

Carol Peterson Haviland
J. Milton Clark

WHAT CAN OUR STUDENTS TELL US ABOUT ESSAY EXAMINATION DESIGNS AND PRACTICES

ABSTRACT: The authors report a three-year study of basic writers' analyses of essay examination questions. The 336 students evaluated a series of sample questions, reported their experiences writing to one or more questions, and offered advice to faculty preparing to write other questions. The data reveal a number of preference patterns that may be weighed both against general teaching and assessment theory and against the particular constraints of specific contexts. The authors conclude that basic writers have much to contribute to the writing of essay examination questions as well as to the final choice of the design of assessment practices.

Our initial interest in the question, "What can basic writers tell us about the essay examination questions we pose for them?" was sparked several years ago by a story one of our colleagues, a medical school faculty member, told at a faculty dinner. He was describing the endocrinology rounds he had been conducting earlier that week and his concern with physicians' increasing reliance on technological data. He said that as the interns, residents, and attending staff gathered at bedside to discuss an elderly diabetic, they presented a brilliant battery of diagnostic information but seemed baffled about

Carol Peterson Haviland and J. Milton Clark are English faculty and codirectors of the basic writing program at California State University, San Bernardino. Professor Haviland's other interests are in writing centers and in the construction of discourse communities; Professor Clark's are in computers and composition and in writing assessment.

© *Journal of Basic Writing*, Vol. 11, No. 2, 1992

what to do as the woman's condition deteriorated. Finally he asked, "Has anyone talked with the patient? What has she told you?" The house staff looked at each other, again shuffled the impressive volume of lab reports, nurses' notes, and research abstracts, and then admitted that no one really had talked with the woman extensively—observing, questioning, and listening. Having trained in an era with less abundant technology and more dependence on history and physical examination, our colleague smiled and said, "Let's see what she can tell us; I suspect that she can help us learn quite a bit."

Clearly our arenas are different, but his question, "Has anyone talked with the patient?" stayed with us, for it focused some of our concerns about the essay questions we have been using for midterm and final examinations in our basic writing programs. These essays were written by students at a common examination time, scored holistically by basic writing faculty, and used both as part of the course grade and as exit criteria.

Like our medical school colleagues, we had assembled a great deal of data about our students. And, like some of our composition colleagues (Ruth and Murphy; Hoetker and Brossell, "A Procedure," Cashin), we had become knowledgeable about the examinations these students had written. We had read the literature, pretested our questions, watched students in the acts of writing, and analyzed the essays they produced. And all of this scrutiny had generated much useful information about question type, length, number, and complexity as well as about what students do while they are writing. But, like our medical school colleagues, we had failed to speak with students directly.¹ Because some of our classes served as admirable models of the student-student and instructor-student collaboration that build solid writing instruction and strong writers, we were particularly chagrined to see that, prior to this investigation, we had largely ignored what students could tell us about their essay examination writing experiences.

Our three-year study was designed to end that shortsightedness. For example, we were curious whether our questions were the clear invitations to produce students' best writing that we intended. We wanted to know whether our beliefs about personal narratives and about the desirability of two questions rather than a single one matched theirs. We wanted to know why certain questions produced thin or formulaic essays. We wondered what we could do to lower the anxiety levels with which students approached these essays, and how to make essay writing a learning experience as well as an assessment exercise.

From 336 students in our fairly typical Southern California state

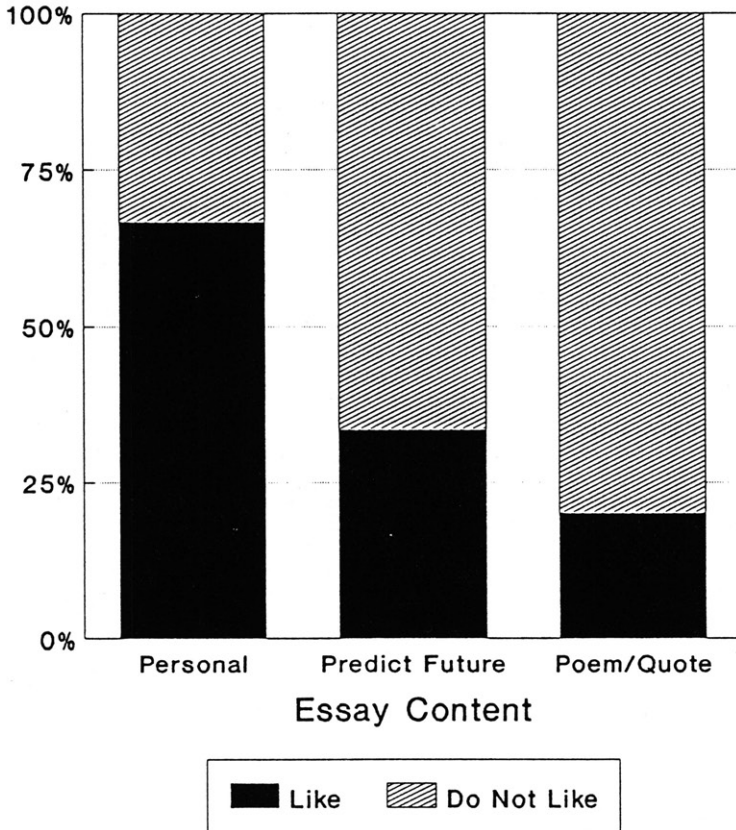
university basic writing population, a mix of Black, White, Latino, Asian, and American-Indian students with one-third of that number ESL in almost every section, we gathered data in three ways:

1. We gave a collection of 25 typical examination questions to four classes of students (56 students) and asked them to select the five they would most like to write on and the five they hoped never to write on; we then invited them to discuss their choices and their reasoning in class.
2. For eight consecutive quarters, on class days prior to essay exams, we asked all 336 students to make suggestions to the instructors who were writing questions.
3. During those same quarters, at the end of the testing periods, we asked students to record their immediate impressions of the questions posed, to describe their writing of the question they selected, and to compare the question they had just answered to other questions on similar examinations.

What we learned separates naturally into three categories: question content, phrasing, and number. (See Figure 1, showing student preferences for each of these.) The responses differed most widely in the content category. Foster suggests that to design a test question that will be accessible to most students, we can “assume no prior knowledge of a particular subject matter [on the part of the test taker]. . . . The only fair topic . . . is one appealing to general adolescent experience” (102). And, indeed, questions growing out of the realm of “general adolescent experience,” the personal experience question, are among the most common basic writing prompts; however, our students’ reactions to such questions varied greatly.

Two-thirds of our students offered some variation of the comment, “My own experiences are available and vivid; they are something I know and care about, so I can write a detailed essay.” Significantly, however, one-third of the students were very uncomfortable with personal questions. When offered two questions, one personal and one not, they chose the less personal one and stated that the topic’s distance from their private selves was the deciding factor. This discomfort was particularly acute for some students for whom sharing themselves with an unknown and thus untrusted audience violates cultural norms. Some of these same students were uneasy with questions that asked them to generalize their own feelings to other students or groups. While she was certain about how she perceived herself, Robin said, “It’s wrong for

Students' Preferences



N = 336

Figure 1

me to say how other people feel; I can only know for sure about myself.”

Their responses reminded us of some learning theories which tell us that combining a difficult task (writing) with a negative personal experience generally spelled failure. Very simply, they saw the combined activities (writing and recalling the experience) as too depressing; they preferred questions that did not ask them to write about failures, bad decisions, or regrettable incidents. Lara said, “I don’t want to relate experiences that make me look bad; if I have to do this, I might get so sad and embarrassed I couldn’t write. I know I won’t say much and you’ll write ‘more specifics . . . why did this

happen . . . explain' in the margin." Eva advised us to, "ask us questions that we are willing to answer, things that make us feel positive and capable, that we want to tell others about."

Students had mixed reactions about questions that asked them to predict the future. One-third liked this option, for it gave them a sense of power and allowed them to be creative. Tina said, "I like to be able to create my future; when I write about it, I come up with great ideas and it makes me think that I can do it." On the other hand, Jeff said, "How can I write about the future? I don't know what it will be until I do it, so I'd have to make it up." His comment reflects the majority view.

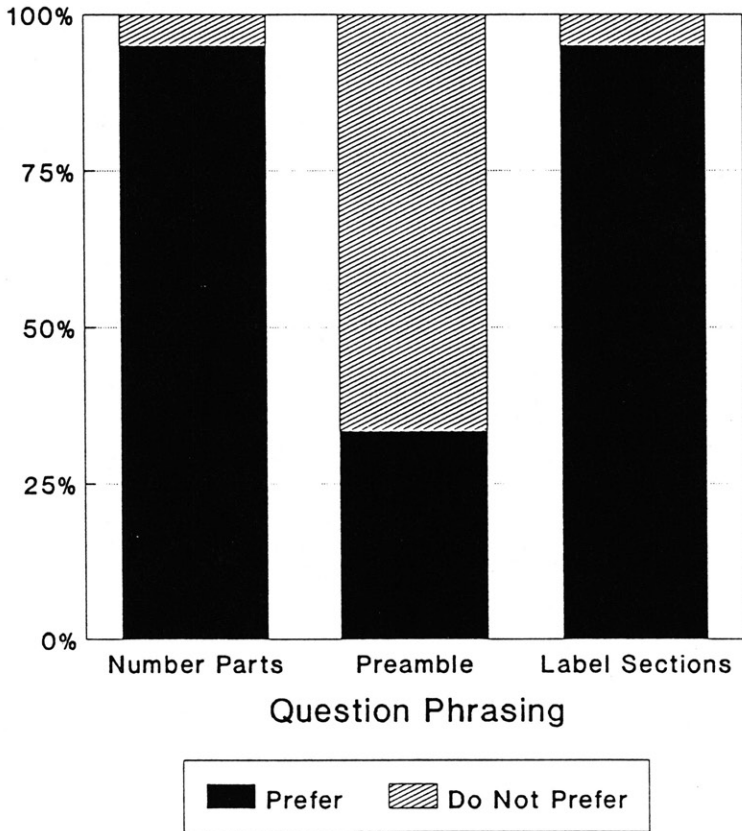
Questions that proceeded from a poem, quotation, or other text also received a mixed response. While a fifth of our students liked these questions, most found them both difficult to read and time-consuming. Interestingly, those who liked this sort of prompt preferred poetry to prose. Jose, for example, said, "The prose talks at me, tells me what to think, but the poetry makes me want to add to it, to create my own version of its pictures. Poetry is music and pictures; prose is just words."

Finally, our ESL students cautioned us that many of them had lived in the United States only a short period of time, so our questions stemming from "common knowledge," i.e., current events, geography, politics, television, and movies often drew blanks for those who had lived on the other side of the world until a few months ago. Tuan, who had come from a Thai refugee camp six months earlier, knew the geography of Vietnam, was fluent in French, and could explain Southeast Asian politics and history, but he hadn't seen the movies or read the books that examination writers thought were "common knowledge" to most U.S. college students.

In addition to their observations about content, students also commented about phrasing questions. (See Figure 2.) Previous studies on the effect of topic phrasing have been limited. Hoetker and Brossell found that holistic scores are not "directly affected by the amount of rhetorical specification in a topic, by the stance the topic invites the reader to take, or by any combination of these features" ("Effects of Systematic Variations" 418). Thus, they conclude that their data "do not support any of the hunches . . . about how low-ability writers can be either helped or hindered by variations in the wording of topics" (420). Kizner's study, in contrast, posits that topic effects do influence both student writing and scores.

Our data source, the students themselves, provided us with a number of phrasing suggestions that they believed would help them

Students' Preferences



N = 336
Figure 2

as writers. Uniformly our students noted how phrasing contributes to clarity, something they all worried about. They knew that they sometimes had lost credit because they had misread questions or failed to answer all of the question, so they asked that we “think like readers” when we write their questions.

Nearly all of our students argued that numbering the several question parts helped them keep track of those parts, though a few wisely observed that numbered questions encourage the “filling in the box” responses that we then told them really weren’t essays. Students also pointed out that questions asking them to tell the how and why of something often created overlap and they spent time

fruitlessly trying to separate them so that they answered all question parts. For example, Kirk said, "It's hard to write part three when I've really already said it in part two."

Essay examination questions are often presented with a preamble, a list of possible applications or other settings designed to stimulate or lead students' thinking. One-third of our students found this design helpful; however, we were surprised to learn that the majority found it derailing. Jessica, reflecting the views of those who liked such lists, remarked, "The question confused me, but the first example was one I experienced just that week, so I knew then what I could write." The larger group's views are illustrated by Rick's comment, "Once I read the examples, I couldn't get away from them, even though they really didn't apply to me. I wasted time trying to invent an experience like the example." The clearest response in this area was LaRonda's, "If the question is clear, you really don't lock us into using your preliminary stuff."

We also were quite surprised to learn that almost all students thought that we routinely phrased questions backwards, beginning with the preamble and ending with the question. Students uniformly believed that this order caused them to spend extra time hunting for the question, sometimes missing it, and always creating anxiety. John complained, "Sometimes when I read those long introductions to questions, I'm so confused that I can't think or write anything." Our students agreed that if we insisted on this order, at least we should label the sections READ and WRITE.

Uniformly, our students cautioned us about using vague or ambiguous terms, idioms, or jargon. When faced with a question that began, "Throughout history, older people and younger people have tended to blame each other for what is wrong with the world," Lisa, one of our 18-year-old students, put it bluntly, "Does old mean as old as you are; does young mean my age, which you think is young, or five years old, which I think is young? I don't want to insult the reading team by calling people their age old people, even though they're old to me!"

In his recent article, "Who Writes These Questions, Anyway?" John White provides two examples of such word-choice problems. The first example comes from an essay question asking students to discuss their "pet peeves"; it elicited one response in which the student "confidently described his 'pet peeve' and explained in detail how he put it on a leash and dutifully took it for a walk every morning" (232). Another example comes from a prompt that used a phrase many of us have used to explain what we want from our students: "provide concrete details in . . . [your] discussion." One

student obliged quite literally and “took the reader step by step through the process of making concrete” (233).

Finally, ESL students offered two directives: Lin said, “No matter how easy the words are to you, if you use too many or too big words, we can’t understand the question.” A second area of concern was brought to our attention by Kim, who pointed out that when given a choice of two questions, the topics were largely inconsequential: she elected the question that required the fewest departures from present tense verbs. Certainly those of us who have tried fumbling in a second language can appreciate this very pragmatic decision.

The third area of student concern was question number. In *Teaching and Assessing Writing*, Edward White details two major objections to giving students more than one question. The first is an issue of time; the selection process takes time away from planning, writing, and proofreading a response (105). White believes that if students have a limited amount of time to write an essay, it is unwise to have them spend part of their time choosing which question they want to answer.

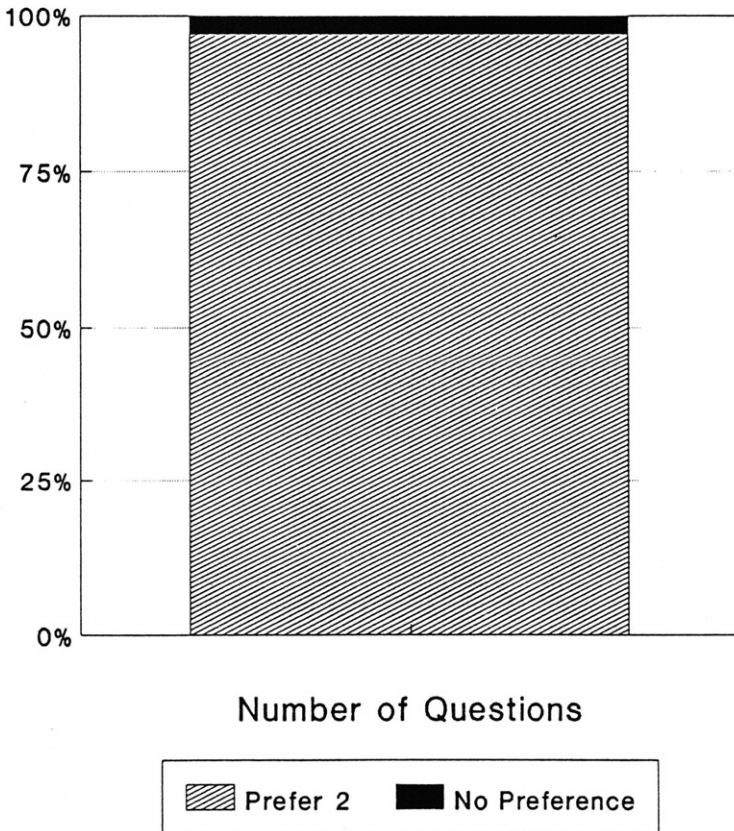
His second objection stems from multiple questions having different degrees of difficulty. He says that the more difficult questions “are the most interesting or most challenging and therefore the most attractive to the best students” (106). White posits that this variation in difficulty produces test results that do not accurately reflect the test taker’s skills within the group being tested, for strong writers may select the more difficult task and receive lower scores while weaker writers may select the less challenging task, write well, and receive higher-than-anticipated marks. Ruth and Murphy see the same kind of change in difficulty even within a single question (419), which suggests that while White’s position may be accurate, the issue may be more complicated than question number alone suggests.

However, even when we shared White’s reasons for using a single examination question, our students refused to back away from their plea for two questions. Anna said, “As soon as I see just one question, I freeze up because I’m afraid it will be one I can’t write on.” Vivian, a student of ours in several classes, laughed when she said, “I know that you all try hard to write good questions and that you’re not trying to trap us, but, face it, you blow it sometimes. If you write two questions, the odds are better that at least one will be writeable.” In fact, at the end of one examination with a single question on which an entire class had done well and had acknowledged that the question had been an excellent one, nearly

all still held the two-question preference (see Figure 3), stating that just seeing that two options existed reduced that initial anxiety.

Many research questions begin, as did this one, with a general concern about the quality of an item or an enterprise coupled with a curiosity about the role of an unknown factor. We believe that we have learned a great deal as we have begun to explore our largely unknown factor—what students can tell us about essay examination questions—by collaborating directly with those students. Methodologically, we are particularly impressed with the outcomes of the extended class discussions of sample questions. These seem particularly consistent with the ethnographic world view that

Students' Preferences



N = 336
Figure 3

emphasizes the import of the native, the actor in the scene, who can generate the “why” and the “whys of the whys” (Spindler 490). But while the natives (in this case, our students) are the experts, they may not recognize the implications of their knowledge and behaviors, and thus the collaboration is critical. Certainly our students continue to tell us much that would be difficult to uncover on our own, but what we do with these data is problematic.

At the simplest level, the data are mixed; if we make changes to satisfy some students, we may create difficulties for others. At a more complicated level, we must mesh student responses with our own writing/testing objectives. And, it is in this arena that following the lead of our medical school colleagues has been most fruitful, for it has caused us to examine our positions regarding pedagogy and assessment in the face of current theory and local context. Thus, while some of our specific findings or conclusions may apply to other essay examination settings, it is this collaborative process with students, rather than the details, that we wish to promote. We will illustrate with an example for each of the three categories of results: phrasing, number, and content.

The first example comes from the responses about question phrasing. Having our vocabulary and tense shifts scrutinized was illuminating and produced the least complex incorporation of student responses with our own writing/testing objectives. Student comments demonstrated that our word choices posed many more problems than we had expected. For example, we were careful to define ambiguous terms, such as, “old.” However, because the testing situation is also a forum for learning, we have not eliminated sophisticated vocabulary, but we are more careful to gloss, parenthetically, words or phrases that might perplex or mislead students, and we are particularly mindful of language (such as idioms) that may confuse an ESL writer.

We agreed to separate the question itself from its elaboration, as our students had requested, using the terms they suggested, “READ” and “WRITE.” In addition, we began to incorporate question-reading strategies into our instruction on in-class essays, for we also wanted to prepare students for those questions that might be written as a single unit in other courses. However, we decided not to meet the ESL students’ implied requests for questions that could be answered entirely in the present tense. Our freshman composition faculty expect students to move reasonably comfortably through multiple tenses, so rather than allow this feature to be hidden, we purposely write questions that will allow us to assess writers’ facility with tenses.

The second example involves the issue of question number.

While we agree with White's directives for large-scale assessment settings, we believe that our local context allows us to accommodate partially our students' requests for two questions.

Our university-imposed and difficult-to-shift time frames allow for 70-minute midterms and 120-minute finals. We have observed that students require the full 70 minutes to compose responses, leaving no time to choose between two questions, but that they have ample time to select and write in the 120-minute blocks. A second contextual factor is that even though all CSU campuses use a common English Placement Test, individual campuses have developed their own exit procedures. Student numbers are relatively small (150–250 Basic Writers/quarter), and our basic writing faculty is relatively stable throughout the year, and year to year. Thus, because we teach and read together frequently, we can accommodate difficulty-of-question differences that may emerge during readings.

Our context, then, allows us to respond to our students' anxieties about question number: we give a single midterm question but offer two on the final. However, this solution should be generalized only in similar examination settings at other universities. Were we, for example, to experience a faculty shift so that we became essentially a new group at each reading, it would be more difficult to accommodate question-difficulty differences. Also, if our midterms and finals had continued to be the principal exit criteria rather than, as they have become, simply two measures incorporated into the course grade, we might have been less comfortable with possibly reducing the reliability of the measure.

This shift in the use of the midterm and final leads to our third example. Our response to students' observations about question content is at once the most complex and the most interesting, for it best demonstrates the import of our decision to consult with students as we write essay questions.

While students differed somewhat on their preference for personal narratives over text-based questions, it seemed clear that Foster's prior-knowledge argument was sound and that students were correct in their belief that writing about positive experiences was easier than writing about negative or troubling experiences. However, because we concurred with Brossell's call for a meshing of assessment and pedagogical objectives, the question of our examination objectives remained: what were we trying to assess, or, even more important, what were we trying to teach? Did we want simply to see what our students could write under the least traumatic conditions? Or, did we want to see what they could produce under stress, how they worked through unresolved

problems? Were we concerned whether they could read questions, read text, respond to text, or all three? Until we could answer our own critical questions, we had no business writing essay questions. Thus, our students' responses forced us to reexamine our thinking about assessment, pedagogy, and essay examination questions, and this reexamination has produced several significant shifts. Although our students did not articulate their overarching anxiety about writing to *any* spontaneous prompt, regardless of how carefully it was composed, this theme became clear to us as we analyzed both the words and the tone of their advice. As we became increasingly willing to trust the collaborative process, we decided to more directly address their anxiety by changing the nature of the essay examination altogether.

When we worked to mesh our assessment with our pedagogy, it became increasingly apparent that having students write multidraft papers throughout the quarter and then using 70 or 120-minute impromptu essays as exit criteria was incongruent. Recognizing that our teaching commitments lay with the multidraft papers, we rethought assessment, concluding that common essay examinations could serve better if they were assigned more teaching value and less assessment value, and yet continued as a vehicle for faculty consensus. Thus, we began on our campus to use the midterm and final time blocks for what are now labeled in-class writing. Students receive a common reading several days before they write and then spend at least one class session discussing the reading and posing possible essay questions and answers. Afterwards, faculty collaboratively write the examination questions given to all sections, which include student suggestions about phrasing and number. Finally, faculty score the papers as before in group holistic sessions. However, in addition to serving their assessment function, these sessions have now become a forum for raising pedagogical issues and for solidifying faculty consensus.

This new protocol offers several advantages over the old. First, these in-class essay examination experiences are more transferable to students' other writing occasions than is practice in writing impromptu essays. Students see the reading-based prompts as similar to those used in their economics, geology, or music courses and begin to view posing and writing such essay questions as a rational, predictable activity. Second, as students see essay examination writing as something they can understand and prepare for rather than as a guessing game for which they can only become anxious, they write more purposefully and successfully. Third, now that scoring sessions no longer focus solely on exit determinations, faculty use them to generate and air other important issues. For

example, one reading provoked an important disagreement over ESL scoring, a hotly debated and rapidly shifting theoretical and social question in Southern California's multicultural population.

In sum, our results suggest that collaborating with students as we write essay examination questions can allow us to write better questions and also create more productive writing environments, ones in which students are more likely to be engaged and to write as well as they are able. In addition, a number of our students have commented that being invited to discuss examination questions and seeing their suggestions incorporated gives credence to our assertions that what they say as writers matters, a particularly important event for basic writers who have had few such experiences. Thus, we agree with Hannah Arendt that, "For excellence, the presence of others is always required" (qtd. in Ede and Lunsford 126). However, even as we commit to working closely with essay examination writers, it is important that we not simply rush to follow each of their directives. Rather, we must allow our collaboration to expand and then to answer jointly conceived questions (Ede and Lunsford 125), in this instance meshing students' suggestions with instructors' testing objectives, incorporating both the existing body of composition theory and local contexts.

Note

¹ The single exception we have discovered is Hayward's 1988 CCCC presentation.

Works Cited

- Brossell, Gordon. "Essay Test Topic Development." ERIC, 1986. ED 279002.
- Cashin, William. "Improving Essay Tests." Manhattan, KS: Center for Faculty Evaluation and Development, 1987. ERIC ED 298151.
- Ede, Lisa and Andrea Lunsford. *Singular Texts/Plural Authors: Perspective on Collaborative Writing*. Carbondale: Southern Illinois UP, 1990.
- Foster, David. *A Primer for Writing Teachers: Theories, Theorists, Issues, Problems*. Upper Montclair, NJ: Boynton, 1983.
- Hayward, Malcolm. "The Rhetoric of Essay Test Prompts." ERIC, 1988. ED 297363.
- Hoetker, James and Gordon Brossell. "A Procedure for Writing Content-Fair

- Essay Examination Topics for Large-Scale Writing Assessments." *College Composition and Communication* 37 (1986): 328-35.
- . "The Effects of Systematic Variations in Essay Topics on the Writing Performance of College Freshmen." *College Composition and Communication* 40 (1989): 414-21.
- Kizner, Charles. "Effects of Topic and Response Variables on Holistic Score." *English Quarterly* 20 (1987): 106-20.
- Ruth, Leo and Sandra Murphy. "Designing Topics for Writing Assessment: Problems of Meaning." *College Composition and Communication* 35 (1984): 410-22.
- Spindler, George. Concluding Remarks. *Doing the Ethnography of Schooling*. Ed. George Spindler. New York: Holt, 1982. 489-96.
- White, Edward M. *Teaching and Assessing Writing*. San Francisco: Jossey-Bass, 1985.
- White, John. "Who Writes These Questions, Anyway?" *College Composition and Communication* 39 (1988): 230-35.