Writing Better Essay Exams

IDEA Paper #76 • March 2019

Allison P. Boye • *Texas Tech University*

Abstract

Essay exams offer many benefits for instructors who seek to vary their assessment methods and engage students in critical discourse, yet they also pose many challenges and require thoughtful construction and evaluation. The author provides an extensive overview of the literature to illuminate best practices for designing and assessing effective essay prompts, includes examples, and offers suggestions for preparing students for success.

Keywords: Essays, constructed response, assessment, examinations

Selecting and designing exams can be one of the most difficult tasks that instructors face. They must not only identify the best measurement of their students' achievement but consider what method is feasible, given course logistics such as class size and time constraints. Instructors who elect to give exams can choose from selected-response formats, such as multiple-choice items, or constructed-response formats, such as essay and short-answer items. (See IDEA Paper 70 for more information on selected-response formats [Haladyna 2018].) Both formats present many benefits as well as drawbacks, and both require careful development to ensure their effectiveness. In this paper, I focus on constructed-response formats—more specifically, on best practices for extended-response essay-test item design, implementation, and evaluation.

A Few Definitions

Before examining the creation and implementation of essay exams, it is worthwhile to clarify some important terms. There are two broad types of "essay" exam items (Clay, 2001; Nilson, 2017). *Restricted-response*, or short-answer, questions likely have expected "correct" responses (e.g., "List the major components of Freytag's triangle in dramatic structure."). *Extended-response* questions (the focus of this paper) are those that typically come to mind when envisioning a traditional essay exam—questions or tasks that could have multiple correct responses or lengthier ones (or both); are often more complex than restricted-response items; and, as Nilson notes, "require professional judgment to assess" (p. 299). (E.g., "Analyze the dramatic structure of Ibsen's *A Doll's House,* providing examples from the text to illustrate and support your analysis.")

But what exactly constitutes an extended-response essay question or test item? Of course, the traditional essay question is an open-ended prompt that requires an in-depth written narrative in response, such as the Ibsen example. However, other non-narrative formats could also fit within these parameters. For instance, an extended open-ended math or engineering problem with a variety of potential valid approaches would certainly qualify, with the traditional writing process supplanted by a demonstration of logical progression and application of principles. Although such an item might have a single "correct" response, the multiplicity of approaches and its call to demonstrate an expansive and thorough higher level understanding and application of course material places it firmly within the realm of the types of test items being considered here.

The Pros of Extended-Response Essay Exams

There is perhaps no perfect assessment tool, and essay exams are no exception, but they do offer much to instructors who might feel that multiple-choice, true-false, or other similar formats don't quite meet all their needs.

Higher level thinking. The first advantage that most educators likely associate with essay exams is their potential for eliciting higher level cognitive skills (Clay, 2001; Center for Research on Learning and Teaching [CRLT], 2016; Halpern & Hakel, 2003; Jacobs & Chase, 1992; Nilson, 2017; Parmenter, 2009; Reiner, Bothell, Sudweeks, & Wood, 2002; Scouller, 1998; Walstad, 2006). For instance, essay-test items can allow an instructor to assess students' reasoning, critical thinking, creativity, or ability to synthesize material or compose an argument (Bean, 1996; Nilson; Ory & Ryan, 1993; Reiner et al.; Walstad). Other researchers agree that essay tests can reward deeper knowledge of course material and assess more complex learning outcomes (Jacobs & Chase; Minbashian, Huon, & Bird, 2004; Parmenter; Scouller).

Interestingly, even students can perceive essay exams to be more appropriate "for the purpose of reflecting one's knowledge in the subject matter" (Zeidner, 1987, p. 357; see also Parmenter, 2009). It is important to understand, however, that while essay prompts have great capacity for assessing higher order thinking skills, they do not inherently or automatically do so. Indeed, essay prompts can certainly be designed to assess nothing but simple recall (Reiner et al., 2002), and it is not unusual for instructors to reward essay responses that demonstrate quantity over quality, or engage in "all about" writing that simply rattles off a laundry list of everything the student can remember about the topic rather than presenting a focused argument (Bean, 1996; Minbashian et al., 2004; Walvoord & Anderson, 1998). Nevertheless, essay items offer instructors an opportunity to engage students in high-level thinking through careful design and evaluation, as I will demonstrate later.

Authentic assessment. Similarly, many see essay exams as a more authentic form of assessment than selected-response tests (Jacobs & Chase, 1992; Lukhele, Thissen, & Wainer, 1993; Nilson, 2017; Reiner et al., 2002; Wiggins, 2011). That is, by posing more complex questions or tasks and requiring responses that students must construct themselves rather than simply recognize the correct response in a predetermined selection (Walstad, 2006), essay exams can more closely emulate tasks that students might be asked to do in the "real world" and help instructors identify student misconceptions more accurately. As such, essay exams can be less prone—though not immune—to student guessing behavior (Clay, 2001; Jacobs & Chase; Parmenter, 2009). Bean (1996) further points out that for those who locate knowledge and mastery "in the ability to join a discourse" rather than in the ability to recall selected information, essay exams are often preferable to objective tests (p. 185).

Avoidance of misinformation. Essay exams can also avoid the perpetuation of misinformation that can arise from multiple-choice tests (Parmenter, 2009; Roediger & Marsh, 2005). Roediger and Marsh found that students taking multiple-choice exams tended to remember an exam's "distracter" answers, or the wrong answers presented as if they might be correct, and thus could actually leave an exam having absorbed false information.

Communication skills. Some instructors also appreciate that essay exams in particular help them emphasize communication as a fundamental skill, regardless of discipline (Jacobs & Chase, 1992). Research has identified writing as a high-impact teaching practice linked to learning, and it is a skill often sought by employers (Walvoord, 2014). Essay exams can certainly aid instructors in gauging students' thought processes, organization ability, and logic (Nilson, 2017; Ory & Ryan, 1993; Walstad & Becker, 1994; Weimer, 2015) and give students the opportunity to "think and compose rapidly," which, as Bean (1996, p. 183) highlights, can also be useful workplace preparation.

Deep-learning study strategies. It is likewise interesting to note that students might actually study differently for essay exams than they do for objective tests, engaging in more "deep learning" methods (CRLT, 2016; Nilson, 2017; Parmenter, 2009; Roediger & Marsh, 2005). Research has demonstrated that students frequently perceive that multiple-choice exams require lower order thinking (not necessarily the case, of course) and thus prepare for those selected-response exams with surface-learning methods such as last-minute cramming, whereas they perceive that essay exams require more higher order thinking and prepare for them less superficially and more thoroughly (Entwistle & Entwistle, 1992; Roediger & Marsh; Scouller, 1998; Scouller & Prosser, 1994). However, Reiner et al. (2002) contend that such preparation might be more dependent upon instructors' expectations than simply on test format. Nevertheless, it is worth considering that deep-learning strategies, however they are inspired, can also lead to greater student satisfaction as well as better performance on higher order learning activities (Parmenter; Scouller & Prosser), and, as such, these study strategies could be an unexpected benefit of essay exams.

Academic integrity. Another conceivable benefit of extended-response essay exams is their potential to complicate traditional cheating methods. That is, students cannot simply memorize essay responses in advance of a test, or create a cheat sheet of sorts (at the very least, it is exceedingly more difficult!). As a result, such test items could reduce the incidence of academic dishonesty (Nilson, 2017).

Quicker construction. Finally, and perhaps most practically, essay exams in particular have the potential to be constructed relatively quickly, compared to multiple-choice exams (Clay, 2001; CRLT, 2016; Jacobs & Chase, 1992; Nilson, 2017; Ory & Ryan, 1993). As any instructor who has even attempted to construct a multiple-choice exam might attest, they can be quite time-consuming and challenging to design, especially those that assess higher order thinking rather than recall (Parmenter, 2009). Essay exams do not require the construction of lures or "distracter" responses (which can lead to the misinformation effect mentioned previously), or the crafting of a long list of questions. In fact, the challenge of creating multiple-choice exams can unintentionally result in more recall-oriented tests (Suskie, 2018) or drive instructors to "protect their questions" for future use by not returning graded exams to students (Parmenter), thereby preventing students from learning from their mistakes. As Reiner et al. (2002) contend, however, effective essay exams absolutely require thoughtful construction, just as effective multiple-choice exams do. For more information regarding effective essay-exam construction, see the later section, Designing Effective Essay Exams.

The Cons of Extended-Response Essay Exams

While essay exams certainly offer numerous advantages, they also include the following limitations.

Restricted content sampling. First, although essay exams may take less time for instructors to compose, time constraints are a factor in other ways for both instructors and students. Exams that consist entirely of essay responses can assess only a limited selection of course content (CRLT, 2016; Ory & Ryan, 1993; Parmenter, 2009; Reiner et al., 2002; Walstad & Becker, 1994). Essay exams necessitate a great deal of writing and response time for students per question and thus restrict the range of content that a given exam can sample. As a result, a student's performance or score might not reflect a

comprehensive knowledge of the course material but rather whether the "right" questions, or those that serendipitously matched with student's knowledge and preparation, were asked (Bean, 1996). In addition, those same testing time constraints are undoubtedly inadequate for fostering productive and thoughtful writing (Bean; Jacobs & Chase, 1992; Walvoord & Anderson, 1998); timed writing certainly does not emphasize process writing and is unlikely to produce a finely wrought essay.

Time constraints in grading. From the instructor's perspective, grading essay exams can be tedious and time-consuming, especially for larger classes (Clay, 2001; CRLT, 2016; Jacobs & Chase, 1992; Nilson, 2017; Reiner et al., 2002; Weimer, 2015). Unlike multiple-choice exams, essay exams cannot be graded quickly with a Scantron machine or simple answer sheet, and the variability in student answers can be a double-edged sword, allowing for latitude but also making the grading process more challenging. Consequently, instructors who grade a large number of essay exams often limit the inclusion of other forms of more effective writing assignments and activities in their courses (Bean, 1996).

Grading inconsistencies. Because the grading process can be so labor-intensive and mentally taxing, the grading of essays can also foster inconsistencies. Reiner et al. (2002) note the potential for variations or deficiencies in both inter-scorer and intra-scorer reliability .Bean (1996) as well as Jacobs and Chase (1992) also warn of the *halo effect*, or the propensity for a scorer's previous impression of a student to influence his or her grading. (In other words, if the instructor believes Jessica to be a good student, her assessment of Jessica's essay might reflect that bias, perhaps undeservedly.) *Location in the stack* can also potentially affect an instructor's response to an essay (Jacobs & Chase); the first essays read in a grading session often receive higher scores, perhaps because the scorer is not yet fatigued, and a reader's assessment of one paper can likewise be influenced by the quality of the papers previously assessed. For example, if an instructor reads a particularly unsuccessful essay, the next one he reads might seem incredibly cogent in comparison, even if it is actually somewhat weak.

Essay exams might also privilege good writers and reward neatness or other factors unrelated to content knowledge during the assessment process (Bean, 1996; Clay, 2001; Jacobs & Chase, 1992; Nilson, 2017). Nilson, however, also argues against cautionary tales about the so-called subjective nature of essay grading, asserting that they undermine the expertise of the instructor, "make a mockery of professional judgment, and give students the mistaken impression that faculty have no clear standards for evaluating their work" (p. 299). Essay exams might not offer the tidy dualistic structure of a selected-response exam (i.e., "right" and "wrong" answers), but with judicious design and the identification of clear evaluative criteria, assessing them does not have to be a free-for-all.

Designing Effective Essay Exams

Although essay exams might appear to be quicker and easier to construct than multiple-choice exams, they are instead deceptively complex and require just as much thoughtful preparation. What follows are some suggestions from the literature of best practices for devising extended-response prompts and essay exams.

Provide clear directions and articulate a well-defined task. It is not uncommon for students to feel as if they must fill an entire blue book to respond to an essay question, particularly when faced with a vague prompt (Reiner et al., 2002). As such, it is imperative to provide clear objectives and distinct tasks for students. Much like guidelines for composing measurable learning objectives, the literature also recommends formulating questions that guide students to the preferred approach, avoiding ambiguous directives such as *discuss* or even *describe,* which can elicit rambling responses (CRLT, 2016; Jacobs & Chase, 1992; Reiner et al.). Instead, instructors should embrace more defined action verbs, such as *justify*, *analyze*, *compare*, or *summarize*. (Bean, 1996, adds that such imperatives should be adequately contextualized for students.) For instance, providing a prompt such as "Discuss the

impact of the Dust Bowl" provides few cues to students regarding instructor expectations; what exactly does "discuss" mean to the instructor? What kind of impact? And on whom? A clearer version of this question might read, "Identify and explain the long-term impact of the Dust Bowl on the American economy." (See Figure 1 for additional examples of potential prompts across a variety of disciplines.)

Transparency in articulating the desired tasks, skills, and knowledge to be demonstrated in assignments has indeed been shown to lead to improved student confidence and success (Winkelmes, Boye, & Tapp, in press). Ultimately, students should not have to speculate about what their instructor wants them to do!

	Art	American literature	Psychology	Physiology
Weaker	Discuss the changes from modern to postmodern art.	Analyze Nick's role as the narrator in *The Great Gatsby.*	Explain humanistic theory.	How do diseases affect renal function?
Stronger	Compare and contrast modernism and postmodernism, identifying and explaining the conceptual differences between the 2 movements. Support your analysis by drawing on at least 2 different artistic works that exemplify each movement.	Construct an argument regarding whether Nick is or is not a reliable narrator in *The Great Gatsby* and how his perspective shapes the narrative for readers. Be sure to include specific examples from the text to support your position.	What are the basic assumptions behind humanistic theory, and how is it a reaction to the behaviorist perspective and the psychodynamic approach?	Identify 2 diseases that affect the kidneys, describing how and why each disease impacts renal function, as well as major symptoms that can help lead to their diagnosis.

Figure 1. Sample essay prompts across disciplines.

Specify expectations and scoring procedures. Just as transparency regarding tasks in assignments is important, so is transparency of expectations and grading criteria (Winkelmes et al., in press). As with any assignment, students will want to know not only the total point value of each response, but also how you will evaluate their work and what components you will prioritize. Transparency also includes clarity about your writing expectations; be specific about the role writing mechanics or other related factors will play in the assessment process (Clay, 2001; CRLT, 2016; Jacobs & Chase, 1992). For instance, will you take into account such elements as spelling, grammar, or use of references? Are you expecting a particular writing style or format? What kind of organization might you be looking for? Should calculations be labeled, or should comments be provided regarding coding?

Bean (1996) suggests that instructors should perhaps learn to live with micro level errors in timed writing exams and instead focus on content and macro level issues, such as thesis, organization, application of principles, and support of ideas, because essay exams are in essence first draft writings. Additionally, it can be beneficial to articulate your expectations to students well in advance, not simply upon grading or even at the time of administering the exam, to allow them ample time to prepare and ask for clarification. (For more information on evaluation procedures, see also the forthcoming section on assessment practices.)

Suggest time allocations and limitations. CRLT (2016) and Nilson (2017) also suggest that you plan for and articulate the amount of time students should spend responding to each essay question. Without adequate limits, students might provide responses that are too long, off task, or incomplete (Reiner et al., 2002). Nilson suggests estimating 15 minutes to one hour of completion time per essay question; more comprehensive questions should likely fall toward the higher end of that range, whereas more focused or limited-content questions might fall toward the lower end. As such, it is beneficial to overtly specify your time expectations for your students; there are vast differences between a 15-minute essay and a 45-minute essay! Furthermore, make sure that you are assigning a reasonable task and estimating a realistic response time for that task—and, of course, the more direct and clear that task is, the greater the likelihood that students will be able to respond effectively in the allotted time.

If possible, sample a range of course objectives and levels of cognitive domain. Although the time constraints of essay-test items can certainly limit content sampling, Jacobs and Chase (1992) still recommend that instructors strive to assess multiple objectives and "think over the spread of topics and the range of cognitive functions that were intended to be developed" (p. 104). For more concrete examples, see Figure 2, which provides some sample question stems focusing on a range of higher level cognitive domains, and which instructors might find helpful during the design process. Nilson (2017) likewise advocates for increasing fairness among students by covering more material rather than concentrating on a single area. However, this suggestion also leads to a bigger debate in the literature.

	Cognitive Domain				
	Lower ←			→	Higher
	Understanding	**Applying**	**Analyzing**	**Evaluating**	**Creating**
Sample question stems	Explain how . . . works. How is . . . an example of . . . ? Compare . . . before and after. . . . Describe an example of the principle of. . . .	Apply the rule/theory of . . . to. . . . Describe . . . from the perspective of. . . . How would you modify . . . ? What approach would you use to . . . and why?	What are some possible causes/ repercussions of . . . ? What ideas justify . . . ? What is the relationship between . . . and . . . ? What are the pros and cons of . . . ?	What criteria could you use to assess . . . ? Which details of . . . are most important and why? Rank the importance of . . . and explain your rationale. What is the most/least . . . ? Explain your reasoning. Do you agree with / that . . . ? Why or why not?	What information would you need to make a decision about . . . ? What might happen if you combined . . . and . . . ? Predict what will happen if. . . . How might . . . benefit/harm society? What do you think is the best solution to the problem of . . . and why?

Figure 2. Sample essay prompt/question stems for higher order thinking. Adapted from Anderson et al., 2001; Haladyna, 2004; and TeachThought, 2018.

Note. The lowest level domain, remembering, has been purposefully omitted here to focus on higher level domains.

Should you pose one or two longer, or more comprehensive, prompts or a larger selection of shorter, more focused prompts? This is one question that doesn't seem to have an obvious or agreed-upon answer. Nilson (2017) certainly favors several shorter prompts that address a larger assortment of material, as do Jacobs and Chase (1992) and Reiner et al. (2002). The rationale for this side of the debate addresses the issue of fairness mentioned previously and the need to assess students' breadth of knowledge, rather than students relying on luck to get a prompt that actually speaks to what they know. (If the exam poses only one question that addresses topic A, but the student is well-versed in topics B, C, and D, she is on the losing end of the assessment.) Many argue that this option supports better sampling and thus more valid assessment practices.

Others (e.g., Clay, 2001) favor essay questions that are more comprehensive in nature rather than those that focus on smaller units of course content. If your primary objective in using an essay exam is to assess the depth of students' knowledge or their abilities to synthesize course information and engage with the discourse in broader ways (particularly if pairing the essay with an objective portion on the exam), more comprehensive prompts might be more appealing, especially for instructors of upper level or graduate students. Bean (1996) further warns instructors to avoid numerous sub questions within prompts, even if well-intentioned, for these might confuse or overwhelm students, and some students might feel as if they must slavishly respond to each sub question rather than simply use them as inspiration for deeper thinking. If you do opt for a smaller number of broader, more comprehensive prompts, the need for clear directions and expectations becomes even more crucial.

Should you allow students to choose the prompts to which they respond? This is yet another question on which there does not seem to be a clear consensus in the literature. Several sources argue against giving students options (Clay, 2001; CRLT, 2016; Jacobs & Chase, 1992; Lukhele et al., 1993; Reiner et al., 2002), contending primarily that this undermines the validity and reliability of the assessment. Jacobs and Chase assert that such a test would not present a "common task," because all students are not asked to jump the same "hurdle" (p. 113), and therefore would not assess students equally. Lukhele et al. add that this practice allows students to avoid topics that they have not learned, and Reiner et al. note that some prompts may be more challenging than others, thereby privileging those students who choose the easier options. These arguments seemingly emphasize the measurement-focused perspective of assessment, perhaps prioritizing the need to compare all students against one another on a common scale or conceiving of exams as a tool for ranking students.

Others support the practice of giving students options. Nilson (2017), for instance, notes that such flexibility can allay student anxieties and allow them to demonstrate "the best of what they have learned" (p. 300). Writing-across-the-curriculum specialists might also champion a more learning-focused perspective and the desire to position and gauge learning in students' ability to participate in the discourse of the field (Bean, 1996). Ideally, to ensure that students are being tested on equal tasks, tests with choice embedded should offer parallel options in terms of both content and difficulty, yet that is likely challenging to achieve absolutely. Furthermore, both Bean and Nilson warn against giving students too many options, lest they waste time or energy making decisions, or even conflate multiple questions in their responses. In other words, if you would like to offer students choices, you should limit their options and allow them to respond to two out of three or four possible prompts, instead of giving them the leeway to choose two out of ten.

Consider posing prompts that call for thesis-governed writing and supporting evidence. Bean (1996) notes that students seem to offer the best responses to essay prompts that ask for a thesis that must be "supported, modified, or refuted," or present a single question for which the answer can serve as the writer's thesis statement (p. 192). Reiner et al. (2002) further suggest situating the essay task within a posed problem to add clarity and focus. For example, to frame the Dust Bowl example cited earlier as a problem to solve, an essay prompt might ask, "What

effects of the Dust Bowl are still impacting the American economy today, and what changes to policy or practice should be considered to mitigate that impact? Provide examples to illustrate." Reiner et al. likewise affirm that the prompt can be effective when posed as a question, as long as it is readily translatable into a clear task. Furthermore, requiring (or reminding) students to support their responses with specific examples and evidence will allow them to not only bolster their thesis, but also better demonstrate their command of the content (Clay, 2001) and potentially diminish the prospect of students bluffing their way through their essays.

Fair Assessment of Essay Exams

Grading essays of any kind can unquestionably be demanding and time-intensive, and, as discussed previously, without careful attention, the practice can fall victim to inconsistencies. However, several strategies are available for maintaining reliability in scoring as well as efficiency.

Develop clear and consistent grading criteria. First and foremost, it is valuable to establish a set of uniform criteria for scoring essays (Bean, 1996), not only for your students, but also for you as the instructor. Rubrics, or documents that articulate the expectations for an assignment, usually in terms of how those expectations will be graded, have been shown to teach as well as evaluate (Andrade, 2000), particularly when provided to students in advance. Rubrics assist students in understanding assignment goals and focusing their efforts, as well as help instructors guide and provide more informative feedback (Andrade, 2005).

Students themselves have reported that rubrics help them focus, produce higher quality work, feel less anxious, and earn better grades (Andrade & Du, 2005). As an instructor, defining your expectations in advance can also help you focus on what is most important during the grading process, and a rubric can by extension help ensure that you are focusing on the same criteria for all students. Figure 3 offers one example of what a general rubric for evaluating a written essay-exam response might look like, broken down by criteria and point value. However, there are many rubric variations that you might adapt and refine for your own purposes.

Criteria	Point value	Points earned	Comments
Thesis: Does the response adequately answer the question? Does it present a clear and logical position/argument based on an appropriate and accurate understanding of course material?	10		
Development: Does the response include sufficient relevant details and at least 3 examples from course material to support the thesis? Does it thoroughly explain the author's ideas and rationale?	10		
Organization: Does the response demonstrate a logical progression from one idea to another, with clear topic sentences? Does it stay focused?	5		
Grammar and mechanics of writing: Is the response generally clear and readable, without an abundance of distracting errors?	5		
Total points:	30		

Figure 3. Sample grading rubric for essay-exam written response.

Score one exam item at a time*, *and consider establishing benchmarks. If you are administering an exam with multiple essay questions, the research suggests that, instead of grading an entire exam before moving on to another, you should evaluate each response to a single prompt to stay focused and consistent (Bean, 1996; Clay, 2001; CRLT, 2016; Jacobs & Chase, 1992; Walstad, 2006). Furthermore, you might consider skimming all responses to a prompt and sorting them into piles based on level of effectiveness before marking or scoring any of them (Clay, 2002). A similar (and perhaps slightly less time-consuming) strategy is to read a random sampling of responses to establish benchmarks for grades and get a sense of what "typical" responses look like (Bean; Jacobs & Chase), thereby facilitating a more uniform and efficient grading process.

***Reshuffle the stack*.** When assessing responses to multiple items, the literature also suggests reshuffling the responses each time you move on to a new item to help counteract the effects of location in the stack (Jacobs & Chase, 1992). As a result, Student A's exam won't always be the first one read, Student Z's won't always be the last one read, and Student M might not suffer from the previous paper quality problem (Jacobs & Chase). By the same token, Suskie (2018) recommends reassessing the first few responses after completing the stack to guard against *rater drift*—i.e., double-checking to make sure your first few assessments are comparable to your last, and everything in between.

***Employ blind grading*.** Just as academics believe in the blind review process for scholarly publication, the literature supports concealing student names while grading essay exams to eliminate potential reader bias (Bean, 1996; Jacobs & Chase, 1992; Suskie, 2018). This can be accomplished by folding down the corner of the exam page where the student's name is listed, old-school style, or by asking students to write their names on the backs of exams; you might also employ student ID numbers generated either by you, the students, or the university, to be correlated with exams after grading is completed. Some online learning-management systems can even facilitate anonymous grading for you.

Provide feedback. While it does add some time to the grading process, offering commentary on your students' essay responses—even just a little—can go far to help them understand their grade and learn from the exam. Ideally, clear communication about what students did well or not-so-well in their essays can support them in monitoring their mastery of the course material and performing better on their next assignment or exam (Parmenter, 2009). Further, providing comments can also help you as the instructor remember your grading rationale, should any grade disputes arise (Clay, 2001). Fortunately, the use of rubrics can also help ease the burden of providing a lot of written commentary.

Operate with efficiency. Clearly, one major drawback to extended-response test items is the time-intensive nature of evaluating them. A rubric can certainly aid in ensuring not only the consistency of the grading process, as noted previously, but also its efficiency, by helping you focus on the most important criteria without getting bogged down in minutiae. To further foster efficiency, you might also consider using a timer. Once you have determined a reasonable average amount of time for grading each response, consider setting a timer for that amount, perhaps 5 to 10 minutes, to keep you aware of the clock and moving through the responses in timely fashion.

Other Suggestions for Student Success

***Purposefully prepare students for writing essay exams*.** One potentially overlooked component of student success is adequate preparation for the specific task of writing effective essay exams (Bean, 1996; Jacobs & Chase, 1992; Nelson, 2010). Nelson finds fault in the assumption that "students should come to us knowing how to read, write, and do essay and multiple-choice questions" and emphasizes that instructors should instead make efforts to teach students how to execute each of these foundational college skills, particularly within the conventions of their distinct disciplines (p. 181).

Fortunately, there are many opportunities and a variety of approaches for helping students learn how to write effective essay exams. For instance, it can be

worthwhile to set aside a bit of class time for students to practice writing sample responses, either on their own or in small groups. As the instructor, you can then offer to provide quick feedback on their responses or facilitate an in-class discussion of the process or desired attributes of a successful response. You might also coordinate a norming session, during which students apply the evaluation criteria to examples that represent a range of scores (Bean, 1996; Clay, 2001; Nelson, 2010). Jacobs and Chase (1992) also recommend helping students learn how to study content for an essay exam, underscoring the importance of making connections and focusing on core ideas rather than excessive detail.

Consider building in opportunities for process. Writing scholar Bean (1996) is, of course, an especially strong proponent of fostering the writing process, which can lead not only to better exams, but also to more focused student learning. While multiple drafts of an essay might be impossible within the constraints of a timed exam, there are several other simple strategies worth considering, one of the simplest being revealing potential essay prompts in advance (Bean; Murray, 1990; Mysliweic, Dunbar, & Shibley, 2005; Nelson, 2010; Parmenter, 2009). Some argue that this practice might simply test students' ability to memorize—especially those educators who favor the testing perspective on assessment, but it can also allow for useful thinking, prewriting, and organizing, resulting in more thoughtful and cogent student responses that make richer connections with the course content than those that might be produced on the fly.

Other suggestions from Bean (1996) include allowing students to revise exams or construct exam prep notes, or giving take-home exams with clear expectations for time and effort (i.e., should students spend several hours on their responses, or several days? Should their responses be 3, 5, or 10 pages long? Should they include external references? Are they allowed to draw upon class notes and texts, or partner with others?). And although some instructors might resist implementing take-home exams for fear of student collaboration or cheating, student collaboration is not necessarily a drawback when considering the end goal of student learning, because cooperative test taking can promote critical thinking and richer reflection, as well as diminish student anxiety (Dallmer, 2004). Further, the most effective take-home exams can allow instructors to ask higher level questions and encourage students to focus on analysis and evaluation, while also fostering a more thorough and thoughtful response process (Murray, 1990; Myseliweic et al., 2005).

Complete the exam yourself. Before administering your essay exam, it can be edifying to predict student responses to the prompts and even draft your own (Clay, 2001; Nilson, 2017; Reiner et al., 2002; Suskie, 2018). Doing so will allow for reflection on the exam's clarity and alignment with desired objectives as well as its feasibility for students, since as a content expert, you might demand unrealistic student responses or underestimate the amount of time needed to provide a thorough response. This practice might also help you discern your own expectations for responses and outline important points that may aid in the articulation of evaluation criteria. You might also consider sharing your response to a practice essay with your students to demonstrate what an effective, expert response might look like, and help further articulate your expectations for them!

Conclusion

There is no one-size-fits-all solution to assessment, and every format presents its own challenges. This is certainly the case for essay exams! Nevertheless, if essays seem to be a good fit with your course logistics and most aligned with your learning objectives, the research-supported practices summarized here should provide some guidance for implementing them with your students in the most effective way possible. And just like anything else that is worth doing—but particularly when it comes to teaching—a little bit of care and thoughtful planning can make your essay exam a valuable experience for all involved.

Allison Boye, Ph.D., is currently Associate Director of the Teaching, Learning, and Professional Development Center (TLPDC) at Texas Tech University. There she oversees several programs for both faculty and graduate students, provides consultations services for instructors, leads workshops and conferences, and teaches. In addition to generational student engagement and communication, her research interests include effective consultation practices, peer observation, graduate student development, syllabus development, and writing pedagogy, as well as burnout and self-care. Her work has appeared in publications such as *To Improve the Academy*, *The Journal on Centers for Teaching and Learning,* and *The Scholarly Teacher*. She is also co-editor of the book *Transparent Design in Higher Education Teaching and Leadership*, forthcoming from Stylus. She has served on the Board of Directors for the Professional and Organizational Development Network in Higher Education (POD) and the Texas Faculty Development Network, and she currently serves as Associate Editor for IDEA Center Staff Papers, and as a consulting editor for *College Teaching*. Allison earned her B.A. from William and Mary, her M.A. from Bowling Green, and her Ph.D. from Texas Tech University, specializing in contemporary American literature, Women's Studies, and Cultural Studies.

References

Anderson, L. W., et al. (Eds.). (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives.* Boston, MA: Allyn and Bacon.

Andrade, H. (2000). Using rubrics to promote thinking and learning. *Educational Leadership, 57*(5), 13–18.

Andrade, H. (2005). Teaching with rubrics: The good, the bad, and the ugly. *College Teaching, 53* (1), 26–30.

Andrade, H., & Du, Y. (2005). Student perspectives on rubric-referenced assessment. *Practical Assessment, Research, and Evaluation, 10*(3), 2–11.

Bean, J. C. (1996). *Engaging ideas: The professor's guide to integrating writing, critical thinking, and active learning in the classroom*. San Francisco: Jossey-Bass.

Center for Research on Learning and Teaching. (2016). *Best practices for designing and grading exams.* Ann Arbor: University of Michigan Center for Research on Learning and Teaching. Retrieved from http://www.crlt.umich.edu/P8_0

Clay, B. (2001). *Is this a trick question? A short guide to writing effective test questions*. Lawrence, KS: Kansas Curriculum Center.

Dallmer, D. (2004). Collaborative test taking with adult learners. *Adult Learning*, *15*(3–4), 4–7.

Entwistle, A., & Entwistle, N. (1992). Experiences of understanding in revising for degree examinations. *Learning and Instruction, 2*, 1–22.

Haladyna, T. M. (2004). *Developing and validating multiple-choice test items* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Haladyna, T. M. (2018). *Developing test items for course examinations* (70). Manhattan, KS: The IDEA Center.

Halpern, D. F., & Hakel, M. D. (2003). Applying the science of learning to the university and beyond: Teaching for long-term retention and transfer. *Change: The Magazine of Higher Learning, 35*(4), 36–41.

Jacobs, L., & Chase, C. (1992). *Developing and using tests effectively: A guide for faculty.* San Francisco, CA: Jossey-Bass.

Lukhele, R., Thissen, D., & Wainer, H. (1994). On the relative value of multiple-choice, constructed response, and examinee-selected items on two achievement tests. *Journal of Educational Measurement*, *31*(3), 234-250.

Minbashian, A., Huon, G. F., & Bird, K. D. (2004). Approaches to studying and academic performance in short-essay exams. *Higher Education, 47,* 161–176.

Murray, J. P. (1990). Better testing for better learning. *College Teaching*, *38*(4), 148–152.

Mysliweic, T., Dunbar, M. E., & Shibley Jr., I. A. (2005). Learning outside the classroom. *Journal of College Science Teaching*, *34*(4), 36–39.

Nelson, C. E. (2010). Dysfunctional illusions of rigor. *To Improve the Academy*, *28*(1), 177–192.

Nilson, L. (2017) *Teaching at its best: A research-based resource for college instructors* (4th ed.). San Francisco: Jossey-Bass.

Ory, J. C., & Ryan, K. E. (1993). *Tips for improving testing and grading*. Newbury Park, CA: Sage.

Parmenter, D. A. (2009). Essay versus multiple-choice: Student preferences and the underlying rationale with implications for test construction. *Academy of Educational Leadership Journal, 13*(2), 57–71.

Reiner, C. M., Bothell, T. W., Sudweeks, R. R., & Wood, B. (2002). *Preparing effective essay questions: A self-directed workbook for educators.* Stillwater, OK: New Forums

Press.
Roediger, H. L., & Marsh, E. J. (2005). The positive and negative consequences of multiple-choice testing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*(5), 1155–1159.

Scouller, K. (1998). The influence of assessment method on students' learning approaches: Multiple choice question examination versus assignment essay. *Higher Education, 35*, 453–472.

Scouller, K. M., & Prosser, M. (1994). Students' experiences in studying for multiple choice question examinations. *Studies in Higher Education, 19*(3), 267–279.

Suskie, L. A. (2018). *Assessing student learning: A common sense guide*. San Francisco, CA: Jossey-Bass.

TeachThought. (2018). 48 critical thinking questions stems for any content area. *Teachthought*. Retrieved from https://www.teachthought.com/critical-thinking/48-critical-thinking-questions-any-content-area/

Walstad, W. B. (2006) . Testing for depth of understanding in economics using essay questions. *The Journal of Economics Education, 37*(1), 38–47.

Walstad, W. B., & Becker, W. E. (1994). Achievement differences on multiple-choice and essay tests in economics. *The American Economic Review, 84*(2), 193–196.

Walvoord, B. (2014). *Assessing and improving student writing in college: A guide for institutions, general education, departments, and classrooms*. San Francisco: Jossey-Bass.

Walvoord, B., & Anderson, V. (1998). *Effective grading: A tool for learning and assessment*. San Francisco, CA: Jossey-Bass.

Weimer, M. (2015, September 22). Advantages and disadvantages of different types of test questions. *Faculty Focus.* Retrieved from https://www.facultyfocus.com/articles/educational-assessment/advantages-and-disadvantages-of-different-types-of-test-questions/

Wiggins, G. (2011). A true test: Toward a more authentic and equitable assessment. *Phi Delta Kappan, 92*(7), 81–93.

Winkelmes, M., Boye, A., & Tapp, S. (Eds.). (In press). *Transparency in higher education teaching and leadership.* Sterling, VA: Stylus.

Zeidner, M. (1987). Essay versus multiple-choice type classroom exams: The student's perspective. *The Journal of Educational Research, 80*(6), 352–358.

T: 785.320.2400
T: 800.255.2757

301 South Fourth St., Ste. 200
Manhattan, KS 66502

Email: info@IDEAedu.org
IDEAedu.org