

WRITING AND GRADING ESSAY QUESTIONS

A hundred years ago, all college course exams were essay exams. The multiple-choice exam had not yet been invented, and students were expected to demonstrate their learning by producing detailed answers to question prompts in clear prose. Today many teachers still consider essay questions the preferred method of assessment. Arguments in favor of the essay exam include:

- Essays require both more effort and deeper understanding on the part of the student than do other types of questions. For example, students typically cannot produce an adequate answer to an essay question simply by recognizing or recalling the correct information, as they often can with multiple-choice or short-answer questions.
- Essay questions can better assess the complexity of students' thought processes, as well as their ability to think critically and solve problems in a particular domain of knowledge, than can multiple-choice or fill-in-the-blank items.
- Essays require students to express their thoughts in grammatical, well-crafted sentences and paragraphs—a skill that all educated persons should be expected to master.

Still, the essay form has its limitations. There are four major arguments against using essay questions to assess student learning:

1. **Essays take much longer than other types of assessment items for students to write and for teachers to grade.** Consequently, for a fixed amount of examination time or grading effort, fewer items can be administered, which weakens the validity of the exam. Specifically, an exam that does not adequately sample the entire domain of knowledge being assessed (i.e., what the student is supposed to have learned) may not yield an accurate indication of how well the student has mastered the course content. One way to achieve adequate domain coverage with essay exams is to administer exams that require many hours or days to complete (e.g., components of the Bar examination that assesses legal knowledge), but this is not a realistic option for most college instructors.
2. **Grading of essay exams can be influenced by extraneous factors.** These factors include such things as handwriting legibility and ink color (Klein & Taub, 2005; Joseph, 2005; Greifeneder et al., 2010). In addition, if the exams are not graded anonymously, teacher expectations regarding the quality of a student's work, stereotypes associated with the student's sex or ethnicity, or the grader's personal feelings toward the student may influence essay scoring (Chase, 1986; Hughes, Keeling, & Tuck, 1983). Contrast or order effects may also play a role; essays preceded in the grading queue by poor quality papers tend to receive higher scores than do the same essays when preceded by high quality papers (Spear, 1997). Because these factors have no systematic relationship to the quality of the ideas expressed in a student's essay, their influence weakens the validity of the assigned score in that the score assigned to the essay does not accurately indicate the degree of subject mastery the student attained.
3. **Essay grades are unreliable.** Due in part to the influence of extraneous factors, there tends to be relatively poor agreement among graders in the score assigned to an essay. Moreover, even individual graders are often inconsistent in the scores they assign to an essay on two different readings. Specifically, in empirical studies of reliability in essay scoring, both consensus (i.e., exact numerical agreement between scores) and correlation (i.e., degree to which the essays in a set are ranked similarly by different graders) have been found to be low (Brown, 2010). The fact that different graders (or even the same grader in the case of multiple readings) often cannot agree on what score a given essay should receive undermines confidence that a student's score on an essay exam accurately reflects the student's mastery of the material being tested.

4. **Essay exam performance conflates course subject matter knowledge with writing skill.** Some authors argue that essay exams are a poor method of assessing subject matter knowledge because the student's performance—and, consequently, her exam score—depend not only on her knowledge of the subject matter being tested but also on her written communication skills.

Given these limitations, if there existed a satisfactory alternative way to assess the highest levels of understanding and reasoning in a domain of knowledge, we might well dispense with the essay exam altogether. At present, however, "objective" forms of assessment such as multiple-choice tests are at best a complement to essay exams and not an adequate substitute for them. Your task as an instructor, then, is to devise essay question prompts and grading procedures that will minimize the threats to validity and reliability described above. Below we offer some suggestions to help you avoid some of the more common pitfalls instructors encounter in creating and grading essay questions.

1. **Determine whether an essay question is the most appropriate format for the type of learning you want to assess.** Essay questions should be used when you want to assess students' ability to think critically and organize their thoughts, or to demonstrate their understanding by taking the factual information they have learned and applying it in some way. For example, an essay question might ask a student to critique an argument, interpret a text, justify a position on some issue, explain the causes of some phenomenon, or predict the effect of an intervention (see Reiner et al., 2003, for an extensive list of the different kinds of tasks you might ask students to perform on an essay exam).
2. **Administer enough different question items to adequately sample the domain of knowledge covered by the exam.** In order for a student's exam score to serve as a valid indicator of how well she has mastered the material covered by the exam, there must be a sufficient number of question items, addressing a sufficiently diverse set of topics, to adequately represent the full range of subject matter the students are expected to have learned. Single- item essay tests rarely meet this criterion unless they are broken down into a number of sub-components, in effect becoming a set of short essays. In many cases it is preferable to use a number of short essay questions to insure that the material has been sampled adequately.
3. **Avoid ambiguous prompts; state the question clearly and precisely and make clear what information the answer should contain.** Do not assume that your students will interpret a vaguely-worded question in the way you intend; it is better to err on the side of providing too much detailed guidance in the prompt than too little. In addition to specifying the information that an essay answer should contain, the prompt can help students allocate their time and effort appropriately by indicating how much time they should spend on each part of the answer and/or how many points each part of the answer is worth, as does the essay prompt from a Physical Anthropology exam shown in Figure 1. Note that the three parts of this prompt ask students to demonstrate progressively more complex and sophisticated forms of understanding and mastery. Part I requires primarily recall and comprehension, Part II requires application and analysis, and Part III requires synthesis and evaluation.

SOURCE: <http://cfe.unc.edu/publications/fyc7.html>