Is This a Trick Question?

A Short Guide to Writing Effective Test Questions

Is This a Trick Question?

A Short Guide to Writing Effective Test Questions

Designed & Developed by:
Ben Clay
Kansas Curriculum Center

Formatting & Text Processing by:
Esperanza Root

This publication was developed by the Kansas Curriculum Center with funds provided by the Kansas State Department of Education.
First printing:
October, 2001

Table of Contents

Preface...

Research indicates...

Teachers tend to use tests that they have prepared themselves much more often than any other type of test. (*How Teaching Matters,* **NCATE,** Oct. 2000)

While assessment options are diverse, most classroom educators rely on text and curriculum-embedded questions and tests that are overwhelmingly classified as **paper-and-pencil** (*National Commission on Teaching and America's Future,* 1996).

Formal training in paper-and-pencil test construction may occur at the preservice level (52% of the time) or as inservice preparation (21%). A significant number of professional educators (48%) report no formal training in developing, administering, scoring, and interpreting tests (*Education Week,* "National Survey of Public School Teachers, 2000").

Students report a higher level of test anxiety over teacher-made tests (64%) than over standardized tests (30%). The top three reasons why: poor test construction, irrelevant or obscure material coverage, and unclear directions. (**NCATE,** "Summary Data on Teacher Effectiveness, Teacher Quality, and Teacher Qualifications", 2001.)

A notable concern of many teachers is that they frequently have the task of constructing tests but have relatively little training or information to rely on in this task. **Is This a Trick Question?** is an information sourcebook for writing effective test questions. The central focus of the sourcebook's content is derived from standards developed by the National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

CRESST's criteria for establishing the technical quality of a test encompasses seven areas: cognitive complexity, content quality, meaningfulness, language appropriateness, transfer and generalizability, fairness, and reliability. Each aspect is discussed in the sourcebook in a straight-forward, jargon-free style.

Part One contains information concerning general test construction and introduces the six levels of intellectual understanding: knowledge, comprehension, application, analysis, synthesis, and evaluation. These levels of understanding assist in categorizing test questions, with knowledge as the lowest level. Since teachers tend to construct questions in the knowledge category 80% to 90% of the time, throughout the sourcebook are examples of or suggestions for developing higher order thinking skills. This supports Kansas' current Quality Performance Accreditation initiative which has established content and performance standards that cannot be measured by low-level tests.

Part Two of the information sourcebook is devoted to actual test question construction. Because of the diversity of assessment options, the sourcebook focuses primarily on paper-and-pencil tests, the most common type of teacher-prepared assessment. Five test item types are discussed: multiple choice, true-false, matching, completion, and essay. Information covers the appropriate use of each item type, advantages and disadvantages of each item type, and characteristics of well written items. Suggestions for addressing higher order thinking skills for each item type are also presented.

This sourcebook was developed to accomplish three outcomes:

- Teachers will know and follow appropriate principles for developing and using assessment methods in their teaching, avoiding common pitfalls in student assessment.

(Continued on next page...)

In Kansas...

The Kansas Commission on Teaching and America's Future (KCTAF), chaired by Dr. Andy Tompkins, Kansas Commissioner of Education, proposes to "develop higher-quality alternative pathways to teaching" as well as to "reinvent teacher preparation and professional development." As secondary and postsecondary institutions are exploring (out of necessity mostly) alternatives to traditional teacher recruitment, the need for training in assessment procedures and paper-and-pencil test construction in particular, become more and more evident.

- Teachers will be able to identify and accommodate the limitations of different informal and formal assessment methods.
- Teachers will gain an awareness that certain assessment approaches can be incompatible with certain instructional goals.

These three outcomes directly support the standards developed by a joint commission established by the National Education Association, the American Federation of Teachers, and the National Council on Measurement in Education. The initial standards were identified in 1990 and revised in 1999. In May 2001, a new listing was issued under the title "Standards for Teacher Competence in Educational Assessment of Students". The first two standards directly reflect the outcomes of this sourcebook:

- Teachers should be skilled in choosing assessment methods appropriate for instructional discussion
- Teachers should be skilled in developing assessment methods appropriate for instructional decisions.

While no one document can thoroughly address the needs and concerns expressed in all of this information, this sourcebook can be a valuable resource for any teacher who is interested in measuring outcomes of significance, tapping into higher-level thinking and problem solving skills, and constructing tests that effectively and fairly capture what a student knows.

Ben Clay, Coordinator
Kansas Curriculum Center

Pre-Test

Two general categories of test items

1. Objective items which require students to select the correct response from several alternatives or to supply a word or short phrase to answer a question or complete a statement

2. Subjective or essay items which permit the student to organize and present an original answer.

Objective items include:

- multiple choice
- true-false
- matching
- completion

Subjective items include:

- short-answer essay
- extended-response essay
- problem solving
- performance test items

Test your knowledge of these two item types by answering the following questions

Test Item Quiz

Circle the correct answer

T = True F = False ? = Unsure

1. Essay exams are easier to construct than are objective exams. **T F ?**
2. Essay exams require more thorough student preparation and study time than objective exams. **T F ?**
3. Essay exams require writing skills where objective exams do not. **T F ?**
4. Essay exams teach a person how to write. **T F ?**
5. Essay exams are more subjective in nature than are objective exams. **T F ?**
6. Objective exams encourage guessing more so than essay exams. **T F ?**
7. Essay exams limit the extent of content covered. **T F ?**
8. Essay and objective exams can be used to measure the same content or ability. **T F ?**
9. Essay and objective exams are both good ways to evaluate a student's level of knowledge. **T F ?**

Answers on next page...

Quiz Answers

1. *Essay exams are easier to construct than are objective exams.*
   TRUE Essay items are generally easier and less time consuming to construct than are most objective test items. Technically correct and content appropriate multiple choice and true-false test items require an extensive amount of time to write and revise.

2. *Essay exams require more thorough student preparation and study time than objective exams.*
   ? (QUESTION MARK) According to research findings it is still undetermined whether or not essay tests require or facilitate more thorough (or even different) student study preparation.

3. *Essay exams require writing skills where objective exams do not.*
   TRUE Writing skills do affect a student's ability to communicate the correct "factual" information through an essay response. Consequently, students with good writing skills have an advantage over students who do not.

4. *Essay exams teach a person how to write.*
   FALSE Essays do not teach a student how to write but they can emphasize the importance of being able to communicate through writing. Constant use of essay tests may encourage the knowledgeable but poor writing student to improve his/her writing ability in order to improve performance.

5. *Essay exams are more subjective in nature than are objective exams.*
   TRUE Essays are more subjective in nature due to their susceptibility to scoring influences. Different readers can rate identical responses differently, the same reader can rate the same paper differently over time, the handwriting, neatness or punctuation can unintentionally affect a paper's grade.

6. *Objective exams encourage guessing more so than essay exams.*
   ? (QUESTION MARK) Both item types encourage some guessing. Multiple choice, true-false and matching items can be correctly answered through blind guessing, yet essay items can be responded to satisfactorily through well written bluffing.

7. *Essay exams limit the extent of content covered.*
   TRUE Due to the extent of time required to respond to an essay question, only a few essay questions can be included on a exam. A larger number of objective items can be tested in the same amount of time, covering more content.

8. *Essay and objective exams can be used to measure the same content or ability.*
   TRUE Both item types can measure similar content or learning objectives. Research has shown that students respond almost identically to essay and objective test items covering the same content.

9. *Essay and objective exams are both good ways to evaluate a student's level of knowledge.*
   TRUE Both objective and essay test items are good devices for measuring student achievement. However, as seen in the previous quiz answers, there are particular measurement situations where one item type is more appropriate than the other.

Generally...

Creating a test is one of the most challenging tasks confronting an instructor. Unfortunately, many of us have had little, if any, preparation in writing tests.

Well constructed tests motivate students and reinforce learning. Well constructed tests enable teachers to assess the students mastery of course objectives. Tests also provide feedback on teaching, often showing what was or was not communicated clearly.

While always demanding, test writing may be made easier by considering the following suggestions for general test construction.

General Tips About Testing

Length of Test

In theory, the more items a test has, the more reliable it is. On a short test a few wrong answers can have a great effect on the overall results. On a long test, a few wrong answers will not influence the results as much. A long test does have drawbacks. If a test is too long, and particularly if students are doing the same kind of item over and over, they may get tired and not respond accurately or seriously. If a test needs to be lengthy, divide it into sections with different kinds of tasks, to maintain the student's interest.

Clear, Concise Instructions

It is necessary to give clear, concise instructions. It is useful to provide an example of a worked problem, which helps the students understand exactly what is necessary. What seems to be clear to the writer may be unclear to someone else.

Mix It Up!

It is often advantageous to mix types of items (multiple choice, true-false, essay) on a written exam or to mix types of exams (a performance component with a written component). Weaknesses connected with one kind of item or component or in students' test taking skills will be minimized.

Test Early

It is helpful for instructors to test early in the term and consider discounting the first test if results are poor. Students often need a practice test to understand the format each instructor uses and anticipate the best way to prepare for and take particular tests.

Test Frequently

Frequent testing helps students to avoid getting behind, provides instructors with multiple sources of information to use in computing the final course grade (thus minimizing the effect of "bad days"), and gives students regular feedback. It is important to test various topics in proportion to the emphasis given in class. Students will expect this practice and will study with this expectation.

Check For Accuracy

Instructors should be cautious about using tests written by others. Often, items developed by a previous instructor, a textbook publisher, etc., can save a lot of time, but they should be checked for accuracy and appropriateness in the given course.

(Continued on next page...)

General Tips About Testing

(Continued from previous page)

Proofread Exams

On written exams, it is important to proofread exams carefully and, when possible, have another person proofread them. Tiny mistakes, such as misnumbering the responses, can cause big problems later. Collation should also be checked carefully, since missing pages can cause a great deal of trouble.

One Wrong Answer

Generally, on either a written or performance test, it is wise to avoid having separate items or tasks depend upon answers or skills required in previous items or tasks. A student's initial mistake will be perpetuated over the course of succeeding items or tasks, penalizing the student repeatedly for one error.

Special Considerations

It is important to anticipate special considerations that learning disabled students or non-native speakers may need. The instructor needs to anticipate special needs in advance and decide whether or not students will be allowed the use of dictionaries, extra time, separate testing sites, or other special conditions.

A Little Humor

Instructors have found that using a little humor or placing less difficult items or tasks at the beginning of an exam can help students with test anxiety to reduce their preliminary tension and thus provide a more accurate demonstration of their progress.

What makes a test good or bad? The most basic and obvious answer to that question is that good tests measure what you want to measure, and bad tests do not.

It is always tempting to emphasize the parts of the course that are easiest to test, rather than the parts that are important to test.

When to Use Essay or Objective Tests

Essay tests are appropriate when:

- the group to be tested is small and the test is not to be reused.
- you wish to encourage and reward the development of student skill in writing.
- you are more interested in exploring the student's attitudes than in measuring his/her achievement.

Objective tests are appropriate when:

- the group to be tested is large and the test may be reused.
- highly reliable scores must be obtained as efficiently as possible.
- impartiality of evaluation, fairness, and freedom from possible test scoring influences are essential.

(Continued on next page...)

When to Use Essay or Objective Tests

(Continued from previous page)

Either essay or objective tests can be used to:

- measure almost any important educational achievement a written test can measure.
- test understanding and ability to apply principles.
- test ability to think critically.
- test ability to solve problems.

Conventional wisdom accurately portrays short-answer and essay examinations as the easiest to write and the most difficult to grade, particularly if they are graded well.

The matching of learning objective expectations with certain item types provides a high degree of test validity: testing what is supposed to be tested

Certain item types are better suited than others for measuring particular learning objectives. For example, learning objectives requiring the student to demonstrate or to show, may be better measured by performance test items, whereas objectives requiring the student to explain or to describe may be better measured by essay test items.

To further illustrate this principle, several sample learning objectives and appropriate test items are provided on the right. Match the most suitable test item with each of the learning objectives.

Matching Learning Objectives with Test Items

Instructions: Below are four test item categories labeled A, B, C, and D. Following these test item categories are sample learning objectives. On the line to the left of each learning objective, place the letter of the most appropriate test item category.

A = Objective Test Item (multiple choice, true-false, matching)
B = Performance Test Item
C = Essay Test Item (extended response)
D = Essay Test Item (short answer)

____1. Name the parts of the human skeleton

____2. Appraise a composition on the basis of its organization

____3. Demonstrate safe laboratory skills

____4. Cite four examples of satire that Twain uses in **Huckleberry Finn**

____5. Design a logo for a web page

____6. Describe the impact of a bull market

____7. Diagnose a physical ailment

____8. List important mental attributes necessary for an athlete

____9. Categorize great American fiction writers

____10. Analyze the major causes of learning disabilities

Answers: 1-A, 2-C, 3-B, 4-D, 5-B, 6-C, 7-B, 8-D, 9-A, 10-C

Planning the Test...

By definition no test can be truly objective: existing as an object of fact, independent of the mind.

In general, test items should...

- Assess achievement of instructional objectives
- Measure important aspects of the subject (concepts and conceptual relations)
- Accurately reflect the emphasis placed on important aspects of instruction
- Measure an appropriate level of student knowledge
- Vary in levels of difficulty

Implying that one type of question is automatically objective and the other necessarily subjective is a faulty assumption, since bias can occur with either type of test.

Criteria for Establishing

Technical Quality of a Test*

1. Cognitive Complexity

Standard: The test questions will focus on appropriate intellectual activity ranging from simple recall of facts to problem solving, critical thinking, and reasoning.

Cognitive complexity refers to the various levels of learning that can be tested. A good test reflects the goals of the instruction. If the instructor is mainly concerned with students memorizing facts, the test should ask for simple recall of material. If the instructor is trying to develop analytic skills, a test that asks for recall is inappropriate and will cause students to conclude that memorization is the instructor's true goal.

Refreshing the old bloom...

During the 1948 convention of the American Psychological Association, a group of educational psychologists decided it would be useful to classify different levels of understanding that students can achieve in a course.

In 1956, after extensive research on educational goals, the group published its findings in a book edited by Dr. Benjamin S. Bloom, a Harvard professor. Bloom's **Taxonomy of Educational Objectives** lists six levels of intellectual understanding:

- Knowledge
- Comprehension
- Application
- Analysis
- Synthesis
- Evaluation

These levels of understanding assist in categorizing test questions. Teachers tend to ask questions in the **knowledge** category 80% to 90% of the time. These questions are not bad, but using them all the time is. Try to utilize higher order level of questions. These questions require much more *brain power. (See the next page for a definition and sample question frames for each level of learning.)*

*Adapted from material developed by the National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

1. Cognitive Complexity (continued)

See **pages 59 & 60** for Cognitive and Affective Domain Guides.

Knowledge

Recognizing and recalling information, including dates, events, persons, places; terms, definitions; facts, principles, theories; methods and procedures

Sample Question Frames

Who invented the...?
What is meant by...?
Where is the...?

Comprehension

Understanding the meaning of information, including restating (in own words); translating from one form to another; or interpreting, explaining, and summarizing.

Sample Question Frames

Restate in your own words...?
Convert fractions into...?
List three reasons for...?

Application

Applying general rules, methods, or principles to a new situation, including classifying something as a specific example of a general principle or using a formula to solve a problem.

Sample Question Frames

How is...an example of... ?
How is...related to... ?
Why is...significant?

Analysis

Identifying the organization and patterns within a system by identifying its component parts and the relationships among the components.

Sample Question Frames

What are the parts of... ?
Classify ...according to...
Outline/diagram...

Synthesis

Discovering/creating new connections, generalizations, patterns, or perspectives; combining ideas to form a new whole.

Sample Question Frames

What would you infer from... ?
What ideas can you add to... ?
How would you create a... ?

Evaluation

Using evidence and reasoned argument to judge how well a proposal would accomplish a particular purpose; resolving controversies or differences of opinion.

Sample Question Frames

Do you agree...?
How would you decide about... ?
What priority would you give... ?

Criteria for Establishing

Technical Quality of a Test

(continued)

2. Content Quality

Standard: The test questions will permit students to demonstrate their knowledge of challenging and important subject matter.

Some important questions need to be answered concerning the content quality of the test. What are the test specifications? What skills do they indicate will be tested. How many questions and how many areas will be covered? How many sections will there be? What formats will be used to test?

If an instructor has focused on the War of 1812 in the majority of the class sessions and activities, this emphasis should be reflected in the test. A test that covers a much broader period will be regarded as unfair by the students, even if the instructor has told them that they are responsible for material that has not been discussed in class. Students go by instructors' implicit values more than their stated ones.

3. Meaningfulness

Standard: The test questions will be worth students' time and students will recognize and understand their value.

To Achieve Meaningfulness...

"In my opinion, students should not be forced to guess what will be on a test, or psych-out the teacher to decide what to study. Research shows that the less able students are heavily penalized by a failure to realize what is required for a test. The more able students seem to sense what the teacher wants, but the students most in need of help are likely to flounder even more painfully if they must guess what to study.

"The obvious solution to this problem is to give students specific study questions, then draw the test from the study questions. Sometimes this is criticized as *teaching the test,* as if having study questions in itself encourages a superficial approach. That may be true if there are very few study questions. However, if a teacher offers questions for all of the most important ideas in an assignment, then **teaching the test is teaching the course.**"

Russell A. Dewey, PhD
Georgia Southern University, Statesboro, GA

To Achieve Content Quality...

The first activity in planning a test is to outline the actual course content that the test will cover. A convenient way of accomplishing this is to take a few minutes following each class to list on an index card the important concepts covered in class and in assigned reading for that day. These cards can then be used later as a source of test items.

An even more conscientious approach would be to construct the test items themselves after each class. The advantage of either of these approaches is that the resulting test is likely to be a better representation of course activity.

It is very easy to write items which require only rote recall but are nonetheless difficult because they are taken from obscure passages (footnotes, for instance).

Preliminary findings by the National Center for Research on Evaluation, Standards, and Student Testing (CRESST):

Results of Applying Language Evaluation Criteria to Standardized Content Test Items

Math and science subsections: 67% percent of items had general vocabulary evaluated as uncommon or used in an atypical manner; 33% of items had syntactic structures evaluated as complex or atypical in their construction.

Reading comprehension: Same as above for vocabulary and syntax; 50% of items also had discourse level demands.

To reduce frustration for good students, avoid **all of these** and **none of these** and **both a & b** answers. These items are acceptable from a theoretical standpoint, but most prepared test-takers dislike them! As an example, the more subject matter a student knows, the easier it is to make arguments in favor of answers that the teacher might regard as wrong.

True-false questions are the worst of all in this regard. Often the truth value of an isolated statement is quite debatable! It all depends on how it is interpreted, the definition of a key term, or the context.

4. Language Appropriateness

Standard: The language demands will be clear and appropriate to the assessment tasks and to students.

Test questions should reflect the language that is used in the classroom. Test items should be stated in simple, clear language, free of nonfunctional material and extraneous clues. Test items should also be free of race, ethnic, and sex bias. Beyond these two qualifications, students' language backgrounds impact their performance on tests. The vocabulary (uncommon usage; nonliteral usage) and the syntax of the test (atypical parts of speech; complex structures) may create language barriers.

Modifications of the test for students that are limited English proficient include: assessment in the native language; text changes in vocabulary; modification of linguistic complexity; addition of visual supports; use of glossaries in native language; use of glossaries in English; linguistic modification of test directions; and additional example items/tasks.

5. Transfer and Generalizability

Standard: Successful performance on the test will allow valid generalizations about achievement to be made.

Presentations, scenarios, projects and portfolios add dimensions to assessment that traditional testing cannot. Teachers can make valid generalizations about achievement more easily using authentic and performance assessments. These generalizations may involve instructional placement decisions, formative evaluation decisions and diagnostic decisions. Well constructed tests—whether they are objective or performance oriented—allow teachers to understand what needs to be taught next. Teachers are also able to monitor a student's learning, while instruction is underway, and can change the instruction program as needed.

Criteria for Establishing

Technical Quality of a Test

(continued)

6. Fairness

Standard: Student performance will be measured in a way that does not give advantage to factors irrelevant to school learning; scoring schemes will be similarly equitable.

Here are a few basic rules of fairness: test questions should reflect the objectives of the unit; expectations should be clearly known by the students; each test item should present a clearly formulated task; one item should not aide in answering another; ample time for test completion should be allowed; and assignment of points should be determined before the test is administered.

Grading constructively requires the instructor to provide feedback (written and/or oral) that helps the students to appreciate what they achieved and did not achieve by taking the test. This feedback could include the following: encouraging comments on a test or paper that convey respect for what the student attempted to accomplish; praise for what the student did accomplish and suggestions for improving performance.

7. Reliability

Standard: Answers to test questions will be consistently trusted to represent what students know.

The whole point of testing is to encourage learning. A good test is designed with items that are not easily guessed without proper studying. It is possible to construct all types of test questions which are not readily guessed and therefore require a student to comprehend basic factual material.

Multiple choice questions are widely scorned as **multiple guess** questions. The solution to this problem is to design multiple choice items so that students who know the subject or material adequately are more likely to choose the correct alternative and students with less adequate knowledge are more likely to choose a wrong alternative. *(On the next page are suggestions on how to defeat the **TEST-WISE** strategies of students who do not study.)*

Five hundred secondary and postsecondary students were surveyed for suggestions on how an instructor could grade fairly and accurately.

Here are the top 10 responses.

- Consider grading based only on mastery of material and not on personalities or perceived effort.
- Do not over emphasize grades. Emphasize learning over grades.
- Keep students informed of their progress throughout the term.
- Clearly state grading policies and procedures in the syllabus and review them with the class during orientation.
- Avoid modifying policies during the term.
- Provide plenty of opportunities for assessment. This will avoid unnecessary pressure and allow for some mistakes.
- Provide some choice in format or topic when assigning work.
- Keep accurate records of grades. Record numerical grades, rather than letter grades, when possible.
- Consider allowing rewrites on papers.
- If many do poorly on an exam, schedule an exam for the following week to retest the class.

How to Defeat the Common Rules of Thumb Which Students Use to Guess Correct Answers

Rule of thumb: Pick the longest answer.

Way to defeat this strategy: Make sure the longest answer is right about a fifth of the time (if there are five alternatives for each question).

Rule of thumb: Pick the 'b' alternative.

Way to defeat this strategy: Make sure each answer is used the same number of times, in random order.

Rule of thumb: Never pick an answer which uses the word 'always' or 'never' in it.

Way to defeat this strategy: Make sure such answers are correct about a fifth of the time.

Rule of thumb: If there are two answers which express opposites, pick one or the other and ignore other alternatives.

Way to defeat this strategy: Sometimes offer opposites when neither is correct.

Rule of thumb: If in doubt, guess.

Way to minimize the impact of this strategy: Use five alternatives instead of three or four.

Rule of thumb: Pick the scientific-sounding answer.

Way to defeat this strategy: Use scientific sounding jargon in wrong answers

---

Rule of thumb: Do not pick an answer which is too simple or obvious.

Way to defeat this strategy: Sometimes make the simple, obvious answer the correct one.

Rule of thumb: Pick a word which you remember was related to the topic.

Way to defeat this strategy: When drawing up distractors (wrong answers) use terminology from the same area of the text as the right answer, but in distracters use those words incorrectly so the wrong answers are definitely wrong.

7. Reliability (continued)

Studies have shown that the grade given to an essay test depend in part upon the neatness of the handwriting. That seems like a poor way to assign a grade. However, if students are asked to do the test on a word processor, it is hard to ensure that the work is original. Studies have also shown that grades for essay tests are influenced by length. If a student rambles on, there is greater likelihood of hitting a few points that the teacher is looking for. But do we want to reward verbosity?

Despite all this, essay and short answer tests have many virtues. Students need practice formulating arguments, expressing things clearly, and integrating ideas. Nobody would argue that **all** testing should be multiple choice. However, for teachers in many situations, a good objective test is both fairer and more efficient than an essay or short answer test. One way to ensure reliability is to share with your students…

General Test Taking Tips

1. Tell students to survey the entire test before they begin. This will help them identify which section will be quick and/or easy and which will require more time and thought.
2. Encourage students to underline important words in the directions such as list, discuss, define, etc.
3. Instruct students that when they take a test, they should do the easy questions first.
4. Help students schedule their time by estimating the total time available compared to the number of questions on the test. They need to recognize that some types of questions will take longer than others.
5. Suggest that students put a checkmark next to any questions which they left blank and will need to come back to for completion later.
6. Prompt students to hold onto their test until they have looked it over thoroughly. They should make sure they have completed each task and have reread the entire test to verify that they have given the answers they intended.

"Remind, remind, remind students to stop and ask for directions or clarification if there is something they don't understand. Directions are the roadmap to their final destination."

Suggestion…
Encourage students to design their own test. This will help them anticipate some of the questions or information to be included on the instructor's exam.

Various kinds of objective and essay test items are presented in the following sections of this document.

Each kind of test item is briefly described in terms of advantages and limitations for use.

General suggestions are also presented for the construction of each test item variation.

Multiple Choice Test Items

"...almost any well defined cognitive objective can be tested fairly in a multiple choice format."

Section Summary

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

The multiple choice item consists of the stem, which identifies the question or problem and the response alternatives or choices. Usually, students are asked to select the one alternative that best completes a statement or answers a question. For example,

Item Stem: Which of the following is a chemical change?
Response Alternatives:
a. Evaporation of alcohol
b. Freezing of water
c. Burning of oil ✓
d. Melting of wax

Multiple choice items are considered to be among the most versatile of all item types. They can be used to test factual recall as well as levels of understanding and ability to apply learning. As an example, the multiple choice item below is testing not only information recall but also the ability to use judgment in analyzing and evaluating.

Multiple choice tests can be used to test the ability to:
1. recall memorized information
2. apply theory to routine cases
3. apply theory to novel situations
4. use judgment in analyzing and evaluating

A. 1 only
B. 1 and 2 only
C. 1, 2 and 3 only
D. 1, 2, 3 and 4 ✓

Multiple choice items can also provide an excellent basis for post-test discussion, especially if the discussion addresses why the incorrect responses were wrong as well as why the correct responses were right. Unfortunately, multiple choice items are difficult and time consuming to construct well. They may also appear too discriminating (picky) to students, especially when the alternatives are well constructed and are open to misinterpretation by students who read more into questions than is there.

Test your knowledge of multiple choice tests by taking the multiple choice test on the next page...

Circle the Most Correct Answer

1. Multiple choice items provide highly reliable test scores because:
A. they do not place a high degree of dependence on the students reading ability
B. they place a high degree of dependence on a teacher's writing ability
C. they are a subjective measurement of student achievement
D. they allow a wide sampling of content and a reduced guessing factor

2. You should:
A. always decide on an answer before reading the alternatives
B. always review your marked exams
C. never change an answer
D. always do the multiple choice items on an exam first

3. The above multiple choice item is structurally undesirable because:
A. a direct question is more desirable than a incomplete statement
B. there is no explicit problem or information in the stem
C. the alternatives are not all plausible
D. all of the above
E. A & B only
F. B & C only
G. A & C only
H. none of the above

4. The above multiple choice item is undesirable because:
A. it relies on an answer required in a previous item
B. the stem does not supply enough information
C. eight alternatives are too many and too confusing to the student
D. more alternatives just encourage guessing

5. The right answers in multiple choice questions tend to be:
A. longer and more descriptive
B. the same length as the wrong answers
C. at least a paragraph long
D. short

6. When guessing on a multiple choice question with numbers in the answer:
A. always pick the most extreme
B. pick the lowest number
C. pick answers in the middle range
D. always pick C

7. What is the process of elimination in a multiple choice question?
A. skipping the entire question
B. eliminating all answers with extreme modifiers
C. just guessing
D. eliminating the wrong answers

7. What should you not do when taking a multiple choice test:
A. pay attention to patterns
B. listen to last minute instructions
C. read each question carefully
D. read all choices

8. It is unlikely that a student who is unskilled in untangling negative statements will:
A. quickly understand multiple choice items not written in this way
B. not quickly understand multiple choice items written in this way
C. quickly understand multiple choice items written in this way
D. not quickly understand multiple choice items not written in this way

Answers: 1-D, 2-B, 3-D, 4-C, 5-A, 6-C, 7-D, 8-C

Suggestions For Writing Multiple Choice Test Items

1. When possible, state the stem as a direct question rather than as an incomplete statement.

Undesirable:	*Desirable:*
Alloys are ordinarily produced by...	How are alloys ordinarily produced?

2. Present a definite, explicit and singular question or problem in the stem.

Undesirable:	*Desirable:*
Psychology...	The science of mind and behavior is called...

3. Eliminate excessive verbiage or irrelevant information from the stem.

Undesirable:	*Desirable:*
While ironing her formal, Jane burned her hand accidently on the hot iron. This was due to a transfer of heat between...	Which of the following ways of heat transfer explains why Jane's hand was burned after she touched a hot iron?

4. Include in the stem any word(s) that might otherwise be repeated in each alternative.

Undesirable:	*Desirable:*
In national elections in the United States the President is officially: A. chosen by the people. B. chosen by members of Congress. C. chosen by the House of Representatives. D. chosen by the Electoral College.✓	In national elections in the United States the President is officially chosen by A. the people. B. members of Congress. C. the House of Reps. D. the Electoral college. ✓

Remember...

- Use at least four alternatives for each item to lower the probability of getting the item correct by guessing.
- Use capital letters (A, B, C, D) as response signs rather than lower case letters ("a" gets confused with "d" and "c" with "a" if the type or duplication is poor).
- Randomly distribute the correct response among the alternative positions throughout the test, having approximately the same proportion of alternatives A, B, C, and D as the correct response.
- Avoid irrelevant clues such as grammatical structure, well known verbal associations or simplistic connections between stem and answer.
- When possible, present alternatives in some logical order (e.g., chronological, most to least, alphabetical).
- Use the alternatives **none of the above** and **all of the above** sparingly. When used, such alternatives should occasionally be used as the correct response.

Remember...

- In testing for definitions, use the term in the stem rather than as an option.
- List alternatives on separate lines (rather than including the options as part of the stem) so that all options can be clearly distinguished.
- Keep all alternatives in a similar format (i.e., all phrases, all sentences, etc.).
- Try to make alternatives for an item approximately the same length. (Making the correct response consistently longer is a common error.)
- Use misconceptions students have indicated in class or errors commonly made by students in the class as the basis for incorrect alternatives.
- Way to judge a good stem: students who know the content should be able to answer before reading the alternatives.

Multiple choice exams provide easier conditions for cheating than essay tests since single letters or numbers are easier to see than extensive text. Cheating can be minimized by using alternative test forms and controlling seating.

5. Use negatively stated stems sparingly. When used, underline and/or capitalize the negative word.

Undesirable:

Which of the following is not cited as an accomplishment of the Kennedy administration?

Desirable:

Which of the following is NOT cited as an accomplishment of the Kennedy administration?

6. Make all alternatives plausible and attractive to the less knowledgeable or skillful student.

Undesirable:

What process is most nearly the opposite of photosynthesis?
A. Digestion
B. Assimilation
C. Respiration ✓
D. Catabolism

Desirable:

What process is most nearly the opposite of photosynthesis?
A. Digestion
B. Relaxation
C. Respiration ✓
D. Exertion

7. Make the alternatives mutually exclusive.

Undesirable:

The daily minimum required amount of milk that a 10 year old child should drink is
A. 1-2 glasses.
B. 2-3 glasses. ✓
C. 3-4 glasses. ✓
D. at least 4 glasses.

Desirable:

What is the daily minimum required amount of milk a 10 year old child should drink?
A. 1 glass.
B. 2 glasses.
C. 3 glasses. ✓
D. 4 glasses.

8. Make alternatives approximately equal in length.

Undesirable:

The most general cause of low individual incomes in the United States is:
A. lack of valuable productive services to sell. ✓
B. unwillingness to work.
C. automation.
D. inflation.

Desirable:

What is the most general cause of low individual incomes in the United States?
A. A lack of valuable productive services to sell. ✓
B. The population's overall unwillingness to work.
C. The nation's increased reliance on automation.
D. An increasing national level of inflation.

Attention Students: Multiple Choice Test Taking Tips

1. Read the directions carefully

The directions usually indicate that some alternatives may be partly correct or correct statements in themselves, but not when joined to the stem. The directions may say: "choose the most correct answer" or "mark the one best answer." Sometimes you may be asked to "mark all correct answers."

2. Do the multiple choice items first

If your exam has types of questions other than multiple choice, just reading the stems and alternatives acts is a warm-up to the material. (The stem is the question and the alternatives are the choices). Also, the ideas embedded in these multiple choice questions will fuel your thinking for doing the other parts of the exam. Use the process of elimination procedure. Eliminate the obviously incorrect alternatives.

3. Read all of the stem and every alternative

Read the stem with each alternative to take advantage of the correct sound or flow that the correct answer often produces. Also, you can eliminate any alternatives that do not agree grammatically with the stem.

Some students find it effective to read the stem and anticipate the correct alternative before actually looking at the alternatives. If you generally do better on essay exams, this strategy may help you a great deal.

4. Consider "all of the above" and "none of the above"

Examine the "above" alternatives to see if all of them or none of them apply totally. If even one does not apply totally, do not consider "all of the above" or "none of the above" as the correct answer. Make sure that a statement applies to the question since it can be true, but not be relevant to the question at hand!

(Continued on next page...)

Pay attention to the words...

- Note qualifying words: **usually, often, generally, may,** and **seldom** are qualifiers that could indicate a true statement.
- Words such as **every, all, none, always,** and **only** are superlatives that indicate the correct answer is an undisputed fact. In general, absolutes are rare.
- If a negative word such as **none, not, never,** or **neither** is in the stem, assume that the correct alternative must be a fact or absolute and that the other alternatives could be true statements, but not the correct answer.

Plan to progress through the exam in three ways:

- Read every question carefully but quickly, answering only those of which you are 100% certain. Put a "?" on those that need more thought.
- Then, examine/study the questions not yet answered. Answer those you are reasonably sure of without pondering too long on each. Erase the "?".
- Finally, study the remaining unanswered questions. If you cannot come to a decision by reasoning or if you run out of time, guess. Erase the "?". Note that some examinations penalize "guessing" by subtracting points for incorrect answers.

If there is no penalty, then a guess is better than a blank.

5. Plan your time

Often you are required to answer up to 70 multiple choice questions in an hour or less. This means you may have less than a minute, on average, to spend on each question. Some questions, of course, will take you only a few seconds, while others will require more time for thought.

6. Changing answers

Research has shown that changing answers on a multiple choice or true-false exam is neither good nor bad: if you have a good reason for changing your answer, change it. The origin of the myth that people always change from "right" to "wrong" is that those (i.e. the wrong ones) are the only ones you will see when you review your exam—you will not notice the ones you changed from "wrong" to "right."

This will pay dividends on future exams...

Study your marked and returned exam in order to learn from your successes and mistakes.

After Your Exam Has Been Returned

1. Examine each question you did get correct. Remember how you knew that the information was important when you studied. How did you study?
2. Examine each question you did not get correct in order to understand the distinction between the correct alternative and the incorrect alternatives. Ask yourself why the correct answer is correct and why the other alternatives are incorrect.
3. Determine the level of thought your instructor expects of you by reading through all of the questions. Are you expected to recognize, analyze, synthesize and/or apply the material that has been presented to you? Study accordingly for the next exam.

Multiple Choice Test Items: Conclusion

Ask yourself:

Why are these multiple choice questions crummy?

1. How frequently do you take a sick day from work?
   A. never
   B. once or twice a year
   C. 3 to 5 times a year
   D. 6 to 12 times a year
   E. at least once a month

2. Identify the issue that you believe is most critical to this country's future.
   A. the economy
   B. education
   C. integrity in government
   D. national defense
   E. some other issue

"Understand that there is always one clearly best answer. My goal is not to trick students or require them to make difficult judgments about two options that are nearly equally correct. My goal is to design questions that students who understand will answer correctly and students who do not understand will answer incorrectly."

John A. Johnson
Dept. of Psychology,
Penn State University

Aim for Higher Levels of Learning

Most teachers find it easier to construct multiple choice items to test recall and comprehension and to use essay items to test higher-level learning objectives. But other possibilities exist. Multiple choice items that require students to do such things as classify statements as fact or opinion go beyond simple recall of facts.

Here are two examples of multiple choice test items designed for higher order thinking skills.

A common goal of the Salt March in India, the Boxer Rebellion in China, and the Zulu resistance in southern Africa was to:
A. overthrow totalitarian leaders
B. force upper classes to carry out land reform programs
C. remove foreign powers
D. establish Communist parties to lead the governments

In western Europe, which development caused the other three?
A. decline of trade
B. fall of Rome
C. breakdown of central government
D. rise in the power of the Roman Catholic Church

One way to write multiple choice questions that require more than recall is to develop questions that resemble miniature "cases" or situations. Provide a small collection of data, such as a description of a situation, a series of graphs, quotes, a paragraph, or any cluster of the kinds of raw information that might be appropriate material for the activities of your discipline.

Then develop a series of questions based on that material. These questions might require students to apply learned concepts to the case, to combine data, to make a prediction on the outcome of a process, to analyze a relationship between pieces of the information, or to synthesize pieces of information into a new concept.

True-False Test Items

There are many situations which call for either-or decisions, such as deciding whether a specific solution is right or wrong, whether to continue or to stop, whether to use a singular or plural construction, and so on. For such situations, the true-false item is an ideal measuring device.

Section Summary

Good for:

- Knowledge level content
- Evaluating student understanding of popular misconceptions
- Concepts with two logical responses

Advantages:

- Can test large amounts of content
- Students can answer 3-4 questions per minute

Disadvantages:

- They are easy
- It is difficult to discriminate between students that know the material and students who do not
- Students have a 50-50 chance of getting the right answer by guessing
- Need a large number of items for high reliability

In the most basic format, true-false questions are those in which a statement is presented and the student indicates in some manner whether the statement is true or false. In other words, there are only two possible responses for each item, and the student chooses between them. True-false questions are well suited for testing student recall or comprehension. Students can generally respond to many questions, covering a lot of content, in a fairly short amount of time.

From the teacher's perspective, true-false questions can be written quickly. They are easy to score. Because they can be objectively scored, the scores are more reliable than for items that are at least partially dependent on the teacher's judgment.

Select or Supply?

True-false questions require the students to select a response (true or false) that shows recognition of correct or incorrect information that is presented to them. These are included among the items that are called **selection,** in contrast to **supply** items in which the student must supply the correct information.

Forced Choice

Another term applied to true-false items is **forced choice** because the student must choose between two possible answers. Educational objectives that specify the student will *identify, select,* and *recognize* material are appropriately targeted to either forced choice questions or more complex matching or multiple choice questions.

Much Maligned and Abused...

Many educators feel that true-false test items serve little or no measurement purposes because true-false items are subject to guessing. (But the likelihood of obtaining a substantially higher than chance score by guessing alone is very small). In general, individual true-false items are less discriminating than individual multiple choice items. There is a tendency to write trivial true-false items, which lead students to verbatim memorization. At the same time, no diagnostic information is available from incorrect responses to true-false items. Finally, true-false items are not amenable to concepts that cannot be formulated as propositions.

Summarizing the Argument for the Value of True-False Test Items

- The essence of educational achievement is the command of useful verbal knowledge.
- All verbal knowledge can be expressed in propositions.
- A proposition is any sentence that can be said to be true or false.
- The extent of students' command of a particular area of knowledge is indicated by their success in judging the truth or falsity of propositions related to it.

Ebel and Frisbie (1991)

Since true-false questions tend to be either extremely easy or extremely difficult, they do not discriminate between students of varying ability as well as other types of questions do.

Making the Case for True-False Items

- Versatility—True-false items are adaptable to the measurement of a wide variety of learning outcomes.
- Scoring accuracy and economy—Scoring keys can be economically applied by machine or clerical assistants.
- Reliability—True-false tests that are highly reliable can be constructed.
- Amenable to item analysis—True-false items are amenable to item analysis, by means of which they can be improved.
- Efficiency—More test responses can be obtained from a given amount of written material and in a given amount of time than from other forms.
- True-false items are useful in testing misconceptions.
- True-false items can be expressed in few words, making them less dependent on reading ability.

Check Your Knowledge of True-False Test Items

Directions: For each question below, circle A or B.

1. Is it recommended to take statements directly from the text to make good true-false questions?
   A. Yes
   B. No
2. Two ideas can be included in a true-false statement if the purpose is to show cause and effect.
   A. Yes
   B. No
3. When a true-false statement is an opinion, it should be attributed to someone in the statement.
   A. Yes
   B. No
4. Underlining or circling answers is preferable to having the student write them.
   A. Yes
   B. No

Circle "Good" if it describes a good practice in true-false questions, circle "Poor" if it characterizes a poor practice.

5. Complex statements are used to measure higher order thinking. **Good Poor**
6. If negatives, such as "not," are used, they should be highlighted in some way. **Good Poor**
7. True and false statements should be approximately the same length. **Good Poor**
8. There should be a recognizable pattern in the answers, such as TFTFTFTF. **Good Poor**
9. The following are examples of words that should be avoided: "all," "none," "never," "sometimes," "generally," and "often." **Good Poor**

Answers: 1-B, 2-A, 3-A, 4-A, 5-Poor, 6-Good, 7-Good, 8-Poor, 9-Good

Suggestions For Writing True-False Test Items

Remember...

- Keep language as simple and clear as possible.
- Use a relatively large number of items (75 or more when the entire test is T/F).
- Be aware that extremely long or complicated statements will test reading skill rather than content knowledge.
- Require students to circle or underline a typed "T" or "F" rather than to fill in a "T" or "F" next to the statement, thus avoiding having to interpret confusing handwriting.
- If a proposition expresses a relationship, such as cause and effect or premise and conclusion, present the correct part of the statement first and vary the truth or falsity of the second part.
- Make true and false items of approximately equal average length throughout the test.
- Randomize the sequence of true and false statements.
- Make use of popular misconceptions/beliefs as false statements.
- Write items so that the incorrect response is more plausible or attractive to those without the specialized knowledge being tested.

1. Base true-false items upon statements that are absolutely true or false, without qualifications or exceptions.

Undesirable:
Nearsightedness is hereditary in origin.

Desirable:
Geneticists and eye specialists believe that the predisposition to nearsightedness is hereditary.

2. Express the item statement as simply and as clearly as possible.

Undesirable:
When you see a highway with a marker that reads, "Interstate 80" you know that the construction and upkeep of that road is maintained by the state and federal government.

Desirable:
The construction and maintenance of interstate highways are provided by both state and federal governments.

3. Express a single idea in each test item.

Undesirable:
Water will boil at a higher temperature if the atmospheric pressure on its surface is increased and more heat is applied to the container.

Desirable:
Water will boil at a higher temperature if the atmospheric pressure on its surface is increased.

4. Include enough background information and qualifications so that the ability to respond correctly to the item does not depend on some special, uncommon knowledge.

Undesirable:
The second principle of education is that the individual gathers knowledge.

Desirable:
According to John Dewey, the second principle of education is that the individual gathers knowledge.

5. Avoid the use of extreme modifiers or qualifiers.

Undesirable:

—**All** sessions of Congress are called by the President. (F)
—The Supreme Court **frequently** rules on the constitutionality of law. (T)
—An objective test is **generally** easier to score than an essay test. (T)

Desirable:

—The sum of the angles of a triangle is **always** 180°. (T)
—The galvanometer is the instrument **usually** used for the metering of electrical energy used in a home. (F)

6. Avoid lifting statements from the text, lecture or other materials so that memory alone will not permit a correct answer.

Undesirable:

For every action there is an opposite and equal reaction.

Desirable:

If you were to stand in a canoe and throw a life jacket forward to another canoe, chances are your canoe would jerk backward.

7. Avoid using negatively stated item statements.

Undesirable:

The Supreme Court is not composed of nine justices.

Desirable:

The Supreme is composed of nine justices.

8. Avoid the use of unfamiliar vocabulary.

Undesirable:

According to some politicians, the raison d'etre for capital punishment is retribution.

Desirable:

According to some politicians, justification for capital punishment is retribution.

Writing Hint...

One method for developing true-false items is to write a set of true statements that cover the content, then convert approximately half of them to false statements. **Remember:** When changing items to false (as well as in writing the true statements initially), state the items positively, avoiding negatives or double negatives.

Extreme Modifiers:

all	none
always	never
only	nobody
invariably	no one
best	absolutely
worst	absolutely not
everybody	certainly
everyone	certainly not

Qualifiers:

usually	frequently
often	sometimes
some	seldom
many	much
probably	a majority
apt to	most
might	a few
may	unlikely

Remember...

- Determine that the questions are appropriately answered by "True" or "False" rather than by some other type of response, such as "Yes" or "No."
- Arrange the statements so that there is no discernible pattern of answers (such as T , F, T, F, T, F and T, T, F, F, T, T, F, F) for True and False statements.
- Avoid the tendency to add details in true statements to make them more precise. The answers should not be obvious to students who do not know the material.
- Be sure to include directions that tell students how and where to mark their responses.

Attention Students: True-False Test Taking Tips

- When you do not know or cannot remember information to determine the truth of a statement, assume that it is true.
- There are generally more true questions on true-false exams than false questions because instructors tend to emphasize true questions.
- If there is specific detail in the statement, it may also tend to be true. For example, the statement "There are 980 endangered species worldwide" has specific detail and is likely to be true.
- Look for extreme modifiers that tend to make the question false. Extreme modifiers, such as always, all, never, or only make it more likely that the question is false.
- Identify qualifiers that tend to make the question true. Qualifiers (seldom, often, many) make the question more likely true.
- Questions that state a reason tend to be false.
- Words in the statement that cause justification or reason (since, because, when, if) tend to make the statement false because they bring in a reason that is incorrect or incomplete.

Variations in Writing True-False Test Items

The True-False-Correction Question...

In this variation, true-false statements are presented with a key word or brief phrase that is underlined. It is not enough that a student correctly identify a statement as being false. To receive credit for a statement labeled false, the student must also supply the correct word or phrase which, when used to replace the underlined part of the statement, makes the statement a true one.

This type of item is more thorough in determining whether students actually know the information that is presented in the false statements. While a student might correctly guess that a statement is false, no credit would be given unless the student could change the statement to a true one by writing word/words to replace underlined word(s).

(Continued on next page...)

Variations in Writing True-False Test Items

(continued)

The teacher decides what word/phrase can be changed in the sentence; if students were instructed only to make the statement a true statement, they would have the liberty of completely rewriting the statement so that the teacher might not be able to determine whether or not the student understood what was wrong with the original statement.

If, however, the underlined word/phrase is one that can be changed to its opposite it loses the advantage over the simpler true-false question because all the student has to know is that the statement is false and change **is** to **is not.**

The Yes-No Variation...

In the yes-no variation, the student responds to each item by writing, circling or indicating yes-no rather than true-false. An example follows:

> *What reasons are given by students for taking evening classes? In the list below, circle **Yes** if that is one of the reasons given by students for enrolling in evening classes; circle **No** if that is not a reason given by students.*

Yes	**No**	They are employed during the day.
Yes	**No**	They are working toward a degree.
Yes	**No**	They like going to school.
Yes	**No**	There are no good television shows to watch.
Yes	**No**	Parking is more plentiful at night.

The A-B Variation...

The example below shows a question for which the same two answers apply. The answers are categories of content rather than true-false or yes-no. This is another form of forced choice question because for each item the student must choose between **A** and **B.**

> *Indicate whether each type of question below is a selection type or supply type by circling **A** if it is selection, **B** if it is supply.*

Select	Supply	
A	**B**	Multiple choice
A	**B**	True-False
A	**B**	Essay
A	**B**	Matching
A	**B**	Short Answer

True-False Test Items: Conclusion

Ask yourself:

Why are these true-false questions crummy?

1. There is no advantage for not using specific determiners in true-false items. **T F**

2. Test validity is a function of test reliability, which can be improved by using fewer items. **T F**

3. A nickel is larger than a dime. **T F**

4. An eagle's range of sight is precisely 1,000 ft. **T F**

5. The telephone was invented a long time ago. **T F**

"A major distinction between the true-false test item and items in a multiple choice format, is that the true-false statement contains no criterion for answering the question. Each examinee must ask the question: *True or false with respect to what?* Each true-false item must be unequivocally true or unequivocally false. It is imperative that proper wording and the elimination of extraneous clues are more crucial with the true-false item than with any other test format."

Writing Test Items, na, Michigan State University Dept. of Education, Dec. 1999

Aim for Higher Levels of Learning

While true-false and other forced choice questions are generally used to measure knowledge and understanding, they could also be used at higher levels. Students could be provided with a set of information new to them, perhaps a portfolio, set of data, or a written work of some type, then asked various forced choice questions related to the content or the presence/absence of certain characteristics in the work.

Anticipate Scoring Ranges

Scores on true-false items tend to be high because of the ease of guessing correct answers when the answer is not known. With only two choices (true or false) the student could expect to guess correctly on half of the items for which correct answers are not known.

If a student knows the correct answers to 10 questions out of 20 and guesses on the other 10, the student could expect a score of 15. The teacher can anticipate scores ranging from approximately 50% for a student who did nothing but guess on all items to 100% for a student who knew the material.

In the final analysis...

The true-false test is probably the best known of the various types of objective test items. It is the easiest to construct and at the same time the most abused. The students learn the weaknesses that are inherent in many such items and are able to obtain high scores by noting the grammatical construction, the choice of words or other clues.

The true-false test can be used effectively as an instructional test to promote interest and introduce points for discussion. This perhaps is the most important use for the plain true-false item. It is a valuable type of test to use in giving short, daily quizzes that may be used to motivate the students for a new assignment, to review a previous lesson, to locate points to be retaught or to introduce controversial points for class discussion.

Matching Test Items

Matching questions provide a most efficient way to test knowledge in courses in which events, dates, names, and places are important. Matching questions are also appropriate for the sciences in which numerous experiments, experimenters, results, and special terms and definitions have to be remembered.

Section Summary

Good for:

- Knowledge level
- Some comprehension level, if appropriately constructed

Types:

- Terms with definitions
- Phrases with other phrases
- Causes with effects
- Parts with larger units
- Problems with solutions

Advantages:

- Maximum coverage at knowledge level in a minimum amount of space/preptime
- Valuable in content areas that have a lot of facts

Disadvantages:

- Time consuming for students
- May not be appropriate for higher levels of learning

A simple matching item consists of two columns: one column of **stems** or problems to be answered, and another column of **responses** from which the answers are to be chosen. Traditionally, the column of stems is placed on the left and the column of responses is placed on the right. An example is given below.

Directions: On the line next to each children's book in Column A print the letter of the animal or insect in Column B that is a main character in that book. Each animal or insect in Column B can be used only once.

Column A	Column B
____1. Charlotte's Web	A. Bear
____2. Winnie the Pooh	B. Chimpanzee
____3. Black Beauty	C. Cricket
____4. Tarzan	D. Deer
____5. Pinocchio	E. Horse
____6. Bambi	F. Pig

The student reads a stem (Column A) and finds the correct response from among those in Column B. The student then prints the letter of the correct response in the blank beside the stem in Column A. An alternative is to have the student draw a line from the correct response to the stem, but this is more time consuming to score.

In the above example notice that the stems in Column A are assigned numbers (1, 2, 3, etc.). The items in Column B are designated by capital letters. Capital letters are used rather than lower case letters in case some students have reading problems. Also there are apt to be fewer problems in scoring the student's handwritten responses if capital letters are used.

Also in the above example, the student only has to know five of the six answers to get them all correct. Since each animal in Column B can be used only once, the one remaining after the five known answers have been recorded is the answer for the sixth premise. One way to reduce the possibility of guessing correct answers is to list a larger number of responses than premises.

Test Your Knowledge of Matching Test Items

1. Problem: Faulty directions.

   Directions: "Place the letter of the term in the right hand column on the line to the left of the definition column."

   Circle the letter(s) that describe the best way to revise these directions:
   A. Add: “Match the following”
   B. Add: “Each term may not be used more than once”
   C. Change the order of the directions provided
   D. No changes needed

---

2. Problem: Unrelated topics.

____1. Year in which WWII began	A. Joseph Stalin
____2. British Prime Minister in WWII	B. Franklin D. Roosevelt
____3. U.S. President during WWII	C. 1939
____4. German dictator in WWII	D. Winston Churchill
	E. Adolf Hitler

   Circle the letter(s) that describe the best way to revise this matching test.

   A. Change one of the descriptions to read: “Russian dictator in WWI”
   B. Add an item to the left hand column
   C. Add a description that reads: “Year in which WWI began”
   D. Remove option C. from the right hand column
   E. Remove all stimuli and responses that do not concern leaders in WWII

---

3. Problem: Mixing matching with completion.

   Directions: On the line to the left of each statement write the letter of the atomic particles from the right hand column that the statement describes. Use each particle only once.

____1. An ____orbits the nucleus.	A. Electron
____2. Positively charged particles are called _____.	B. Neutron
____3. A _____ has no charge.	C. Protons
____4. The _____ is located in the center of an atom.	D. Nucleus
	E. Ion

   Circle the letter(s) that describe the best way to revise this matching test.
   A. Edit all the stimuli on the left to be complete statements.
   B. Remove all the blanks from the stimuli on the left.
   C. Change the order of the responses on the right.
   D. Edit the stimuli to be grammatically unbiased (i.e. singular/plural)

Answers: 1-C, 2-E, 3-A, B & D

Test Your Knowledge of Matching Test Items

4. **Directions:** The four statements presented below refer to the structure of the matching test, specifically what elements should be in Column A and what elements should be in Column B. At the left of each statement are the letters **A** and **B.** Circle **A** if Column A is the best choice; circle **B** if Column B is the best choice.

A B I. When presenting words and their definitions, which column should contain the definitions, which are longer than the words?
A B 2. Items arranged in chronological order would be found in which column?
A B 3. Premise is the term applied to the items in which column?
A B 4. Items are designated by numbers in which column?

---

5. **Directions:** For the four learning objectives listed below, decide whether a matching exercise would be an appropriate method of assessment (Assume that you can construct a list of 6-8 items for the matching question.) Circle **YES** if appropriate; circle **NO** if not appropriate.

YES **NO** A. The student will be able to recognize the cities in/near which the major battles in the American Revolution took place.
YES **NO** B. The student will be able to differentiate between words that are spelled correctly and those spelled incorrectly.
YES **NO** C. The student will be able to identify the elements with their symbols from the periodic table.
YES **NO** D. The student will be able to identify the English words for various fruits that are represented by their Spanish language counterparts.

---

6. **Directions:** On the lines following this matching question supply four recommendations to improve this question.

Hitler	A. Year in which WWII began
Mulroney	B. A Canadian Prime Minister
Tank	C. A German dictator during the WWII
1939	D. An armored vehicle used originally to break the trench war stalement in WWI

Recommendations:

1._____

2._____

3._____

4._____

Answers: 4-A, B, A, A; 5-Yes, No, Yes, Yes; 6-Examples: Need directions, reverse Column A and Column B, make items similar, increase the number of responses

Suggestions For Writing Matching Test Items

Remember...

- Review your teaching objectives to make sure that a matching component is appropriate.
- Keep matching items brief, limiting the list of stimuli to 10 - 15.
- When possible, reduce the amount of reading time by including only short phrases or single words in the response list.
- Use the more involved expressions in the stem and keep the responses short and simple.
- Arrange the list of responses in some systematic order if possible (chronological, alphabetical).
- Make sure that there are never multiple correct responses for one stem (although a response may be used as the correct answer for more than one stem).
- Avoid breaking a set of items (stems and responses) over two pages. (Students go nuts flipping pages.)

1. Include directions which clearly state the basis for matching the stimuli with the responses.

Explain whether or not a response can be used more than once and indicate where to write the answer.

Undesirable:

Directions: Match the following.

Desirable:

Directions: On the line to the left of each identifying location and characteristics in Column I, write the letter of the country in Column II that is best defined. Each country in Column II may be used more than once.

2. Use only items that share the same foundation of information.

Unrelated topics included in the same matching item may allow for obvious matches and mismatches.

Undesirable:

Directions: Match the following.

1. Water	A. NaCl
2. Discovered Radium	B. Fermi
3. Salt	C. NH3
4. Ammonia	D. 1942
5. Year of the first Nuclear Fission	E. H_20
	F. Curie
	G. 1957

Desirable:

Directions: On the line to the left of each compound in Column I, write the letter of the compound's formula presented in Column II. Use each formula only once.

Column I	**Column II**
____1. Water	A. H_2S0_4
____2. Salt	B. HCl
____3. Ammonia	C. NaCl
____4. Sulfuric Acid	D. H_20
	E. H_2HCl

3. Avoid grammatical or other clues to the correct response.

Undesirable:

Directions: Match the following in order to complete the sentences on the left.

____1. Plato insisted that government was	A. The Prince.
____2. Machiavelli wrote about achieving political unity in	B. desirable and inevitable
____3. Hobbes argued that human nature made absolute monarchy	C. a science requiring experts.
____4. Marx was a German philosopher and economist who founded	D. organized along industrial lines.
	E. Communism.

Desirable:

Directions: On the line to the left of each statement write the letter of the philosopher from the right hand column that the statement describes. Use each philosopher once.

____1. Thought government was a science requiring experts.	A. Hobbes
____2. Described methods of achieving political unity.	B. Marx
____3. Founded Communism.	C. Machiavelli
____4. Believed that human nature made absolute monarchy desirable and inevitable	D. Durkheim
	E. Plato

4. The column of stimuli on the left should set the question clearly.

Undesirable:

Directions: Match the following.

____1. City dwellers	A. Wild animals
____2. Hunter-gatherers	B. Farm
____3. Pastoral nomads	C. Apartment buildings
	D. Graze animals

Desirable:

Directions: On the line to the left of each definition, write the letter of the term in the right hand column that is defined. Use each term only once.

____1. Live in areas of high population density.	A. Pastoral nomads
____2. Move from one place to another in search of wild animals.	B. Ranchers
____3. Move from one place to another with grazing animals.	C. Hunter-gatherers
____4. Till land for cash crops.	D. City dwellers
	E. Farmers

Attention Students: Matching Test Taking Tips

- Read the directions. There are usually two lists that need to be matched. Take a look at both lists to get a feel for the relationships and build your confidence.
- Use one list as a starting point and go through the second list to find a match. This process organizes your thinking. It will also speed your answers because you become familiar with the second list and will be able to go straight to a match that you saw when looking through the lists a previous time.
- Move through the entire list before selecting a match. If you make a match with the first likely answer, you may make an error, because an answer later in the list may be more correct.
- Cross off items on the second list when your are certain that you have a match. This seems simplistic, but it helps you feel confident and stay organized.
- Do not guess until all absolute matches have been made. If you guess early in the process, you will likely eliminate an answer that could be used correctly for a later choice.

How to Study For a Matching Test

If your instructor usually includes a matching section in a typical exam, here is one way to prepare for it. As you read the textbook, be alert for facts and ideas that are associated with people's names. On a separate sheet, list the names and facts opposite each other, resulting in two distinct vertical columns, as in the following example.

NAMES	FACTS OR IDEAS	SUBJECT
Susan B. Anthony	Women's movement	Sociology
Jack London	Call of the Wild	Literature
George Washington Carver	Agricultural chemist	Science
Lewis and Clark	American explorers	History
George A. Miller	Magic number seven	Psychology
William James	Pragmatism	Philosophy
Mozart	Marriage of Figaro	Music

To master your list, cover the fact column with a sheet of paper. Look at each item in the name column, and recite and write the corresponding fact or idea. Then, to make sure that you learn the material both ways, block out the name column and use the facts as your clues. The example given above includes items from various subject areas.

Variations for Creating Matching Tests

Keylists or Masterlists Example

A (Na^{11})	B (Cl^{17})	C (H^{1})
22.9898	35.453	1.00797
I	± 1, 3, 5, 7	I
(Ne)35	(Ne) $3s^2 3p^5$	Is'
892	-34.7	-252.7
97.5	-102	-259.2
0.97	1.56	0.071

Refer to the chemical symbols above to answer the following:

___1. Which of the above elements has the largest atomic weight?
___2. Which of the above elements has the largest atomic number?
___3. Which of the above elements has the lowest boiling point?
___4. Which of the above elements has the lowest melting point?
___5. Which of the above elements has the highest density?
___6. Which of the above elements has the least number of electrons?
___7. Which of the above elements has the least number of protons ?
___8. Which of the above elements represents chlorine?
___9. Which of the above elements represents sodium?

Ranking Example

TOPIC: Social Studies, Western Civilization

Directions: Number (1-8) the following events in the history of ancient Egypt in the order in which they occurred, using 1 for the earliest event.

_____Egypt divided; ruled by Libyan kings, Nubian pharaohs, Assyrians, and Persians
_____Seizure of power by Hyksos kings
_____Upper and Lower Egypt are united by Menes
_____Alexander the Great conquers Egypt
_____Reunification of Egypt under pharaoh Mentuhotop II
_____Rise of feudal lords leads to anarchy
_____Thutmose III expands empire to the Euphrates
_____Many kings with short reigns; social and political chaos

Note in the above example the implied column of responses is 1, 2, 3, 4, 5, 6, 7, 8.

Aiming for Higher Order Thinking Skills

Usually matching items measure recognition of factual **knowledge** rather than higher order thinking skills such as analysis and synthesis. This does not mean, however, that variations cannot be constructed to aim for higher levels of thinking.

One variation, presented below, combines elements of a multiple choice test item with a matching component.

Item Components

A. Correct answer(s)
B. foil(s)
C. option(s)
D. stem(s)

___1. The components of a multiple choice item are a

___2. a. stem and several foils.
___3. b. correct answer and several foils.
___4. c. stem, a correct answer, and some foils.
___5. d. stem and a correct answer
___6.

In the above example it is necessary to answer the multiple choice item in order to answer the matching item. Note also that the responses (item components) in the list at the top have an (s) added to each response to eliminate singular-plural clues.

Answers: 1-D, 2-B, 3-B, 4-A, 5-B, 6-C

Completion or Fill-in-the-Blank Test Items

Section Summary

Good for:

- Knowledge levels
- Recall and memorization of facts

Advantages:

- Good for **who, what, where, when** content
- Minimizes guessing
- Encourages more intensive study. Student must know the answer vs. recognizing the answer.
- Can usually provide an objective measure of student achievement or ability

Disadvantages:

- Difficult to assess higher levels of learning because the answers to completion items are usually limited to a few words
- Difficult to construct so that the desired response is clearly indicated
- May overemphasize memorization of facts
- Questions may have more than one correct answer
- Scoring is time consuming

No-Hint Test Construction

Completion items are especially useful in assessing mastery of factual information when a specific word or phrase is important to know. They preclude the kind of guessing that is possible on limited-choice items since they require a definite response rather than simple recognition of the correct answer. Because only a short answer is required, their use on a test can enable a wide sampling of content.

A completion item requires the student to answer a question or to finish an incomplete statement by filling in a blank with the correct word or phrase. For example,

> **According to Freud, personality is made up of three major systems, the_____, the _____ and the _____.**

What About Synthesis and Evaluation?

Completion items tend to test only rote, repetitive responses and may encourage a fragmented study style since memorization of bits and pieces will result in higher scores. They are more difficult to score than forced-choice items and scoring often must be done by the test writer since more than one answer may have to be considered correct.

Is Short Answer the Same Thing?

A distinction should be made between completion—often referred to as fill-in-the-blank—and short answer questions. With completion questions the response is usually one or two words that fit on a line provided by the tester. Short answer questions may require one sentence or even a paragraph to fully answer the question.

Short answer questions are appropriate in measuring a student's understanding of principles or the ability to solve problems or apply principles. Short answer questions go beyond simple recall or recognition. They require students to consider various factors and to arrive at solutions, whether they deal with mathematical or other situations.

(Continued on next page...)

Short Answer **(continued from previous page...)**

Strategies for developing short answer questions are similar to those concerning completion but have an added dimension requiring strategies appropriate for essay questions. As an example, scoring completion questions can be more *objective* than scoring short answer questions which require a subjective *interpretation* on the teachers part. The information contained in this section primarily focuses on **completion** or **fill-in-the-blank** questions.

On the whole, completion test items have little advantage over other item types unless the need for specific recall is essential.

Test Your Knowledge of Completion Items

Directions: Fill in the blanks.

1. A fill-in-the-blank question asks students to supply rather than _____ the answer.
2. The main problem in constructing completion items is to limit the number of possible _____.
3. Put blanks at the _____ of the statement rather than the _____.
4. Completion items are faster to answer than _____ items because there are no alternatives to consider.
5. Make the _____ of equal length.
6. A direct _____ is often more desirable than an incomplete _____.
7. When doing fill-in-the-blank test items, read the _____ with the intent to give an answer that is _____ correct.
8. Always concentrate on the _____ of blanks to fill in.
9. When you do not know the exact _____, provide a descriptive answer.
10. Scoring completion items is less _____ than multiple choice or true-false because the student supplies the response.

Answers: 1-select; 2-answer; 3-end, beginning; 4-multiple choice; 5-blanks; 6-question, sentence; 7-question, grammatically; 8-number; 9-response; 10-objective

Suggestions For Writing Completion Test Items

Remember...

- When possible, provide explicit directions as to what amount of variation will be accepted in the answers.
- Give much more credit for completions than for true-false or matching items.
- Avoid using a long quote with multiple blanks to complete.
- When working with definitions, supply the term, not the definition, for a better judge of student knowledge.
- For numbers, indicate the degree of precision/units expected.
- Facilitate scoring by having the students write their responses on lines arranged in a column to the left of the items.

Suggestion...
It is difficult to write completion items so that there is only one correct answer. When preparing a scores key, list the correct answer and any other acceptable alternatives. Be consistent in using the key; it would not be fair to accept an answer as right on one paper and not accept it on others.

1. Omit only significant words from the statement.

Undesirable:
Every atom has a central _____ called a nucleus.

Desirable:
Every atom has a central core called a(n) _____.

2. Do not omit so many words from the statement that the intended meaning is lost.

Undesirable:
The _____ were to Egypt as the _____ were to Persia and as _____ were to the early tribes of Israel.

Desirable:
The Pharaohs were to Egypt as the _____ were to Persia and as _____ were to the early tribes of Israel.

3. Avoid obvious clues to the correct response.

Undesirable:
Most of the United States' libraries are organized according to the _____ decimal system.

Desirable:
Which organizational system is used by most of the United States' libraries? _____.

4. Be sure there is only one correct response.

Undesirable:
Trees which shed their leaves annually are _____.

Desirable:
Trees which shed their leaves annually are called _____.

5. Avoid grammatical clues to the correct response.

If the indefinite article is required before a blank, use **a(n)** so that the student does not know if the correct answer begins with a vowel or a consonant.

Undesirable:
A subatomic particle with a negative electric charge is called an _____.

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

6. If possible, put the blank at the end of a statement rather than at the beginning.

Asking for a response before the student understands the intent of the statement can be confusing and may require more reading time.

Undesirable:
_____ is the measure of central tendency that is most affected by extremely high or low scores.

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

Ask yourself:

Why are these completion items crummy?

- The_____ of _____ took place in the year _____.
- _____ was a crucial event to German history.

Beware of Clever Students

Nudity, *infancy*, and *bliss* are some of the answers for the following completion item:

- George Washington was born in the state of _____.

Attention Students: Completion Test Taking Tips

- Read the question with the intent to give an answer and make the sentence grammatically correct. In this process it is important to focus on how the sentence is written. For example, if the blank is preceded by the article "an," you know the word that goes in the blank must start with a vowel.
- Concentrate on the number of blanks in the sentence and the length of the space. The test maker is giving you clues to the answer by adding spaces and making them longer.
- Provide a descriptive answer when you cannot think of the exact word or words. The instructor will often give you credit or partial credit when you demonstrate that you have studied the material and can give a credible answer, even when you have not given the exact words.

Essay Test Items

Essay tests present a realistic task to the student. In real life, a person is required to organize and communicate thoughts rather than respond to multiple choice questions.

Section Summary

Good for:

- Application, synthesis and evaluation levels

Types:

- Extended response: synthesis and evaluation levels; a lot of freedom in answers
- Restricted response: more consistent scoring, outlines parameters of responses

Advantages:

- Students less likely to guess
- Easy to construct
- Stimulates more study
- Allows students to demonstrate ability to organize knowledge, express opinions, show originality.

Disadvantages:

- Can limit amount of material tested, therefore has decreased validity.
- Subjective, potentially unreliable scoring.
- Time consuming to score.

A typical essay test usually consists of a small number of questions to which the student is expected to recall and organize knowledge in logical, integrated answers. An essay test item can be an extended response item or a short answer item. An example of each type follows.

Extended Response

Compare the writings of Bret Harte and Mark Twain in terms of settings, depth of characterization, and dialogue styles of their main characters. (10 pts. 20 minutes)

Short Answer

Identify research methods used to study the S-R (Stimulus-Response) and S-O-R (Stimulus-Organism-Response) theories of personality. (5 pts. 10 minutes)

The Benefits of Essay Tests

The main advantages of essay and short answer items are that they permit students to demonstrate achievement of such higher level objectives as analyzing and critical thinking. Written items offer students the opportunity to use their own judgment, writing styles, and vocabularies. They are less time consuming to prepare than any other item type.

Research indicates that students study more efficiently for essay type examinations than for selection (multiple choice) tests. Students preparing for essay tests focus on broad issues, general concepts, and interrelationships rather than on specific details. This studying results in somewhat better student performance regardless of the type of exam they are given. Essay tests also give the instructor an opportunity to comment on students' progress, the quality of their thinking, the depth of their understanding, and the difficulties they may be having.

Unfortunately...

Essay tests consisting only of written items permit only a limited sampling of content learning due to the time required for students to respond. Essay items are not efficient for assessing knowledge of basic facts and provide students more opportunity for bluffing, rambling, and *snowing* than limited choice items. They favor students who possess good writing skills and neatness. They are pitfalls for students who tend to go off on tangents or misunderstand the main point of the question.

Essay question:

"Discuss the importance of the nature/nurture controversy in the shaping of current developmental theory."

Student answer:

"The nature/nuture *(sic)* controversy was very impotent *(sic)* in shaping current developmental theory becuse *(sic)* it was needed to help people who were doing work in that area to come up with their current theories."

I'd like to use essay tests, but...

Marilla Svinicki, University of Texas at Austin
The Professional & Organizational Development Network in Higher Education

Do you cringe when you read the kind of tortured prose and fractured thinking that is represented by the example on the left? Or plow through paragraph after paragraph of detail in a student's answer in hopes of finding an original thought.

It has been our habit in the past to blame the students, the school system and the English department for not teaching our students how to write a solidly argued, concisely worded essay answer. But we must face the fact that we are as much to blame for their imprecise prose as those other entities.

Part of the problem may lie in the way instructors help (or fail to help) students prepare for writing essay tests. Learning specialists have known for a long time that the kind of preparation needed for responding to essay questions is different from that needed for objective tests. Unfortunately, many of our students prepare for all exams with the same learning strategies, and then are ill-equipped to tackle the kind of thinking needed during essay tests.

*(See **page 56** for a continuation of this essay.)*

Read 'Em and Weep Essay Test Items*

History: Describe the history of the papacy from its origins to the present day, concentrating especially, but not exclusively, on its social, political, economic, religious, and philosophical impact on Europe, Asia, America, and Africa. Be brief, concise, and specific.

Biology: Estimate the differences in subsequent human culture if our form of life had developed 500 million years earlier, with special attention to its probable effect on the English parliamentary system. Prove your thesis.

Psychology: Based on your knowledge of their works, evaluate the emotional stability, degree of adjustment, and repressed frustrations of each of the following: Alexander the Great, Rameses II, Gregory of Nicia, Hammurabi. Support your evaluation with quotations from each man's work, making appropriate references. It is not necessary to translate.

Sociology: Estimate the sociological problems which might accompany the end of the world. Construct an experiment to test your theory.

Economics: Develop a realistic plan for refinancing the national debt. Trace the possible effects of your plan in the following areas: Cubism, the Donatist controversy, the wave theory of light.

Physics: Explain the nature of matter. Include in your answer an evaluation of the impact of the development of mathematics on science.

Philosophy: Describe everything in detail. Be objective and specific.

*Just kidding

Test Your Ability to Construct Essay Test Items Requiring Higher Order Thinking Skills

Directions: In the blank to the left of the essay item statement, place the letter of the learning level that is best indicated by the words contained in this statement.

A = Knowledge **B = Comprehension** **C = Application**
D = Analysis **E = Synthesis** **F = Evaluation**

_____1. Essay items may begin with *modify, prepare,* or *solve.*

_____2. Essay items may begin with *define, label, outline,* or *state.*

_____3. Essay items may begin with *convert, predict,* or *estimate.*

_____4. Essay items may begin with *appraise, interpret,* or *criticize.*

_____5. Essay items may begin with *categorize, compile,* or *re-arrange.*

_____6. Essay items may begin with *diagram, illustrate,* or *separate.*

Directions: Use the above letters to identify the learning level of the essay test items listed below. Place your answer in the blank to the left of each item.

_____1. In the president's State of the Union Address, which statements are based on facts and which are based on assumptions?

_____2. How would you restructure the school day to reflect children's developmental needs?

_____3. Why is Bach's Mass in B Minor acknowledged as a classic?

_____4. Calculate the deflection of a beam under uniform loading

_____5. Summarize the basic tenets of deconstructionism.

_____6. List the steps involved in titration.

Directions: Write a test item for each learning level.

My topic:_____

1. Knowledge_____
_____

Example Topic: Asbestos: What is asbestos?

2. Application_____
_____

Example Topic: Asbestos: Consider the crystal structures of chrysotile and crocidolite. Why should the most common mineral be the less hazardous?

3. Synthesis_____
_____

Example Topic: Asbestos: Design a study to reasonably demonstrate the dangers posed by asbestos to the general populace.

4. Evaluation_____
_____

Example Topic: Asbestos: The “asbestos hazard” is either (1) nothing more than a costly bureaucratic creation or (2) a hazard that accounts for tens of thousands of deaths annually. Which of the two controversial arguments has the best scientific support?

Suggestions For Writing Essay Test Items

Standard Phrases for Writing Essay Test Items

Agreement or Disagreement: The student is being asked to assert and support a thesis with evidence.

Analyze: Analyzing is a picking apart of the whole.

Classification and Division: Grouping items into a category according to a consistent principle.

Compare/Contrast: Comparing shows similarities, while contrasting points out differences.

Cause and Effect: Establishes a link between two things and also to describe the outcome.

Define: Consists of three parts: term, class, and differentiating characteristics.

Define and give an example of: Asks students to not only define the term, but to supply an example.

Describe: Requires students to explain something in detail.

Discuss: Too vague and may elicit vague, overgeneralized, unsupported responses.

Illustrate: Give examples and/or analogies to demonstrate a particular process/idea or steps in a series.

Summarize: The overall view of some process, speech, play, concept, etc.

1. Formulate the question so that the task is clearly defined for the student.

Use words that *aim* the student to the approach you want them to take. Words like discuss and explain can be ambiguous. If you use *discuss,* then give specific instructions as to what points should be discussed.

Undesirable:

Discuss Karl Marx's philosophy.

Desirable:

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

2. Pay attention to the number of items.

In order to obtain a broader sampling of course content, use a relatively large number of questions requiring shorter answers (one-half page) rather than just a few questions involving long answers (2-3 pages).

3. Avoid the use of optional questions on an essay test.

When students answer different questions, they are actually taking different tests. If there are five essay questions and students are told to answer any three of them, then there are ten different tests possible. It makes it difficult to discriminate between the student who could respond correctly to all five, and the student who could answer only three.

Use of optional questions also affects the reliability of the scoring. If we are going to compare students for scoring purposes, then all students should perform the same tasks. Another problem is that students may not study all the course material if they know they will have a choice among the questions.

(Continued on next page...)

5. Write essay items at different levels of learning.

The goal is to write essay items that measure higher cognitive processes. The item should represent a situation that tests the student's ability to use knowledge in order to analyze, justify, explain, contrast, evaluate, and so on.

Try to use verbs that elicit the kind of thinking you want the students to demonstrate. Instructors often have to use their best judgment about what cognitive skill each question is measuring. Ask a colleague to read the questions and classify them according to Bloom's taxonomy.

5. Choose a scoring model.

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

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

5. Prepare students to take essay exams.

Essay tests are valid measures of student achievement only if students know how to take them. Many college freshmen do not know how to take an essay exam, because they have not been required to learn this skill in high school.

Take some class time to tell students how to prepare for and how to take an essay exam. Use old exam questions and let students see what an "A" answer looks like and how it differs from a "C" answer.

Remember...

- Make essay questions comprehensive rather than focused on small units of content.
- Provide clear directions as to the expectations.
- Allow students an appropriate amount of time. (It is helpful to give students some guidelines on how much time to use on each question, as well as the desired length and format of the response, such as full sentences, phrases only, outline, etc.)
- Inform students, in advance of answering the questions, of the proportional value of each item in comparison to the total grade.
- Require students to demonstrate command of background information by asking them to provide supporting evidence for claims and assertions.
- Students should be informed about how you treat such things as misspelled words, neatness, handwriting, grammar, etc.
- Decide how to treat irrelevant or inaccurate information contained in students' answers.
- Write comments on the students' answers. Teacher comments make essay tests a good learning experience for students. Comments serve to refresh your memory should the student question the grade.

A Four-Step Process in Grading Essay Test Items

Step One

When the assignment is given...

- Figure out what the purpose of the assignment is, and generate grading criteria based upon that purpose.
- Share the criteria you decide upon with your students: hand it out in class, and post it on your door.
- Provide models of your grading criteria to your students.

Step Two

When the assignments are turned in...

- Quickly overview a percentage of the papers to get an overall sense of how the group did on the assignment.
- Skim some papers that you feel are representative of the range of quality in the student work.
- Use these papers to start four piles: High, Medium High, Medium Low, and Low.

Step Three

Start the grading...

- Always use a pencil on your first run through: as you develop your sense of how the students did, you will probably go back and fine-tune the papers you graded first!
- Having separated the papers into piles (high, medium, low: not letter grades yet), do an initial read through and assign a preliminary, holistic grade based upon a general impression of the work. Do not get bogged down in details yet, short of marking a plus (+) or minus (-) in the margins next to issues that strike you.
- Now re-read each paper for how it addresses the criteria identified for the assignment. Two papers may address the same criteria differently. Focus first on what the paper does, before getting to what it does not. After a sympathetic read, give it a critical read.

Step Four

Mark up the papers...

- Interactive grading poses questions and presents problems the student needs to resolve. For example: "Is this (x) what you mean? How does this connect to your main point?"

Research Shows That a Number of Factors Can Bias the Grading of Essay Tests

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

Attention Students:
Essay Test Taking Tips

1. Organize your thoughts before you begin to write

A short outline on a separate piece of paper will improve your thinking. There is usually a main idea or issue, several supporting issues and examples to illustrate the issues.

2. Paraphrase the original question to form your introductory statement

This benefits you in two ways. First, it helps you get the question straight in your mind. Second, it may protect you from the teacher. If you have rephrased the question, the teacher can see how you understood the question. Perhaps you understood it to mean something other than the teacher intended. If so, the teacher may give you credit for seeing another perspective.

3. Remember: Neatness counts!

Write your answer clearly, so the reader will be able to decode your writing and understand your ideas. Without clearly written words your chances of a good grade are severely diminished. Write or print clearly, using a dark-colored erasable ball point pen. Avoid crossing out words or sentences, and do not smudge your paper.

4. Verb alert

Read each essay question with the intent to identify the verbs or words that give you direction. These are the verbs that describe the task you are expected to complete. Circle the direction verbs in the question to make sure that you are focusing on the desired task.

5. Use the principles of good English composition

Form a clear thesis statement (statement of purpose) and place it as near to the beginning as possible. Provide supporting issues to back up the main concept you present. Underline or highlight the main and supporting issues. Examples will improve your answers and set them apart from other students' answers. Remember to save some space for a brief but adequate summary.

Additional Types of Test Items

Advantages in Using Problem Solving Items

- Minimize guessing by requiring the students to provide an original response rather than to select from several alternatives.
- Easier to construct than are multiple choice or matching items.
- Can most appropriately measure objectives which focus on the ability to apply skills or knowledge in the solution of problems.
- Can measure an extensive amount of content or objectives.

Limitations in Using Problem Solving Items

- Require an extensive amount of instructor time to read/grade.
- Subject to scorer bias when partial credit is given.

Remember...

- State in the directions whether or not the student must show the work procedures for full or partial credit.
- Ask questions on which experts could agree that one solution and one or more work procedures are better than others.
- Work through each problem before classroom administration to double check accuracy.

Problem Solving

An essay is not the only form of a subjective test item. Another form is the problem solving or computational exam question. Such items present the student with a problem situation or task and require a demonstration of work procedures. Problem solving is classified as subjective due to the procedures used to score item responses. Instructors can assign full or partial credit to either correct or incorrect solutions depending on the quality and kind of work procedures presented. An example of a problem solving test item follows:

> *It was calculated that 75 men could complete a strip on a new highway in 70 days. When work was scheduled to commence, it was found necessary to send 25 men on another road project. How many days longer will it take to complete the strip? Show your work for full or partial credit.*

Suggestions for Writing Problem Solving Test Items

1. Provide directions which clearly inform the student of the type of response called for.

Undesirable:

An American tourist in Paris finds that he weighs 70 kilograms. When he left the United States he weighed 144 pounds. What was his net change in weight?

Desirable:

An American tourist in Paris finds that he weighs 70 kilograms. When he left the United States he weighed 144 pounds. What was his net weight change in pounds?

2. Separate item parts and indicate their point values.

A man leaves his home and drives to a convention at an average rate of 50 miles per hour. Upon arrival, he finds a telegram advising him to return at once. He catches a plane that takes him back at an average rate of 300 miles per hour.

Desirable:

If the total traveling time was 1¾ hours:

(1) How long did it take him to fly back? (1 pt.)

(2) How far from his home was the convention? (1 pt.)

Show your work for full or partial credit.

Authentic Assessments

How well do multiple choice tests really evaluate student understanding and achievement? Many educators believe that there is a more effective assessment alternative. These teachers use testing strategies that do not focus entirely on recalling facts. Instead, they ask students to demonstrate skills and concepts they have learned. This strategy is called authentic assessment.

Authentic assessment aims to evaluate students' abilities in 'real-world' contexts. In other words, students learn how to apply their skills to authentic tasks and projects. Authentic assessment goes beyond rote learning and passive test-taking. Instead, it focuses on students' analytical skills; ability to integrate what they learn; creativity; ability to work collaboratively; and written and oral expression skills.

Use Authentic Assessment For...

1. Performance Tests

Performance tests assess students' ability to use skills in a variety of authentic contexts. They frequently require students to work collaboratively and to apply skills and concepts to solve complex problems. Short- and long-term tasks may include:

- writing, revising, and presenting a report to the class
- conducting a week-long science experiment and analyzing the results
- working with a team to prepare a classroom debate

In theory, a performance test could be constructed for any skill and real life situation. In practice, most performance tests have been developed for the assessment of vocational, managerial, administrative, leadership, communication, interpersonal and physical education skills in various simulated situations.

Suggestions for Writing Performance Test Items

- Prepare items that elicit measurable behavior
- Clearly identify and explain the simulated situation
- Make the simulated situation as *life-like* as possible
- Directions should indicate the type of response called for
- Clearly state time and activity limitations in the directions

Authentic assessment values the learning process as much as the finished product.

Examples...

In authentic assessment, students:

- do science experiments
- conduct social-science research
- write stories and reports
- read and interpret literature
- solve math problems that have real-world applications

Advantages in Using Authentic Assessment

- Can measure objectives related to the ability of the students to apply skills or knowledge in real life situations.
- Provide a degree of test validity not possible with standard paper and pencil test items.
- Useful for measuring objectives in the psychomotor domain.

Limitations in Using Authentic Assessment

- Difficult and time consuming to construct.
- Primarily used for testing students individually and not for testing groups. Consequently, they are relatively costly, time consuming, and inconvenient forms of testing.

Use Authentic Assessment For...

(continued)

2. Short Investigations

Many teachers use short investigations to assess how well students have mastered basic concepts and skills. Most short investigations begin with a stimulus: a math problem, cartoon, map, or excerpt from a primary source.

The teacher may ask students to interpret, describe, calculate, explain, or predict. These investigations may use enhanced multiple choice questions. Or they may use concept mapping, a technique that assesses how well students understand relationships among concepts.

3. Open Response Questions

Open response questions, like short investigations, present students with a stimulus and ask them to respond. Responses include:

- a brief written or oral answer
- a mathematical solution
- a drawing
- a diagram, chart, or graph

4. Portfolios

A portfolio documents learning over time. This type of authentic assessment is a purposeful collection of work that shows the achievement or growth of a student. A portfolio is not a specific test but rather a cumulative collection of a student's work.

Students decide what examples to include that characterize their growth and accomplishment over the term. While most common in composition classes, portfolios are beginning to be used in other disciplines to provide a fuller picture of a students' achievement.

This long-term perspective accounts for student improvement and teaches students the value of self-assessment, editing, and revision. A student portfolio may include:

- journal entries and reflective writing
- peer reviews
- artwork, diagrams, charts, and graphs
- group reports
- student notes and outlines
- rough drafts and polished writing
- electronic, video, and/or digital items

For additional information...

Using Portfolios for Authentic Assessment

For a free copy contact...

Kansas Curriculum Center
(785) 231-1010 x1534
esperanza.root@washburn.edu

Grading Authentic Assessments

A rubric enhances the quality of authentic assessment...

Advantages in Using Rubrics

- Many experts believe that rubrics improve students' end products and therefore increase learning.
- When teachers evaluate papers or projects, they know implicitly what makes a good final product and why.
- When students receive rubrics beforehand, they understand how they will be evaluated and can prepare accordingly.
- Developing a grid and making it available as a tool for students' use will provide a frame of reference for the self-assessment of the quality of their work.

Once a Rubric is Created...

An established rubric can be used or modified and applied to many activities. For example, the standards for excellence in a writing rubric remain constant throughout the school year; what does change is students' competence and your teaching strategy. Because the essentials remain constant, it is not necessary to create a completely new rubric for every activity.

Many educators find that authentic assessment is most successful when students know *just what is expected.* For this reason, teachers should clearly define standards and expectations. Educators use **rubrics,** or established sets of criteria, to assess a student's work.

A rubric is a scoring guide that seeks to evaluate a student's performance based on the sum of a full range of criteria rather than a single numerical score. Rubrics can be created in a variety of forms and levels of complexity, however, they all contain common features which:

- focus on measuring a stated objective (performance, behavior, or quality)
- use a range to rate performance
- contain specific performance characteristics arranged in levels indicating the degree to which a standard has been met

Rubrics can be created for any content area including math, science, history, writing, foreign languages, drama, art, music, and even cooking! Once developed, they can be modified easily for various grade levels. The following information presents general guidelines for developing a rubric. To Illustrate the various steps a sample rubric is used. This rubric was created by a group of postgraduate education students at the University of San Francisco, but could be developed easily by a group of elementary students.

Steps in Rubric Development

1. Determine Learning Outcomes

- Determine which concepts, skills, or performance standards you are assessing
- List the concepts and rewrite them into statements which reflect both cognitive and performance components
- Identify the most important words within the concepts or skills being assessed in the task

Chocolate Chip Cookie Rubric

The cookie elements the students chose to judge were:

- Number of chocolate chips
- Texture
- Color
- Taste
- Richness (flavor)

2. Determine Measurable Criteria

- On the basis of the purpose of the task, determine the number of points to be used for the rubric (example: 4-point scale or 6-point scale)
- Starting with the desired performance, determine the description for each score remembering to use the importance of each element of the task or performance to determine the score or level of the rubric

Chocolate Chip Cookie Rubric

The students developed a 4-point scale with the following descriptions:

- Delicious (4)
- Good (3)
- Needs improvement (2)
- Poor (1)

The measurable criteria for each point of the scale follows:

4—Delicious:

- Chocolate chip in every bite
- Chewy
- Golden brown
- Home-baked taste
- Rich, creamy (high-fat flavor)

3—Good:

- Chocolate chips in about 75 percent of the bites taken
- Chewy in the middle, but crispy on the edges
- Either brown from overcooking, or light from being 25 percent raw
- Quality store-bought taste (medium-fat content)

2—Needs Improvement:

- Chocolate chips in 50 percent of the bites taken
- Texture is either crispy from overcooking or does not hold together because it is at least 50 percent uncooked
- Color is either dark brown from overcooking or light from undercooking
- Tasteless (low-fat content)

1—Poor:

- Too few or too many chocolate chips
- Texture resembles a dog biscuit
- Burned
- Store-bought flavor with a preservative aftertaste—stale, hard, chalky nonfat contents

Terms to Use in Measuring Range/Scoring Levels

- Needs Improvement
  Satisfactory
  Good
  Exemplary
- Beginning
  Developing
  Accomplished
  Exemplary
- Needs work
  Good
  Excellent
- Novice
  Apprentice
  Proficient
  Distinguished
- Numeric scale ranging from 1 to 5, for example

Concept Words That Convey Various Degrees of Performance

- Presence to absence
- Complete to incomplete
- Many to some to none
- Major to minor
- Consistent to inconsistent
- Frequency: always to generally to sometimes to rarely

As students become familiar with rubrics, they can assist in the rubric design process. This involvement empowers the students and as a result, their learning becomes more focused and self-directed. Authentic assessment, therefore, blurs the lines between teaching, learning, and assessment.

Steps in Rubric Development

(continued)

3. Develop a Grid

Criteria	4-Delicious	3-Good	2-Needs Improvement	1-Poor
Number of chips	Chocolate chip in every bite	Chips in about 75% of bites	Chocolate in 50% of bites	Too few or too many chips
Texture	Chewy	Chewy in middle, crisp on edges	Texture either crispy/crunchy or 50% uncooked	Texture resembles a dog biscuit
Color	Golden brown	Either brown from overcook-ing or light from being 25% raw	Either dark brown from overcooking or light from undercooking	Burned
Taste	Home-baked taste	Quality store-bought taste	Tasteless	Store-bought flavor, preservative aftertaste—stale, hard, chalky
Richness	Rich, creamy, high-fat content	Medium fat contents	Low-fat contents	Nonfat contents

4. Compare Student Work to the Rubric

- Assign a rating to the various criteria you have identified as important.
- Revise the rubric descriptions based on performance elements reflected by the student work that you did not capture in your original rubric
- Rethink your scale: Does the number of points differentiate enough between types of student work to satisfy you?
- Adjust the scale if necessary. Reassess student work and score it against the developing rubric.

To assist in the initial development of a rubric, sample criteria (on a 5-0 point score range) are presented on the next page.

	Criteria for Scoring
5	• This is the highest rating • The student is extremely knowledgeable about the topic • The student demonstrates in-depth understanding of important ideas • The student shows a depth of understanding of important relationships • The answer is fully developed and includes specific facts or examples • The answer is organized around big ideas, major concepts/principles • The response is exemplary, detailed and clear
4	• The student is knowledgeable about the topic • The student has a good understanding of the topic • The student includes some of the important ideas related to the topic • The student shows a good understanding of the important relationships • The answer demonstrates includes adequate supporting facts or examples • The answer demonstrates some organization around big ideas, major concepts/principles • The response is good, has some detail, and is clear
3	• This is the middle score of the scale • The student demonstrates some know ledge and understanding of the topic. The overall answer is OK but may show apparent gaps in his/her understanding and knowledge • The student includes some of the important ideas related to the topic • The student shows some but limited understanding of the relationships • The answer demonstrates satisfactory development of ideas and includes some supporting facts or examples • The response is satisfactory, containing some detail, but the answer may be vague or not well developed and may include misconceptions or some inaccurate information
2	• The student has little knowledge or understanding of the topic • The student does not develop the ideas or deal with the relationships among the ideas • The response contains misconceptions or inaccurate information • The student may rely heavily on the group activity • The response is poor and lacks clarity
1	• The student shows no knowledge or understanding of the topic. • The student either: (1) writes about the topic using irrelevant or inaccurate information (2) recalls the steps of the Group Activity in Part II of the performance assessment, adding no new or relevant information and showing no understanding of how the activity relates to the general topic.
0	• The student either: (1) left the answer blank (2) wrote about a different topic (3) wrote “I don’t know”

Etc...Etc...Etc...

Remember...

Purpose of Testing

- To provide a record for assigning grades
- To provide a learning experience for students
- To motivate students to learn
- To communicate to students their level of understanding of the course objectives and serve as a guide for further study
- When utilizing pretests, feedback is provided regarding the knowledge students bring to the content
- To assess how well students are achieving the stated goals and course objectives
- To provide the instructor with an opportunity to reinforce the stated objectives and highlight what is important for students to remember

Tips on Test Construction

1. Assess information indicative of the material stressed in class, not trivial information
2. Have students submit 1 or 2 test questions and give extra credit for appropriate questions. Have them write a question with a correct answer and source
3. To determine how much time the student will need to take the test use the following:
   - 30 seconds per true-false item
   - 60 seconds per multiple choice item
   - 120 seconds per short answer item
   - 10-15 minutes per essay question
   - 5 to 10 minutes to review the work
   - Or, allow triple the amount of time it takes you to complete the exam
4. Select items that at least 50 to 70% of the students can correctly answer, or are of average difficulty
5. In terms of test reliability, longer tests are considered more reliable than shorter tests
6. Be aware that many of the test banks and/or reviews in textbooks rarely assess higher levels of learning

Test Layout Tips

1. Include simple, succinct directions to include the following:
   - How to record answers if they are not to write on the exam
   - Whether or not to show work on problems
   - The point value for different items
   - Directions on how to use an answer sheet if provided
2. Avoid splitting a test item between two different pages
3. Leave the appropriate amount of space for each item
3. Leave wide enough margins for your comments, points, etc.
4. Group similar items together
5. If it is a large exam, it might be worthwhile to group items according to content as well
6. Leave space for the students name if they write on the exam.
7. Start with your easiest items in each section

Returning Tests and Giving Students Feedback Regarding Tests

1. Return exams promptly. If this is not possible, post a corrected copy immediately after the exam

2. When you return a test make sure the score is not showing (turn the test over; put the score on the last page)

3. Give feedback to the class as a whole regarding the following:
   - Items most missed
   - Mistakes most frequently made
   - What was done particularly well

4. When going over the test with the class, ask the students to refer to their class notes (For example, "Look back at your notes on 12-5-01. What do you have regarding this topic?")

5. Do not respond to specific questions regarding the details of an individual students answer.

6. Consider having students prepare their case in writing if they want you to give them credit for a question

7. Consider full or partial credit for valid arguments.

8. Ask students to come up with specific questions versus "Why is my test score so low?"

9. When giving one-on-one feedback, do not overwhelm a student whose performance was overall poor with so much information that they do not know where to begin.

10. Let students know when they have improved, even if it did not result in extremely high marks.

11. Instead of explaining to a student why they missed a question, ask them to "think out loud". In other words, have them answer the question and tell you out loud their thinking process.

12. To account for missed tests/quizzes you might want to drop the lowest quiz score or double the highest quiz score.

Have Students Complete an Exam Evaluation

Include some or all of the following:

- How well did the exam questions reflect the content and what was emphasized?
- What questions challenged you to think?
- Which questions seemed like trick questions and why?
- How difficult did you find the exam?
- How much time did you spend studying for the exam?
- Were you clear as to what the questions were asking? List the numbers of those questions you were unclear about.
- Are you satisfied with your answers to the questions?
- What grade would you assign to this test?

Alternative Testing Modes

1. Take-Home Tests

Take-home tests allow students to work at their own pace with access to books and materials. Take-home tests also permit longer and more involved questions, without sacrificing valuable class time for exams.

- Problem sets, short answers, and essays are the most appropriate kinds of take-home exams.
- The instructor should avoid designing a take-home exam that is too difficult or an exam that does not include limits on the number of words or time spent.
- The take-home test should have explicit instructions on what the students can and cannot do: for example, are they allowed to talk to other students about their answers?
- A variation of a take-home test is to give the topics in advance but ask the students to write their answers in class. Some instructors hand out ten or twelve questions the week before an exam and announce that three of those questions will appear on the exam.

2. Open-Book Tests

Open-book tests simulate the situations professionals face every day, when they use resources to solve problems, prepare reports, or write memos.

- Open-book tests tend to be inappropriate in introductory courses in which facts must be learned or skills thoroughly mastered if the student is to progress to more complicated concepts and techniques in advanced courses.
- On an open-book test, students who are lacking basic knowledge may waste much of their time consulting their references rather than writing.
- Open-book tests appear to reduce stress, but research shows that students do not necessarily perform significantly better on open-book tests.
- Open-book tests seem to reduce students' motivation to study. A compromise between open- and closed-book testing is to let students bring an index card or one page of notes to the exam or to distribute appropriate reference material such as equations or formulas as part of the test.

Creating Fair Tests and Testing Fairly*

Many students with and without identified disabilities need support when taking tests. The type and extent of adaptations for fair test administration will vary from student to student and, possibly, from subject to subject for the same student. In addition, as the student gains skills, fewer accommodations may be needed.

A number of possible test administration adaptations are listed below. Educators should choose the best combinations of strategies for student success based on individual needs.

- Provide oral and/or written time checks during the test and provide breaks during long tests
- Give oral interpretation of directions
- Confirm correct responses with a nod, thumbs up, or correct mark on the page
- Explain the meaning of key vocabulary words
- Provide additional examples of the expected answer
- Trigger associations: "Remember when we..."
- Use a student-generated reference sheet (i.e., a legitimate "cheat sheet")
- Review just prior to the test
- Display reference charts in the classroom
- Excuse a student from answering specified test questions or sections (i.e., omit the essay or the short answer)
- Require fewer answers (evens or odds only when appropriate)
- Remove the pressure to rush through a test by agreeing to base the grade on the number of correct answers out of the total number of questions answered
- Provide a word bank/outline
- Read the test orally
- Allow use of calculators, computers, dictionaries, electronic spell-checkers, and/or tape recorders
- Allow enough time for completion of tests in one sitting or break the testing into two days
- Give a re-test
- Avoid adding additional pressure during testing by stating negative consequences of a poor score
- Allow students to tape record answers to essay questions or to outline the answer

*Information from: **Including All Students: A General Educator's Guide to Teaching a Diverse Student Population** For a free copy contact: Kansas Curriculum Center, (785) 231-1010 x1534, zzroot@washburn.edu

A number of decisions must be made about testing accommodations...

- How much and what kind of help will be given?
- Who will give the help (e.g., general or special education teacher, para-educator, or volunteer)?
- Where will the student be tested (e.g., in the regular classroom, a resource or conference room, the library, or the cafeteria)?
- When will the test be given (e.g., time of day, in one sitting or broken into short time periods, during the regularly scheduled class, after school, during recess, with or without additional time, etc.)?
- What adaptations should be made depending on the student's disability, the subject, the type of test, and the student's increasing skill in reading, processing, and writing independently?
- Adaptations must be individualized and kept private between teachers and students. Adaptations should parallel the accommodations made during instruction. For instance, if a student commonly uses taped books, then tests should be presented orally. If a student uses a calculator for completing daily assignments, then the calculator should be allowed during tests.

I'd like to use essay tests, but...

Marilla Svinicki, University of Texas at Austin
The Professional & Organizational Development Network in Higher Education
(continued from page 39)

If we want the students to be able to deal with the complex nature of essay tests and other forms of spontaneous writing, there are some things we can do in our instruction that will prepare them more adequately.

1. Help them think differently about the material

Students are conditioned from an early age to think in terms of discrete facts and "correct" answers rather than looking for the relationships which are characteristic of essay answers. One of the first steps toward improved essay answers is to adopt a different perspective on the nature of what is to be learned from the material presented and read. To help students think about the material differently, the instructor can:

- Encourage students to integrate material from class to class and unit to unit. For example, have the students answer some of the questions listed below each time they begin a new topic:

  —How does this topic compare with/relate to what has gone before?
  —How is it different? How is it similar?
  —Why is it included in the course? Why at this point?
  —What are its main points, its strengths, its weaknesses?
  —How does it apply to the overall goal of the course?

- Have them write their own sample essay questions for each lecture or reading assignment and then in class. Discuss those that most closely parallel what **you** would ask.

- Explain the levels of cognitive complexity (such as Bloom's taxonomy) which might be expected of them in the course and differentiate between knowledge of facts and ability to analyze and critique material.

- Emphasize process during classtime itself, so that the students begin to understand how conclusions are reached rather than focusing on the conclusions alone.

2. Help them study the material differently

Studying for essay exams is much different from studying for objective exams. Instructors should encourage students to:

- Create outlines of readings and lecture notes which emphasize the relationships among the ideas. Paraphrase or create an executive summary for each reading or lecture.

- Draw concept maps, which are visual diagrams of how terms, principles, and ideas interconnect.

(Continued on next page…)

3. Help them write structurally sound answers

To help students compile the information they have learned into answers which are written more effectively and efficiently, an instructor can:

- Provide a list of key words used in essay questions and what they imply in terms of answer content and structure. *(See page 41 of this document.)*
- Give students opportunities to practice writing essay answers in class and discussing the structure of the answers.
- Assign brief out-of-class essay questions with which to practice and provide individual feedback on the writing. You may wish to develop a feedback phrase sheet, which lists your most commonly used comments and an extended description of what that comment means.
- Give the students an opportunity to grade an essay answer using the system (rubric) you normally use so that they will understand how they are being evaluated.
- Provide examples of good and poor answers to essay questions with an explanation of why they are evaluated that way.

4. Help them learn time management techniques

Here are some examples of efficient time management techniques that the student could possibly benefit from in completing an essay exam:

- Scanning all the items and parceling out an appropriate amount of time to spend on each according to weight or importance
- Spending a few minutes outlining an answer before writing (the teacher could possibly give some credit for content which appears on an outline, but was not included in the answer due to time constraints)
- Having a checklist for quickly evaluating answers before completing the exam (such as “did you answer the question?” “are the transitions clear?” “is evidence provided for each assertion?” and so on).

5. Why should we bother?

There is actually an additional selfish motive for improving students’ essay writing skills: it makes the grading process much easier. If students learn how to read and interpret the structure of an essay question, they can create an answer that is comprehensive and well-organized. The task of grading those essay answers becomes less one of interpretation and more one of evaluation.

Test Administration Assignment

Read Through the Following Description of a Teacher Giving a Test

1. Students enter the classroom. Before the students have a chance to put away their things, the teacher announces that they will be having a test. No notice has been given of the test. In response to student complaints about the test, the teacher responds "it will show who is really paying attention in class and keeping up with the reading."

2. Before the students have time to remove their books, notebooks, etc. from their desks, the teacher starts handing out the quiz. (Some books, etc. remain on the desks.) Once the first students receive the test, they start busily taking it, while the teacher is handing out the rest. The teacher announces that the students will have twenty-five minutes to take the fifty-item test.

3. One minute into the test, one of the students raises her hand and asks whether to mark the correct response to the multiple choice item by circling or placing a check beside the correct response. Later, someone asks how to respond to the true-false items.

4. The desks in the room are close together. During the test, Billy glances over to Juan's paper and sees what his answers are. Billy sees about five answers.

5. After about twenty minutes, two students have completed the test and start rustling papers and whispering. Some students complain about the noise.

6. At the end of the time, the teacher announces that time is up. Some students complain that they did not know how much time was left and that they are in the middle of answering an item. The teacher collects the papers anyway.

7. While scoring the test, the teacher notices that some students did not have the last page. For those students, the teacher decides to score the test using only 40 items instead of 50.

List problems associated with the way the teacher administered the test.	*Indicate ways the teacher could have avoided the problems.*

Cognitive Domain Guide

Use this chart when the major topic or task primarily involves the acquisition and processing of knowledge.

Level	If the student must...	Then use these key words in objectives, assignments and evaluations		
Knowing	...recall or recognize this knowledge; giving it back in nearly the same form as it was received.	define describe identify label	list match name recall	recite recognize select state
Comprehending	...demonstrate an understanding of this knowledge, seeing relationships and telling in their own words what it means.	associate compare discriminate	distinguish explain match	paraphrase restate illustrate translate
Applying	...use this knowledge in a concrete situation other than in which it was learned.	apply compute convert demonstrate dramatize employ estimate extend	generalize illustrate infer interpret modify operate perform practice	predict produce relate rewrite show how solve transform use
Analyzing	...analyze or break down this knowledge into its essential parts, and differentiate between facts, opinions, assumptions, hypotheses and conclusions.	analyze break down categorize compare contrast detect determine	diagram differentiate distinguish find isolate outline point out	separate separate out show how subdivide trace the logic
Synthesizing	...produce something unique and original from this knowledge by synthesizing or combining the elements from an analysis into a new structure or organization.	arrange a new assemble combine construct create design develop from	devise formulate generalize harmonize integrate organize parallel	plan problem solve reconcile relate summarize
Evaluating	...form judgments about the value or worth of this knowledge.	access appraise conclude critique defend	evaluate judge justify rank by refute	select based on support test validate

Affective Domain Guide

Use this chart when the major topic or task is primarily concerned with acquiring new attitudes, values or beliefs.

Level	If the student must...	Then use these key words in objectives, assignments and evaluations	
Receiving	...receive information about or give attention to this new attitude, value or belief.	be alert to be aware of be sensitive to experience listen to look at	perceive existence receive information on take notes on take notice of willingly attend
Responding	...participate in, or react to this new attitude, value or belief in a positive manner.	allow other to answer questions on contribute to cooperate with dialog on	discuss openly enjoy doing participate in reply to respect those who
Valuing	...show some definite involvement in or commitment to this new attitude, value or belief.	accept as right accept as true affirm belief/trust in associate himself with assume as true consider valuable	decide based on indicate agreement influence others justify based on seek out more detail
Organizing	...integrate this new attitude, value or belief, with the existing organization of attitudes, values and beliefs, so that it has a position of priority and advocacy.	advocate integrate into life judge based on place in value system	prioritize based on persuade others systematize
Characterizing	...fully internalize this new attitude, value or belief so that it consistently characterizes thought and action.	act based on consistently carry out consistently practice fully internalize	know by others as characterized by sacrifice for view life based on

Bibliography

"The Advantages of Rubrics," The Learning Network
http://www.teachervision.com/lesson-plans/lesson-4522.html

"Assessing Student Performance," Ohio State Office of Faculty and TA Development, Ohio State University, Columbus, OH
http://www.osu.edu/education/ftad/Publications/TeachingHandbook/chap-6.html

"Authentic Assessment Overview," Pearson Education Development Group
http://www.teachervision.com/lesson-plans/lesson-4911.html

Border, Barbara. "The Status of Alternative Assessments Through the 1990s: Performance and Authentic Assessments in Relation to Vocational-Technical Education, Technical Skills, Workplace Skills, and Related Academic Skills" (1998) V-TECS

"Common Characteristics of Authentic Assessment," Aurbach & Associates
http://www.aurbach.com/alt_assess.html

"Creating Better Student Assessments," The U.S. Department of Education, Improving America's Schools http://www.ed.gov/pubs/IASA/newsletters/assess/pt1.html

Davis, Barbara G. "Quizzes, Tests, and Exams," **Tools for Teachers,** Jossey-Bass (1999)
http://uga.berkeley.edu/sled/bdg/quizzes.html

"Designing Test Questions," Walker Teaching Resource Center, The University of Tennessee at Chattanooga http://www.utc.edu/Teaching-Resource-Center/test-questions.html

Dewey, Russell A. "Writing Multiple Choice Items Which Require Comprehension," Georgia Southern University, Statesboro, GA
http://www.psywww.com/selfquiz/aboutq.htm

"Essay Tests," Office of Teaching Effectiveness, University of Colorado at Denver
http://thunderl.cudenver.edu/OTE/nn/

"General Tips For Writing Essay Prompts," Online Writing Lab, Rogue Community College, Grant's Pass, OR http://www2.rogue.cc.or.us/OWL/teaching/prompt.htm

"Guidelines for Rubric Development," EdWeb, San Diego State University
http://edweb.sdsu.edu/triton/july/rubrics/Rubric_Guidelines.html

Hitchcock, David. "How Can I Use Multiple Choice Tests for More Than Just Rote Recall of Facts?" http://www.mcmaster.ca/learning/teaching_tips/tip14.htm

"How to Write Tests: Matching Questions," University of Lethbridge, Alberta, Canada
http://www.edu.uleth.ca/courses/ed3604/HTWT.html#mat

(Continued on next page...)

"Improving Your Test Questions," Division of Measurement and Evaluation, Office of Instructional Resources, University of Illinois at Urbana-Champaign http://www.oir.uiuc.edu/dme/Exams/ITQ.html#sec23

"Including All Students" Kansas Curriculum Center, Washburn University

Jacobs, Lucy C. "How to Write Better Tests," Indiana University, Bureau of Education Studies & Testing http://www.indiana.edu/~best/write_better_tests.shtml

Johnson, John A. "Levels of Understanding Assessed by Multiple Choice Questions" http://www.personal.psu.edu/faculty/j/5/j5j/psy002/MCQ.html

Kehoe, Jerard. "How to Write Multiple Choice Questions," QuizPlease http://www.robmccormack.com/qphelp/qphowto5.htm

"Lack of Training on Standards and Tests". **Education Week on the Web** http://www.edweek.org/sreports/qc01/articles/qc01chart.cfm?slug=17tools-c1.h20

"The Multiple Choice Exam," University of Victoria, British Columbia Learning Skills Program http://www.coun/uvic.ca/learn/program/hndouts/multicho.html

"Performance Assessment," Education Research Consumer Guide, Office of Educational Research and Improvement (OERI), US Department of Education http://www.ed.gov/offices/OERI/

"Performance Assessment Scoring Rubric," National Center for Research on Education, Standards, and Student Teaching (CRESST) http://www/cse.ucla/edu/CRESST/pages/Rubrics.htm

"Principles of Good Practice for Assessing Student Learning," University of Montana at Bozeman, Office of Institutional Research http://www.montana.edu/aircj/assess/GoodPractices.html

Rowe, Allan D. "Rubric Basics," http://www.rubrics.com: 19815/4DAction/w_start/1

Rudman, Herbert C. "Integrating Testing with Teaching," ERIC Clearinghouse on Assessment and Evaluation http://ericae.net/pare/getvn.asp?v=1&n=6

"Standards for Teacher Competence in Educational Assessment of Students," Buros Institute, University of Nebraska, Lincoln http://www.unl.edu/buros/article3.html

"Student Evaluation of Test Item Quality," University of Illinois at Urbana - Champaign http://www.oir.uiuc.edu/dme/ices_cat/NewICESfaq.html

"Summary Data on Teacher Effectiveness, Teacher Quality, and Teacher Qualifications," National Council for Accreditation of Teacher Education (NCATE), 2000 http://www.ncate.org/resources/factsheettq.htm

Svinicki, Marilla. "I'd Like to Use Essay Tests, But..." The Professional & Organizational Development Network in Higher Education http://www.ucet.ufl.edu/ProgramService/topic6-2.htm

(Continued on next page...)

"Test Construction," Teaching@Indiana, Indiana University
http://www.indiana.edu/~handbook/evalsp-t.htm

"Test Taking Strategies," Lawrence University, Appleton, WI
http://www.lawrence.edu/dept/student_acad/college_place/before_the_exam.html

"Test Taking Strategies," Marquette University, Milwaukee, WI
http://www.eop.mu.edu/study/

"Test Techniques," The Center for Teaching Excellence, Lansing Community College, Lansing, MI
http://www.lansing.cc.mi.us/~cte/workshops/testingtechniques/tt9_references.html

"True-False Item Construction," College of Education, Texas A&M University
http://empowermentzone.tamu.edu/true_false_item/page2.html

"Writing Matching Test Items," Alabama Professional Development Modules
http://web.utk.edu/~mccay/apdm/mod3b/mod3b_a.html

"Writing Test Items," Michigan State University
http://www.msu.edu/dept/soweb/writitem.html

"Writing True-False Questions and Evaluating Responses," Alabama Professional Development Modules http://web.utk.edu/~mccay/apdm/mod3at-f_b.html

"Writing True-False Items," Office of Measurement Services, University of Minnesota
http://www.ucs.umn.edu/oms/truefalse.htmlx

Writing Constructed Response Items and Scoring Guides

For some desired outcomes of mathematics and science education, constructed response items provide more valid measures of achievement than do multiple-choice items. The quality of constructed response items depends largely on the ability of scorers to assign scores consistently and reliably within and across countries. Thus, it is essential that each constructed response item and its scoring guide be developed together.

Please keep the guidelines for writing constructed response questions in mind. In particular, ask a clear question, and develop a scoring guide for the question at the same time as the question is developed.

Constructed response items usually require students to give a numerical result, provide a short explanation or description given in one or two phrases or sentences, create a list, complete a table, or provide a sketch. They are scored as either 1 or 2 points for fully-correct answers.

- 1-point constructed response items are scored as correct (1 score point) or incorrect (0 score points).
- 2-point constructed response items are scored as fully correct (2 score points), partially correct (1 score point), or incorrect (0 score points). For example, a response demonstrating thorough understanding of concepts and processes will receive full credit (2 points). These responses show a complete or deeper understanding than a response that will receive partial credit (1 point). (Developing scoring guides is explained in the next section.)

Constructed response items should be used when it is desirable that the student be required to think of an answer without the possible cues provided by an option in a multiple-choice item. If too few plausible distracters are available for a multiple-choice item, it may be better framed as a constructed response item.

Developing a constructed response item accurately targeted on the ability to be assessed, along with the accompanying scoring guide, is not a straightforward task. Care in writing constructed response items is especially important for two reasons. First, if the task is not well specified students may interpret the task in different ways and respond to different questions. Second, a constructed response item may carry more score points than a multiple-choice item.

Guidelines for Writing Constructed Response Items

1. Write questions in easily accessible language appropriate to the age and experience of the target population. Use simple vocabulary and sentence structure, and avoid using complicated names for the subjects in the item.
2. Make what is expected of students as clear as possible without compromising the intent of the item. Give an indication, where appropriate, of the extent, or level of detail, of the expected answer (e.g., "Give three reasons …" rather than "Give some reasons …" and "Draw a labeled diagram illustrating the water cycle" rather than "What is meant by the term 'water cycle'?"). Select real life problem settings that are likely to be "real" to students at the target grade levels, and that involve quantities that are realistic for the situations.
3. Avoid asking questions that could give rise to answers that cannot be scored strictly in terms of accuracy of mathematical or scientific understanding (e.g., "What are satellites used for?").

4. Students should be able to complete the task in the time allocated for each constructed response item, that is, a maximum of 3 minutes.
5. Write an appropriate answer to the question in terms of the language, knowledge, and skills that a student at the target grade could be expected to possess. This tests the clarity of the question and is also an essential first step in producing a scoring guide for the item. It is also helpful for those who are reviewing the question.
6. Produce a scoring guide (see below). This action usually results in amendments to the item to clarify its purpose and improve the quality of information that can be obtained from student responses.

Response Types for Constructed Response Items

It is important to select the best response mode for each constructed response item. For most constructed response items, students will write the results of calculations, descriptions, or explanations. It is also possible to ask students to draw graphs or diagrams.

Scoring Guides

To ensure reliability, constructed response items need scoring guides with well-defined categories for allocating score points. It also is important to collect information of value for educational improvement. Students' answers can provide insights into what they know and are able to do, including common misconceptions.

General Scoring Guidelines for TIMSS Constructed Response Items

General scoring guidelines used for 1-point and 2-point constructed response items are described in Exhibit 6.

Exhibit 6: TIMSS General Scoring Guidelines for Constructed Response Items

Scoring 1-Point Items
1 Point (Full Credit)
A 1-point response is correct. The response indicates that the student has completed the task correctly.
0 Points (No Credit)
A 0-point response is incorrect, irrelevant, or incoherent.

Exhibit 6: TIMSS General Scoring Guidelines for Constructed Response Items (Continued)

Scoring 2-Point Items
2 Points (Full Credit)
A 2-point response is complete and correct. The response demonstrates a thorough understanding of the concepts and/or procedures embodied in the task.
• Indicates that the student has completed all aspects of the task, showing correct application of concepts and/or procedures
• Contains clear, complete explanations, supporting work, or evidence when required
1 Point (Partial Credit)
A 1-point response is only partially correct. The response demonstrates only a partial understanding of the concepts and/or procedures embodied in the task.
• Addresses some elements of the task correctly, but may be incomplete
• May contain a correct answer, but an incomplete explanation when required
• May contain an incorrect answer with an explanation or supporting work indicating a correct understanding of the concepts
0 Points (No Credit)
A 0-point response is inaccurate or inadequate, irrelevant, or incoherent.

The TIMSS Two–Digit Diagnostic Scoring System

The TIMSS diagnostic scoring system uses two digits. For example, 10, 11, 70, or 79.

The **first digit** is the score indicating the degree of correctness of the response as described in the generalized scoring guidelines.

The **second digit** is used to classify the method used in solving a problem, or perhaps to track common errors or misconceptions. The information from the second digit addresses questions such as: Do approaches that lead to correct responses to the item vary across countries? Is there one approach that students have more success with than others? What are the common misconceptions that students have about the matter being tested? What common errors are made?

The First Digit

The first digit for **correct or partially correct responses** signifies the number of score points given to the response. Thus:

- The first digit for a correct response is 1 for 1-point items or 2 for 2-point items

However, for an **incorrect or blank** response, the first digit is not zero. Instead:

- The first digit for an incorrect response is 7
- The first digit for a blank response is 9

The Second Digit

The **second digit** for **correct or incorrect responses** provides diagnostic information. Thus:

- The second digits used for diagnostic purposes with either correct or incorrect responses can be 0 through 2 (codes 20–22, 10–12, and 70–72)
- An incorrect response not fitting a pre-defined incorrect code is given a 79 for "other incorrect." If no diagnostic categories are defined, all incorrect responses receive **code 79**.
- **Code 99** means a **completely** blank response.

It is unusual for an item to give rise to more than two commonly used correct methods, or more than one common error or misconception. Frequently no more than one or two categories are required. In other words, **the specific diagnostic codes should capture only the predominant correct and incorrect approaches/strategies used by students**. Scoring constructed response items is a significant cost factor for national centers, so care should be taken not to provide codes for response types that do not have apparent value for educational improvement.

Writing Scoring Guides

It is essential that item writers provide scoring guides that include:

- The student response or responses that will be awarded full credit (1 or 2 points)
- For 2-point items, the student response or responses that will be awarded partial credit (1 of 2 points)
- If applicable, and absolutely necessary, diagnostic codes
- Example student responses for each code used

The example constructed response items and their scoring guides presented in Exhibits 7 and 8 illustrate the two-digit scoring system used in TIMSS and include the required information described above.

Exhibit 7: Example Constructed Response Item and Scoring Guide – Grade 4 Mathematics

$43 \div 5 =$

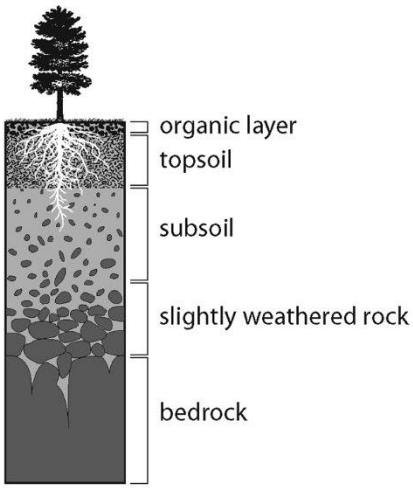
Answer: _____

M061272

Code	Response	Item: M061272
	Correct Response	
10	8 with a remainder of 3 OR $8\frac{3}{5}$ or equivalent	
	Incorrect Response	
70	8 (does not include remainder)	
79	Other incorrect (including crossed out, erased, stray marks, illegible, or off task)	
	Nonresponse	
99	Blank	

Exhibit 8: Example Constructed Response Item and Scoring Guide – Grade 8 Science

The diagram shows layers in the soil.



Most plants have roots that grow in the topsoil, but some have roots that reach into the subsoil.

Write two advantages for a plant to have long roots that go down into the subsoil.

1.

2.

S052116

Exhibit 8: Example Constructed Response Item and Scoring Guide – Grade 8 Science (Continued)

Code	Response	Item: S052116
	Correct Response	
20	Identifies any **two of these acceptable** advantages. • Long roots can reach more nutrients (minerals). *Examples:* *The plant will be able to receive more nutrients that are hidden in the subsoil.* *More soil nutrients can be reached* • Long roots can reach water (when topsoil is dry or frozen). *Examples:* *They can reach deep underground water.* *Will get more water.* *So that in time of drought they can reach the subsoil water.* • Long roots can anchor the plant better. *Examples:* *They are not easily blown over by strong winds.* *Will make the hold stronger in the ground.* *It holds the stem of the tree firmly.*	
	Partially Correct Response	
10	Identifies **one** acceptable advantage only. The second advantage may be incorrect or missing.	
	Incorrect Response	
79	Incorrect (including crossed out, erased, stray marks, illegible, or off task)	
	Nonresponse	
99	Blank	