:**
*Questions that require student recall over recognition.* Examples include assessing the correct spelling of items, or in cases when it is desirable to ensure that the students have committed the information to memory (medical students, for example, will require recall of information more than recognition by the nature of their jobs).

Short answer questions typically use these three levels of Bloom’s Taxonomy
Developing Short Answer Items

Short answer items (also called supplied response or constructed response items) are an effective measure of a student's ability to accurately recall specific, target information. Short answer items require students to either complete a statement (fill-in-the-blank or completion items) or answer a direct question using a single word or brief phrase. The nature of supplied response items lends itself well to the effective assessment of lower level learning objectives such as knowledge or comprehension of terms and definitions. As opposed to traditional objective measures (true-false, matching, multiple-choice, etc.) that assess the recognition of correct information, short answer items require students to independently generate their own response. While this type of recall assessment is more cognitively demanding, the independent nature of the responses makes scoring much more subjective. Due to the subjective interpretation of short answer items and the increased challenges in assessing the accuracy of responses, instructors should carefully examine the utility of short answer items in relationship to their instructional objectives.

<table>
<thead>
<tr>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scores less likely to be influenced by guessing</td>
<td>Accuracy of assessment may be influenced by handwriting/spelling skills</td>
</tr>
<tr>
<td>Requires increased cognitive ability to generate answers</td>
<td>Subjective nature can make scoring difficult and time consuming</td>
</tr>
<tr>
<td>Provide diagnostic information when looking at types of errors</td>
<td>Difficult to write items so that desired knowledge is clear</td>
</tr>
<tr>
<td>Promotes more in-depth study as students must recall answers</td>
<td>May overestimate learning due to bluffing</td>
</tr>
<tr>
<td>Effective for assessing who, what, where, and when information</td>
<td>Generally limited to knowledge and comprehension questions</td>
</tr>
<tr>
<td>Relatively easy to construct</td>
<td>Not suitable for item-analysis</td>
</tr>
<tr>
<td>Effective as either a written or oral assessment</td>
<td>Often criticized for encouraging rote memorization</td>
</tr>
<tr>
<td>Quicker for students to complete than</td>
<td></td>
</tr>
</tbody>
</table>
Tips for Writing Short Answer Items:

- **Questions must be clearly worded so that students understand the nature of the information being requested.** In order to facilitate understanding, phrase the item so that the required answer is brief and specific. In addition, provide clear clues to indicate the expected response.
- To ensure that a short answer question is an effective measure of student knowledge, require students to fill in important terms or phrases. For example, when assessing understanding of definitions, have students supply the term.
- **When utilizing short answer questions that require a numerical response, specify the degree of precision that is expected and the relevant units of measurement.**
- To prevent confusion and make scoring more precise, phrase question so there is only one answer or a limited range of answers possible. If multiple answers will correctly complete the item, ensure that there is a pre-established scoring rubric to deal with variations in response.
- Leave information to be filled in at or near the end of the question. This type of arrangement allows for ease of reading and enhances the efficiency of scoring.
- Utilize clear, explicit instructions that specify the format of the target answer (one word, multiple words, etc.) as well as the amount of acceptable variation (spelling, synonyms, etc.).
- To prevent confusion and ensure requested information is clear, limit the number of blanks within each short answer question. In addition, ensure that blanks are the same physical length to prevent context clues to the correct answer.
- Limit the influence of extraneous clues to the correct answer by utilizing correct, neutral grammar. Avoid providing grammatical clues to the correct answer (plurals, "a" versus "an," specific modifiers, etc.) and make certain that all correct responses can fit grammatically in the blank.
- **To reduce the emphasis on rote memorization of trivial information, do not use direct quotes from the text or lecture. Rather, phrase short answer items using unique or novel wording.**

Ideas to Enhance the Effectiveness of Short Answer Assessments:

*To ensure an accurate measure of target information, use direct questions rather than fill-in-the-blank or incomplete statements. This type of wording reduces confusion or ambiguity concerning the request information and directs students toward the relevant information.*
One of the benefits of short answer items is that they often encourage more intensive study of information due to the increased cognitive demands of recall over recognition. To promote this type of invested studying, award more credit for short answer items than for lower level recognition items (true-false, matching, etc.).

While short answer items often target knowledge or comprehension understanding, effectively developed completion items can also be utilized to assess application, synthesis, analysis, and evaluation levels. One means of measuring this type of higher-order understanding is to utilize combinations of short answer statements within a given paragraph. When implementing the paragraph format, be sure that desired knowledge is clearly specified.

Review Checklist:

_____ Is a short answer item an appropriate assessment of the learning objective?
_____ Does the content of the short answer question measure knowledge appropriate to the desired learning goal?
_____ Is the item clearly worded and stated in language appropriate to the student population?
_____ Is there only one clearly correct answer?
_____ Can the item be answered briefly and concisely using a single word or short phrase?
_____ Does the positioning of the item blank promote efficient scoring?
_____ Does the desired knowledge represent a key word or phrase?
_____ Is there a limited number of blanks in the short answer item?
_____ Do instructions clearly specify the desired knowledge and specificity of response?

The information above was taken from the following website - http://www.park.edu/cetl/quicktips/shortanswer.html

-----------------------------------
Short answer content

- Design for recall of specific names, facts, and basic knowledge.
- Focus the question so there a limited number of possible correct responses (or only one) in order to discriminate achievement.
- Make sure that question provides enough information so that students know what constitutes a correct response. For example, if you discussed four books written by a certain author and you want students to name them, be specific on how many constitute a correct answer (i.e., name two of four, or name any three, etc.).
Short answer protocol

- Make sure that the answer will require only a few words.
- **If a list is expected, limit the number of items to be listed to 6 or so.**
- Write the stem in your own words; that is, don’t use wording directly from a text or reading assignments.
- Make the stem complete enough so that the meaning is clear.
- When the stem is an incomplete statement, place the “blank” for the correct choice at the end of the statement.
- **When a negative item is used, emphasize the negative word or phrase, that is, underline, capitalize, or italicize, for example, “DOES NOT.”**
- Avoid clues to the correct answer such as “a” or “an” as the last word of the stem or plural verbs that match plural choices.

The information above was taken from the following website -
http://tlt.psu.edu/suggestions/questionwriting/sa_construct.shtml

---

4.2. Designing good marking schemes

*The aim* in summative assessment of course work is to assign a mark that closely reflects the ability of the student with respect to the learning objectives that are set out in the course module manual. (If there are several items of course work, each item of assessment may cover a subset of the complete list of learning objectives.) **So, good assessment criteria will ensure that no matter who marks the students' work, the final mark will properly reflect the student's ability**

- **There should be a clear correspondence between the markers notes and the specification for the work that was given to the students**
- **If specific information or specific conclusions should have been given in the students work provide the marker with a check list**
- There should be an indication of scale, i.e. how bad does it have to be to score 1 out of 5?
- Be realistic about the granularity of marking. For instance, if you have an assessment category where markers award marks for the clarity with which concepts were explained, it may be unreasonable to set 10 as the highest available mark for that category. If two reliable and well briefed markers marked the same set of work it is very unlikely that that they would agree on the assessment of many students at that level of precision. Marking would probably be more consistent if the available marks were set to 5 for this category and the job of marking would be a little less onerous. **If the overall maximum mark for the SAP (Short Answer Paper) is low then it is pointless setting high available marks on individual questions because the fine distinctions between students will be**
rounded away when the marks for all the questions are summed and the final mark as a proportion of the overall maximum mark for the SAP

• Avoid double penalties. If you have a category that assesses the ability to reach conclusions and include relevant facts and another category that assesses clarity of the English you should remind the marker not to penalize the student on the second category if the work was factually incorrect

The information above was taken from the following website -
http://www.oucs.ox.ac.uk/ltg/vle/docs/short_answer_papers/index.xml.ID=body.1_div.4

The following site I think might be very useful. It describes short answer questions and gives tips on how to develop them, but the most important tool on this page may be the “practice quiz” and practice activity that they have set up. It also has a classroom application. Some of these activities you might find useful, but I think that the main drawback from them could be that they are too easy. Many of their ideas for short answer questions relate to using fill-in-the-blanks and other items that some may not consider to be true “short answer” questions.  
Introduction to Short Answer and Completion Items (Alabama Professional Development Modules)

The following site is a newsletter from the University of Wisconsin-Madison, which discusses Short Answer questions. It discusses how to create and assess them.

http://testing.wisc.edu/LL11-03.pdf

The University of Wisconsin-Madison also provides a short course online"Exam Question Types and Student Competencies." This course will examine what kinds of question types are best for assessing desired student learning and comprehension. Through informed selection of test questions, you can ensure a more accurate evaluation of how well your students have mastered course material. This would be a useful site to review prior to creating your tests—it provides a quick overview of Bloom’s taxonomy and exam types.

http://teachingacademy.wisc.edu/Assistance/course/questions.htm