

Part I: Testing and Grading – abridged and edited

*Stanford C. Ericksen
University of Michigan*

Fair play is the first and final requirement in matters of testing and grading. Students will accept pressures for hard work but object strenuously and rightly so, to signs of unfairness in a teacher's assessment of their efforts.

An important distinction must be made between evaluation and grading. *Evaluation* is information provided to the student about particular aspects of what was said or done during the effort to learn, to solve a problem, or to organize and integrate facts and concepts. As they move into unknown intellectual territory, students must have guideposts to confirm that they are moving in the right direction. The qualitative comments about particular aspects of a term paper are far more constructive aids for the specifics of learning and remembering than is the grade on the cover page. Evaluation, therefore, is indispensable to students for gaining understanding and to fix what is learned in memory. A *grade*, on the other hand, is a gross index which typically comes too late for the student to take corrective measures about the specifics of learning.

Testing as a Tool for Instruction

Students tend to concentrate their study effort in preparation for an exam, and they structure this effort in anticipation of the nature of the questions they will be asked. If students anticipate the need to know unassimilated facts, they will concentrate on memorizing information; if they expect to be asked to integrate, extend, and evaluate information, they will try to prepare themselves along those lines. The management of testing is an opportunity for the instructor to underline the essential elements making up the course.

The Diagnostic Use of Tests

Placement testing is commonly used at the department and college level, but within our own courses we can also make effective use of similar testing for making a grade-free diagnostic appraisal of what information is already known by the students or is not known but should be. Diagnostic testing is an excellent instructional tool because when a student says, in effect, "I don't see why the question was scored that way," an inquiry is started toward unscrambling the false connections. In this close-up look, the teacher may note a pattern of mistakes showing a misunderstanding of a particular rule, procedure, or principle. It may also appear that a student has the right answer but for the wrong reasons.

When students realize the significance to themselves of grade-free probing, they are more likely to open up and reveal low points in their preparation profile, anxieties, misconceptions and deficiencies in knowing how to do certain tasks. A sprinkling of short, diagnostic quizzes early in the term suggests to students that the teacher cares about how they are doing and is taking corrective steps to help them along - an excellent climate for starting the semester.

Assessing Achievement

Although test scores in any setting are affected by students' aptitude, study skills, motivation, background preparation, and the influence of the teacher, our classroom examinations should be designed primarily to measure subject-matter achievement. To this end, the teacher and student seek the same wavelength within an assigned domain of knowledge. A frustrated student expressed a contrary state of affairs quite clearly, "I don't like to play the professor's game: I've got a secret, see if you can guess what it is."

Effective classroom instruction is central to student learning, but students are short-changed if the examinations are trivial, irrelevant, confusing and tangential to the substance of the course. College teaching is not complete without an accurate and fair assessment of students' achievement during the term and at its conclusion.

Objective Tests

Students are best served when their instructors develop exams tailored to their specific course and based on sound principles.

Two basic concepts need to guide the development of classroom examinations:

1. *Validity* refers to whether an instrument measures what it is supposed to measure. A valid test, therefore, samples what students should have learned from your course offering. A valid test is an unambiguous reflection of what is worth knowing and remembering.
2. *Reliability* refers to the consistency of an instrument's results. A good short quiz is better than a poorly constructed long test but, assuming equal quality of items, a 50-item test is more reliable (stable, consistent) than a 10-item quiz. The easiest way to reduce the unreliability in the measuring instrument is simply to increase the length of the test.

Objective tests come in many forms, but the multiple-choice format carries most of the burden. When carefully worded, multiple-choice items can probe a student's understanding of factual information, skills and procedures, concrete and abstract concepts, and the implications from different scales of values.

To strengthen the quality of the set of items used, a complete item analysis should be made of each new test. This test-of-the-test is mainly to determine and adjust the difficulty level of each item. It is normal to find that many of our carefully conceived questions turn out to be too easy or too difficult or just seem to ride along as excess baggage.

Because ambiguity of meaning is a persistent problem, the wording of test items is critical. Careful editing of the draft exam includes close attention to such pitfalls as cluing the right answer, overlapping correct alternatives, or asking for a positive answer to a negative question. Good test items are parsimonious in meaning and simple in wording

Subjective Evaluation

The distinctive value of essay exams or term papers is the freedom they offer for students to probe and develop the personal meaning of ideas and to express these thoughts in their own words. To organize an integrated chain of thought, to elaborate on findings, and to communicate ideas to others are stronger tests of achievement than is the recognition or recall of isolated units of information.

1. *Essay Exams.* Th[e] ability to recall [material for an essay exam] is a more demanding test of memory than simply to recognize something. As essay examination elicits the ability to retrieve information but with little help from presently given cues. The perceptive teacher (reader) can evaluate the strong and weak points in a written argument even when the student's perception of a question differs from the teacher's. Evaluative permissiveness can, of course, go only so far.

2. *Term Papers.* In some respects, the term paper is the essence of what a student has gained from the course. It sets forth what the individual student has learned and how the student has pulled together all the information for comprehension and understanding. This, in turn, serves to keep the knowledge available in long-term memory.

A written handout is a useful guide regarding the due date, length, use of references, comments about style, and any other restrictions or suggestions about the assignment. The heavy task of reading these papers is counterbalanced, somewhat, by the satisfaction of reading the better papers - some of which can be truly exciting.

stay tuned next week for part II "Grading"

Material on this topic and others can be found in the CTL library in the Mason Room at the Smith House.

