

ASSESSING STUDENT OUTCOMES

Compiled and Presented

By

Dr. Steve Atkins

Introduction

- Grades assigned to student should be an accurate reflection of the amount they learned in the course.
- The assignment of grades often reflects complex, subjective and irrelevant variables such as neatness, attendance, gender, tardiness, etc.
- Common testing and grading practices, which appeal to our common sense, often produce inaccurate information which may not give a true representation of student learning

- The highest grades should go to the students who have achieved the greatest number of instructional objectives in the course, whereas, the lowest grade should go to the students who have achieved the least number of objectives.
- Often, there is a low correlation between the amount of learning a student achieves and the grade assigned
- Grades are often assigned according to an identical ritual, year after year, with little or no attempt to stop and examine the basis for the system being used.
- Lectures are routinely examined and updated but grading systems rarely, if ever, are

Definition of “Grade”

- Grade: A short-hand code for communicating with others your professional judgment as to the competence, skill, or knowledge a student possesses.
- Grading Practice: The set of rules and procedures followed in converting measures of students performance to a grade

Purpose of Grades

- To provide a formal systematic process for transmitting value assessments made by teachers to students and others directly concerned with their welfare and development.
- These value judgments provide a basis for making important decisions which are faced in the normal course of an individual's development in society

Why grades are important

- Serve to motivate students
- Provide evaluative feedback to students and instructors
- Provide useful information to other parties such as parents, colleges, graduate schools and prospective employers
- Help students plan a career path

Summative and Formative Evaluation

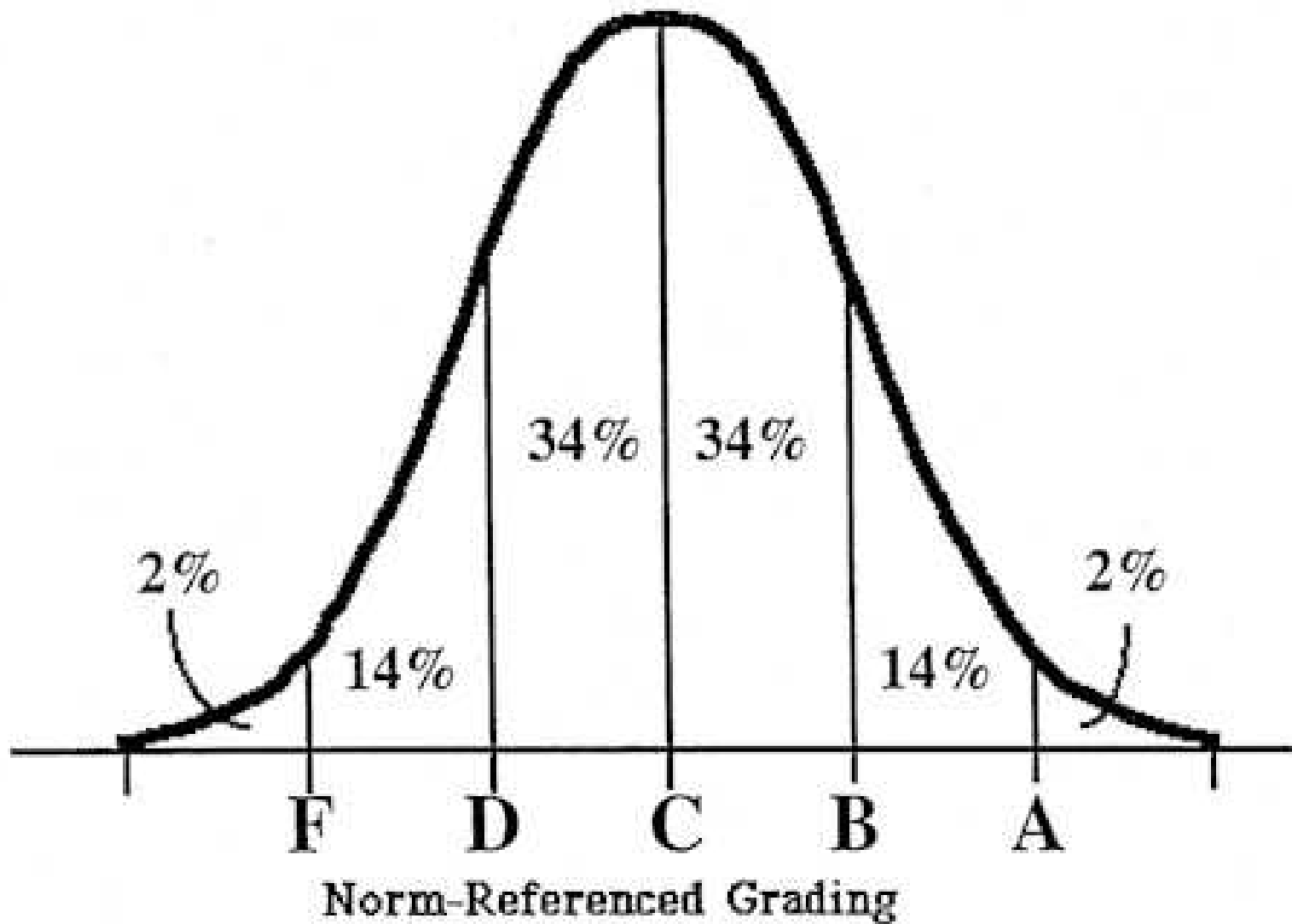
- **Summative evaluation** refers to judgments made at the end of a course to determine if objectives have been achieved. The focus is on determining if the course or program was successful and effective.
- **Formative Evaluation** refers to judgements made during the course so that improvement may be made immediately. The focus is on modifying an ongoing course .

GRADING SYSTEMS

Norm-Referenced Grading

- Norm Referenced Grading: Assigning grades based on the student's relative position within the class. Technically, in norm-referenced grading it is not possible for everyone to get an "A" or an "F"
- Grading on the "curve" (norm referenced grading) is appropriate only when you wish to divide students into homogenous ability for placement purposes.

Grading on the “Curve”

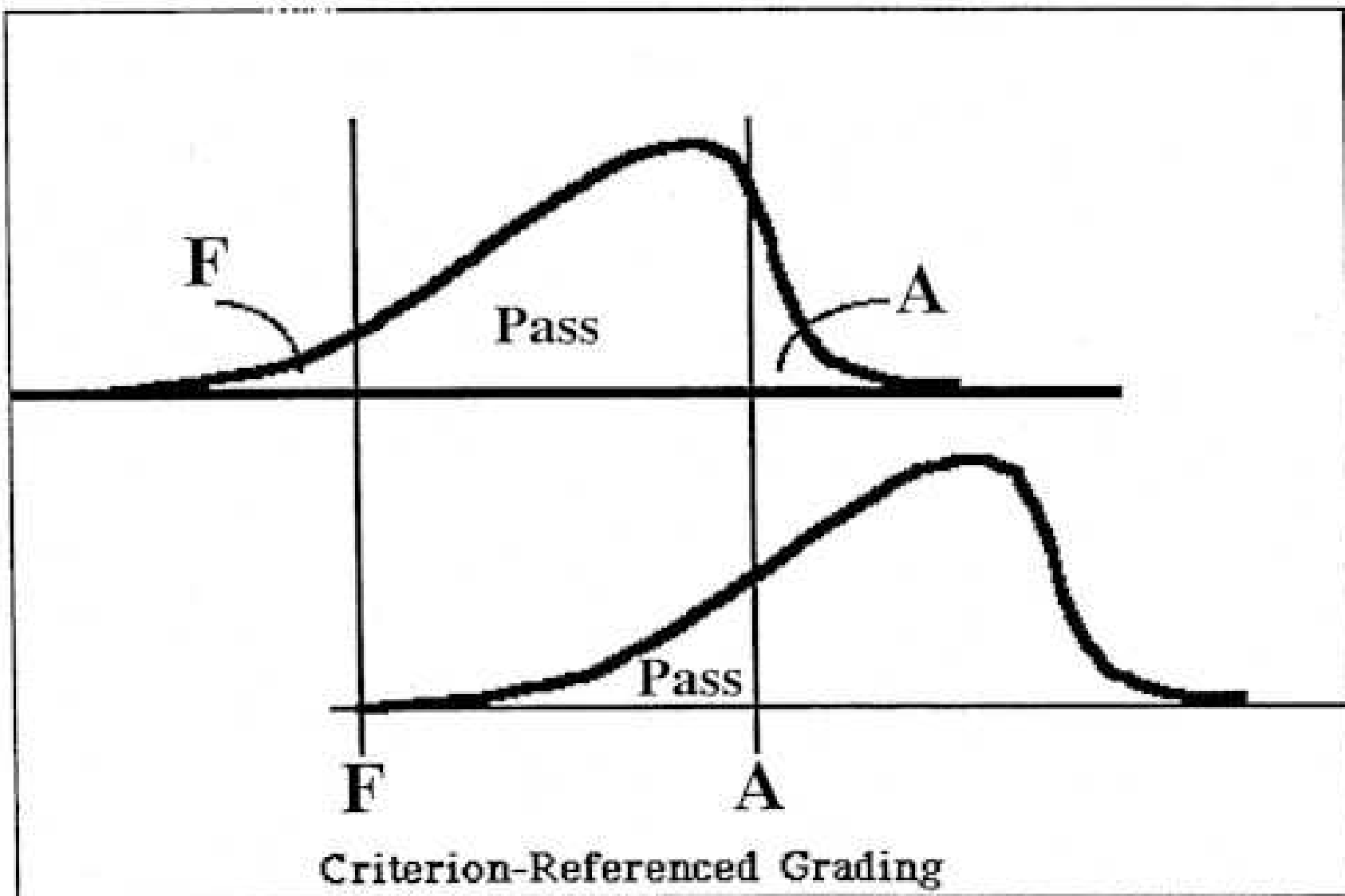


Criterion-Referenced Grading

- Criterion-Referenced Grading: Assigning grades based on a student's performance relative to some predetermined standard. Any student who meets the performance standard for an “A” receives an “A”. It is possible for all students to receive an “A” or an “F”.
- The correct procedure to use in most instructional situations is criterion-referenced grading.

Criterion-Referenced (cont.)

- Evaluated performance in comparison to a fixed standard.
- Identifies individuals who have mastered the institutional objectives.
- Geared to providing information to be used in planning instruction
- Is content-specific



Grading Systems

- The instructor should clearly and explicitly convey to students all of the factors constituting the assignment of their grade, in writing, at the beginning of the course and then adhere to them.
- A major problem with many grading systems is that the definitions are either vague, too short, or subject to a variety of interpretations. Vague grading system definitions can result in the problem of one faculty member's "A" being worth another faculty member's "C". This type of confusion may lead to grade inflation.

An appropriate system should clearly define the **performance standards** on which grades should be based, but include a statement as to the **instructors professional judgment** as to how the student may perform in the future.

Objective-Referenced Grading

- A = The student has demonstrated mastery of the course objectives, is well qualified for a subsequent course in this area and may be expected to perform well in such a course.
- B = The student has demonstrated a clear understanding of the course objectives, is prepared to take a subsequent course in this subject area and may be expected to perform adequate in such a course.

- C = Basic comprehension of course objectives, is qualified to take a subsequent course in the subject area but may be expected to perform only at a minimally adequate level in such a course.
- D = The student has demonstrated only an elementary familiarity with the course objectives, is technically qualified to take the next course, but may not be expected to perform well.
- F = Does not demonstrate knowledge or understanding of any meaningful portion of course objectives, not qualified...

Typical Grading System

- A = Consistently Outstanding
- B = Exceeds Expectations for Satisfactory Performance/Occasionally Outstanding
- C = Satisfactory
- D = Marginal
- F = Failure

TEST VALIDITY AND RELIABILITY

Test Validity

- Validity: The extent to which a test measures what it is supposed to measure. Or, are the test items, the lab situations, related to the objectives?

Steps for Improving Test Validity

- Teachers may be able to improve classroom tests' validity by using the list below as a guide to test construction
- **A. Directions and test format**
 1. Are directions clear and complete?
 2. Are students told how to record answers?
 3. Can students record answers without making errors?
 4. Are directions for specific items located near the items?
 5. Are students told how items will be scored? How many points? Is it permissible to guess?
 6. Has the student been given an example of how to respond?
 7. Is test printing large and bold enough to be easily read?
 8. Is there adequate space for students to write answers?

Improving Test Validity (cont.)

B. Test items

9. Do test items measure the instructional objectives and ideas to be learned in the unit?
10. Do test items and instructional objectives correspond to what was taught in the unit?
11. Is there an adequate number of items to reduce the effect of guessing?
12. Are items inappropriately difficult or easy?
13. Is the vocabulary and sentence structure within the ability of students?
14. Do test items violate any of the important rules for construction of items?
15. Are there distractions such as noise, movement, or student activity? Is there adequate time allowed?
16. Are students giving or receiving answers from each other inappropriately?

Test Reliability

- **Reliability**: How consistently the test measures what it is supposed to measure. Or, how well can we depend on the score as a representation of the students “real” or “true” achievement?
- If a test gives the same results when measuring an individual or group on two different occasions then the scores are reliable. If different teachers rate the same essay, for example, on the same criteria and obtain the same score then we say the scores are reliable from one rater to another. In both cases we are interested in consistency or trustworthiness.
- Objective tests usually have coefficients in the range of .80-.95.

Factors Influencing Reliability

- **Objectivity**: Scores for objective test items are less subject to the opinions or values of the scorers and are thereby more reliable. In essay testing, when relying on observations of students' performance, or when rating the products of their work, scores tend to be unreliable. Later you will learn of some ways to increase reliability in such situations.
- **Difficulty of the test**: Tests which are too easy or too difficult tend to be less reliable.
- **Length of test**: Generally speaking, the longer the test the more reliable. More specifically, the more items testing an idea or a skill the more reliable the test score will be. This is because 1) more items reduce the chance that guessing will effect the score 2) greater sampling of content.

Factors Influencing Reliability (cont.)

- **Reliability of Judgements**: If judges adopt common criteria, then the inter-rater reliability can be good.
- What is needed is a set of criteria, or standards, such that any teacher rating a student would be looking for the same thing.
- Everyone involved should know the criteria.

ITEM ANALYSIS

Item Analysis

- Each test should have an item analysis conducted every time it is administered. The analysis should include (1) Item Difficulty and (2) Item Discrimination.

Item Difficulty


- The difficulty of an item is computed as the percentage of people who got the item right.
- A very easy item may have a difficulty index of .87, and a very difficult item may have a difficulty index of .13.
- In constructing your exam you may wish to have the bulk of the items with difficulty indices in the range of about .4 to .6.

Item Discrimination

- The discriminating ability of an item is the item's ability to have the people who know the most get it right and the people who know the least get it wrong.
- The value for the discrimination index can range from -1.00 to $+1.00$
- -1.00 means the low scores on the test got the item right while the high scores got it wrong
- 0.00 means equal number of low scorers and high scorers got it right
- $+1.0$ means high scorers got it right, low scores got it wrong.

Item Discrimination (cont.)

- As a “rule of thumb” any item with a discrimination value of $+0.19$ or less is not considered a good item.
- Many teacher made tests include items which do not discriminate well and which possess unknown difficulty levels. This results in a test which may not be reliable and which may produce scores that are not accurate reflections of the student's learning.



LEARNING AND PERFORMANCE OBJECTIVES

Specifying Learning Objectives for Assessment

- Learning is a psychological process. Thus, the assessment of learning requires assessment of various psychological processes.
- Bloom has identified six levels of increasingly higher levels of cognitive learning that can serve as a useful tool for test construction

Bloom's Taxonomy

- **Basic Knowledge:** - To recall and memorize - Assess by direct questions.
- **Comprehension:** - To translate from one form to another - Assess by having student's (1) restate material in their own words, (2) reorder or extrapolate ideas, predict or estimate.
- **Application:** To apply or use information in a new situation - Assess by presenting students with a unique situation (not identical to that used during instruction) and have them apply their knowledge to solve the problem.
- **Analysis:** To examine a concept and break it down into its parts - Assess by presenting students with a unique situation of the same type but not identical to that used during instruction, and have them analyze the situation and describe the appropriate procedure or solution to the problem.

Bloom's Taxonomy (cont.)

- **Synthesis**: To put information together in a unique or novel way to solve a problem - assess by presenting students with a unique situation NOT of the same type used during instruction and have them solve a problem by selecting and using appropriate information.
- **Evaluation** - To make quantitative or qualitative judgements using standards of appraisal - Assess by presenting the students with a situation which includes both a problem and a solution to the problem and have them justify or critique the solution.

Performance Objectives

- A performance objective is a statement of what students will be able to do when they have completed instruction.
- A performance objective has three major components:
 1. A **description** of what the student will be able to do
 2. The **conditions** under which the students will perform the task.
 3. The **criteria** for evaluating student performance.

Why are performance objectives important?

- 1. Selection of content
- 2. Development of an instructional strategy
- 3. Development and selection of instructional materials.
- 4. Construction of tests and other instruments for assessing and then evaluating student learning outcomes

Three Essential Characteristics

- **Behavior** - The choice of a verb is all important here. Such frequently used terms as know, understand, and appreciate do not meet this requirement.
- **Criterion** – An objective should make clear how well a learner must perform to be judged adequate. This can be done with a statement indicating a degree of accuracy, a quantity or proportion of correct responses – etc.
- **Conditions.** What tools, references, or other aids will be provided or denied should be make clear.

Sample Performance Objective

- Given a set of data the student will be able to compute the standard deviation.
- Condition – Given a set of data
- Behavior – the student will be able to compute the standard deviation
- Criterion- (implied) –the number computed will be correct.

TEST CONSTRUCTION

Constructing an examination (Learning Assessment Instrument)

- **1. Develop Table of Item Specifications**
- **2. Determine Test Length** – Depends on the time available and amount of instruction to be covered.
- **Thumb rules:**
 - 50 – 60 sec per multiple-choice or completion item
 - 30 sec per true/false
 - 10 min per ½ page essay
 - (Useful technique – give test to a knowledgeable colleague, note the time that he/she takes, and then multiply that time by 2.5 to determine amount of time for students to complete test)

- **3. Plan Item Arrangement**
- Items of same type or topic should be grouped together
- Multiple-choice or true-false items should be arranged in such a way that correct answers do not follow some set pattern
- Completion items should be grouped in fours or fives
- On matching items all of the options should appear on the same page
- **4. Write the Test Directions**
- **Each test should have a written set of directions at the beginning, even if direction are provided orally. Directions should include time limits, method of recording answers, and the way the test is to be scored**

Write the Test Items – Multiple Choice

- **Stem** – presents the basic question. It should supply all the information and qualifications for making the selection among the response alternatives. The stem should be written in such a way that it ends with a question mark. The student should not have to read all of the responses in order to understand the question.
- **Distracters** – Distracters must be plausible enough to attract the attention of the under prepared students, yet not provide them with unintentional clues to the keyed response. One method to use in writing distracters is to develop “wrong” answers to the question by systematically leaving out an essential piece of information in thinking through the solution.

General Guidelines for writing multiple-choice Items

- 1. The stem should clearly formulate a problem
- 2. Include as much of the item content as possible
- 3. Don't load the stem down with irrelevant material
- 4. Be sure that there is one and only one correct or clearly best answer
- 5. Beware of “Clang” associations
- 6. Beware of grammatical cues
- 7. Beware of the use of “none of these”, “none of the above”, “all of the above” (if used then should be an answer as often as not)
- 8. Do not lift a statement, verbatim, from the text or lecture notes

- 9. If an item is based on opinion or authority, indicate whose opinion or what authority
- 10. Watch out for interlocking or interdependent items (don't make it necessary for students to get one item correct prior to considering the next item)
- 11. Don't try to write "trick" or "catch" questions.
- 12. Avoid ambiguity in the stem

Guidelines for Completion and True/False

- 1. Be sure an item can be unequivocally classified as either true or false.
- 2. Beware of “specific determiners” (give cue to the probable answer – “all”, “never”)
- 3. Beware of ambiguous and indefinite terms of degree or amount (Expressions such as “frequently”, “greatly”)
- 4. Beware of negative statements and particularly of double negatives. (the negative is often overlooked in hurried reading of an item)
- 5. Beware of items that includes more than one idea in the statement
- 6. Beware of giving cues to the correct answer by the length of the item.

Matching Items

- To prevent guessing, use more items in one column than in the other

Completion

- Good completion items usually cannot be made merely by leaving out one or more words from a sentence or passage from the textbook, because such words often make sense only as a result of their context. Without the context the completion item becomes ambiguous.
- Be careful not to leave out too many words or overly mutilate a passage so that the item becomes nonsense, e.g. “The is the of the

Essay Exams

- The best evaluation for higher order concepts
- Advantage
 - High order
 - Creative ability
 - Improves writing ability

Defining Essay Exams

- An essay question may be defined as an item which requires an original thoughtful response composed by the examinee, in the form of several sentences

Which of the following satisfy the criteria for an essay item?

- 1. List the three ways of testing for a starch.
- (NO. This is a recall item)
- 2. What is meant by the statement “all physical and chemical changes are accompanied by changes in energy?”
- YES
- 3. Tell what you know about active transport.
- (NO. Allows the student to write about anything for full credit.
- 4. How do you feel about the position that the US Government seems to be taking on environmental issues?
- (NO. This asks for an opinion)
- 5. Why is the following assumption so important to our understanding of geologic history?
- (YES)

Appropriate Essay Items

- 1. Draw warranted conclusions from evidence.
- 2. Interpret a written selection, art work, etc..
- 3. Decide what information is needed to solve a problem
- 4. Express a reasoned judgment and defend it.
- 5. Compare and contrast methods, authors, etc.
- 6. Identify the assumptions implicit in...
- 7. Integrate knowledge from related areas
- 8. Evaluate the worth, merit, utility of...
- 8. Justify methods, procedures, points of view, etc.
- 10. Summarize the main ideas in...
- 11. Explain the implications of
- 12. Explain which principle or generalization are relevant
- 13. Identify common exceptions to a generalization and explain the reasons for these exceptions.

Guidelines for Writing Essay Questions

- 1. Delimit the scope of the content to be covered.
- 2. Define the students' task as clearly, specifically, and completely as possible.
- 3. Appropriate words or phrases: analyze, classify, evaluate, interpret, explain why, justify the use of, cite example of, give reasons to support, predict what would happen if are appropriate to help the student identify the task you want him or her to address.
- 4. Avoid using verbs "discuss", "comment on", "elaborate on", "write all you know"
- 3. Don't use essays to measure learning outcomes that can be better measured by objectively scored items

- 4. Use relatively short essay questions rather than one long one.
- 5. For each essay item, tell the student the point value possible and the approximate time limit students should observe in responding.
- 6. Before administering the question, always write a model answer or a scoring guide – an outline or list of important elements that should be included in an ideal answer.
- 7. Before administering any of the questions you have written, ask at least one colleague to review critically each question
- 8. After administering the test, review the range of answers you receive and the manner in which students appear to have interpreted your question.

Preparing students for an Essay Exam

- 1. From the first day of class, emphasize the concept of writing-to-learn.
- 2. Make available samples of your essay exam, with examples of good responses.
- 3. During lecture or a class discussion say something like, “Now this is a question you might find on an essay exam...”
- 4. Prior to exams, provide students with study questions, out of which will grow specific exam questions.

Grading Essay Exams

- 1. Develop a model answer or list component that the ideal answer should include, and how much credit is to be allowed for each point. The components should include major ideas, but may also include specific features.
- 2. Read all answers to one question before going on to the next.
- 3. Grade the papers as nearly anonymously as possible.