THE STUDENT ESSAY AS DUBLOON: DISCREPANCIES IN HOLISTIC EVALUATION

DONALD A. DAIKER

In what several commentators take to be the central chapter of *Moby-Dick*, Ishmael describes the varied reactions to the gold coin—a dubloon—that Captain Ahab nails to the mainmast as a reward for the first sailor who sings out for the white whale. But "The Dubloon" does more than record the contrasting ways in which Ahab, Starbuck, Stubb, and other members of the *Pequod* respond to the precious coin; it also suggests that *any* complex or valuable object is likely to evoke amazingly different responses. And if the object contains writing or other forms of symbolic expression—if the object becomes, in other words, a *text*—then the responses are likely to be even more discrepant.

The dubloon does in fact become a text for Ishmael and his shipmates. Appropriately so, because the coin consists of both words and symbols. It bears the inscription "REPUBLICA DEL ECUADOR: QUITO" as well as signs of the zodiac and images of three Andes peaks. After the Manxman reads the zodiac signs as prophesizing the destruction of the *Pequod*—"Ship, old ship! my old head shakes to think of thee"—Stubb comments, "There's another rendering now; but still one text."

Indeed, the text of the dubloon produces a different rendering from each member of the crew—largely because each projects

himself upon the object he views. In the signs of the zodiac, Stubb sees support for his cheerily fatalistic philosophy whereas Ahab finds reinforcement for his own pain-centered view of life. Once Ahab goes below, Starbuck says, "Let me read." But it's as if Starbuck were reading an entirely different text: he misses the zodiac altogether and looks instead at the Andes peaks and valleys. There, in the contrasts between height and depth, light and dark, he sees a reflection of the tension between faith and doubt in his own soul. Flask's reading of the dubloon is different still: he sees the coin only in material terms, only as the means of buying himself 960 cigars. As the other crew members in turn advance to examine the dubloon—the Manxman, Queequeg, Fedallah, and finally Pip—the message repeats itself: what we see and what we understand and what we value is colored by our beliefs, our temperament, our experience, our hopes and fears. When Pip summarizes the chapter with "I look, you look, he looks; we look, ye look, they look" and then comments, "And I, you, and he; and we, ye, and they are all bats," he is playing upon the cliche "blind as a bat" to imply that each of us is blind to what the other sees.

Holistic Evaluation

The message of "The Dubloon" tends to make those of us who participate in the holistic evaluation of student essays a bit uneasy. After all, the premise of holistic evaluation is that readers of like background and experience can be trained to look at a text in pretty much the same way. English teachers can be trained—Paul B. Diederich, Charles R. Cooper, and Edward M. White assure us—to assign an essay the same score as our fellow readers, at least in the vast majority of instances. According to Cooper, "When raters are from similar backgrounds and when they are trained with a holistic scoring guide—either one they borrow or devise for themselves on the spot—they can achieve nearly perfect agreement in choosing the better of a pair of essays; and they can achieve scoring reliabilities in the high eighties and low nineties on their summed scores for multiple pieces of a student's writing" (19).

There is now widespread agreement that the holistic evaluation of student essays provides the most valid and reliable form of writing assessment. Yet even the most enthusiastic proponents of holisticism acknowledge that what works well for groups may not always work

well for individuals. White, for example, speaks of the "real possibility of substantial disagreement about any one particular paper" (178). He cites a research study in which 699 essays from the 1976 California State University English Equivalency Examination were rescored by a similar but not identical group of trained readers one year later. Although the two holistic scoring sessions produced a "remarkable stability of scores" for the group as a whole, the scores for eight percent of the students "changed radically" as a result of the rescoring (180). That is, one out of every twelve students produced writing that—like the dubloon—drew markedly different reactions from those who looked at it.

One important question raised by White's study is whether certain papers can be said to "invite" discrepant readings. Are there student essays that, by their very nature, cause even trained readers from similar backgrounds to see them in significantly different ways? If so, does it make sense to talk about a discrepant essay as well as, or instead of, a discrepant reading? Are some student essays the epistemological equivalent of Ahab's dubloon?

THE WRITING ASSIGNMENT AND RATING CRITERIA

To learn whether certain student essays produce widely diverging assessments from trained readers, we conducted a study during the 1986 Early English Composition and Assessment Program (EECAP).[1] Organized by Miami University's Ohio Writing Project for the Ohio Board of Regents, EECAP evaluates compositions written under controlled conditions by high school juniors. The compositions are scored by high school teachers trained in the procedures used by the Educational Testing Service for scoring Advanced Placement (AP) examinations. That is, the teachers learn to use a scoring scale and to apply rating criteria by reading, scoring, and then discussing a wide range of sample essays.

Both the sample essays and the "live" papers were written in response to the following assignment, which students had 35 minutes to complete:

> If you think about it, you're really not the same person you were four or five years ago. Your ideas, tastes, attitudes, and perhaps even your goals have changed—probably in several ways. Choose any one person (a relative, a teacher, a friend, or anyone else) or any event or experience (a course, a trip,

a conversation, or any other event or experience) that has made a difference in your life, and explain as fully as you can how the person or event has changed you. Be as specific as you can in showing how you are different now because of the person or event.

All papers were evaluated on a standard six-point scoring scale described in Table 1:

TABLE 1. Scoring Scale

Grades	Description
Grade 5/6:	Clearly above-average papers that demonstrate strength in virtually all the criteria. Rarely are these flawless papers, but they are usually substantial in content and often original in idea and/or expression. A "5" tends to be thinner or weaker in some ways than a clearly superior "6."
Grade 3/4:	Papers ranging from slightly below average ("3") to slightly above average ("4"), either combining strengths with weaknesses in the various criteria or showing an over-all sense of under-development.
Grade 1/2:	Clearly below-average papers which fail to demonstrate competence in several of the criteria (often because the paper is too short) or which are generally empty or which fail to respond to the question. A "2" tends to have redeeming qualities absent in a "1."
Grade O	Papers which are *wholly* off topic. Such papers neither state nor imply that a change of any kind has taken place.

The rating criteria of the scoring scale were 1) ideas 2) supporting details 3) unity and organization, and 4) style. These criteria were defined for the raters in Table 2:

TABLE 2. Rating Criteria

Criterion	Description
Ideas	The extent to which the thoughts and content of the essay are original, insightful, and clear.
Supporting Details	The extent to which the ideas of the essay are supported by examples and details which are specific, appropriate, original, and well developed.
Unity and Organization	The extent to which the parts of the essay are all connected to each other and all help achieve the goal of the essay.
Style	The extent to which the language of the essay is used creatively and correctly and helps achieve the writer's goals.

LOCATING "DISCREPANT" PAPERS

Once the training period was completed and readers began scoring "live" papers, we began to look for discrepant essays. First, we set aside essays whose two holistic scores differed by three points or more on the six-point scoring scale—essays, in other words, that received one of the following sets of holistic scores: 1 and 4; 1 and 5; 1 and 6; 2 and 5; 2 and 6; or 3 and 6.

We next sought to determine if the wide difference in scoring was caused primarily by a single reader who, for one reason or another, had not been properly "calibrated." To do so, we distributed the initially discrepant essays to two additional readers. In most cases, the third and fourth readings confirmed one of the two initial scores; that is, if the paper had initially received, say, a 2 and a 5, the third and fourth readings would likely produce two additional scores of "5" or two additional scores of "2." In these instances what we had was not a discrepant paper, but a reader who was not using the rating criteria and scoring scale in the same way as the rest of us.

But we did locate several papers that seemed discrepant even after a third and fourth reading. These were papers that fell into one of the following three categories:

1. Two pairs: the paper received two pairs of scores that differed by three or more points (two 1's and two 4's; two 2's and two 6's).

2. Four different scores: either a straight (1,2,3,4; 3,4,5,6) or an even more varied series (1,2,3,6; 2,4,5,6).

3. A three- or four-point spread: with two scores differing from the two other scores by at least two points (1,2,4,4; 2,3,6,6,).

We chose four papers from the above categories, and we asked all 61 participating readers to read each paper and then to complete a "Rater Questionnaire" (Table 3). On the questionnaire readers assigned each paper a holistic and then four analytic scores, and they responded to four questions that we hoped would help explain their evaluation.

TABLE 3. Rater Questionnaire

Paper #39215060

1. What holistic score (1-6) would you give this paper?__

2. What analytic scores (1-6) would you give this paper?
   Ideas__ Supporting details__
   Unity & Organization__ Style__

3. Please explain as clearly and specifically as you can the major reasons for your holistic score.

4. Was this an easy or difficult paper for you to score? Please explain in detail.

5. What would you say if you knew that several raters gave this paper a score at least two points different from your own score?

6. Any other comments or responses?

Here is the student paper that, when the rater questionnaires were analyzed, turned out to be the most discrepant of all:

> During my fifth grade year in elementary school, an event occurred that changed my life. My brother, having fought a

long and severe battle with cancer, died at the youthful age of twenty-three. Being the youngest in a family of six children I was often doted on by my older brothers and sisters, and I enjoyed being in their company immensely. When my oldest brother fell sick nearly a year before his death, my families' concerns and thoughts focused on his welfare. Too young to understand the complicated reports on his conditions, I was never completely aware of how serious his illness had become. Bused back and forth to my other siblings' homes I missed the attention that I was so used to receiving. I spent as much time as was possible with my brother during his short stays at home, and waited endless hours in hospital lobbies for relatives visiting him since I was too young. I enjoyed playing nurse with him while he laid helplessly in bed. I did not know how much he was hurting because he never complained and he always was willing to give me love and attention. I began to mature during those long periods without my mother's sole attention and without the cares of my brothers and sisters who had other things on their minds. I did not resent them for shifting their attention from me to him, he was the most wonderful person I ever knew. I wish I could thank him for the opportunity that it gave me to mature from a spoiled brat to a loving young adult. I will never forget him, he will always occupy a large part of my heart. He gave till it was no longer possible and provided my family and I with one of the most valuable lessons in life—love.

ANALYSIS OF THE RATER QUESTIONNAIRE

Analysis of the 61 rater questionnaires confirmed what the first 4 readings had suggested: this is a paper that, like the coin nailed to the masthead of the *Pequod*, evokes very different responses from those who look at it. As Table 5 makes clear, four different scores on the six-point scale—"3," "4," "5," and "6"—were relatively popular with the readers. Although "4" was the most frequently assigned score, it was chosen by well under half the readers. Even more to the point, the readers could not agree whether this was an upper-range or middle-range paper: 42.6% of them placed it in the upper range, 57.4% in the middle or lower range. If one out of every six readers gave the paper a "6," the highest possible

score, exactly the same number gave it a "3" or "2," scores appropriate for papers either "slightly" or "clearly" below average. What was it about the paper, we began to ask ourselves, that led to such significant differences in the responses of trained readers? What was it that made this student essay as slippery in value as Ahab's dubloon? To begin answering these questions, we analyzed the written responses on the rater questionnaire.

TABLE 4. Scoring Results

Holistic score	Frequency	Percent
6	10	16.4%
5	16	26.2%
4	25	41.0%
3	9	14.8%
2	1	1.6%

What became clear, first of all, was that there were significant areas of agreement among the raters. There was a general consensus, for example, that the addition of concrete, specific details would have strengthened the paper. As Debbie Meyer put it, "I can see how the death of a brother would change your life, but the writer does not really show the changes using specific detail." Most raters agreed with Carole Knight that the paper "needs more thorough use of detail and more development." Thelma Swihart speaks for the vast majority of raters in saying that "the paper had a good beginning and ending with too little in between."[2] But there were other areas of agreement as well. Almost every rater agreed that the paper was on topic, that its chronological organization was effective, and that the details that were present—"bused back and forth," "endless hours in hospital lobbies"—constituted one of its strengths. There was further agreement that the paper would have improved had it consisted of more than a single paragraph and that in diction, usage, and sentence structure it was at least adequate. The one issue which most consistently created disagreement was the paper's emotional impact.

Readers who scored the paper a "5" or "6" found the story of the brother's death to be a moving one. Bob Dizney referred to the "emotional impact" of the piece, and Sue Sherman called it

"gripping." Sister Marie Irene Schneider described the essay's ending as "a powerful culmination" of the writer's changes. Bob Kiracofe acknowledged that the essay "drew me to it" and "made me change along with the writer." As Tally Harwood saw it, "the writer obviously experienced trauma," and "the emotions of the writer came through." Nancy Simmons testified that she "felt the writer's pain. If a writer can draw me into a feeling of sympathy," she continued, "I think they are doing a superior job of writing."

But there were many readers who were not moved by the story, and they tended to assign the paper a "3" or "4." Tina Meinke asserted there was "No Way!" this could be an upper-range paper because of "too little supporting detail and human emotion." Carol Schiavone reacted much the same: the essay "lacked the detail and 'emotion' a paper of this type would need to really grab me." Cindy Briggs thought the subject "could have led to powerful writing and it just didn't deliver it." For Winifred Cairns, the paper simply failed to "provide 'moving' moments of the brother's illness."

Readers who scored the paper a "5" or "6" felt that much of its power came from its honesty and genuineness. For Sister Mary Dolores Schneider, the story's "warmth" and "sincerity" was a major reason it earned the score of "6." Gary Gilmer "liked its honesty," and several readers agreed with Brenda Pansing that "the student's voice really comes through." Sharon Freyhof admired "the way this writer makes her brother (and the rest of her/his family) seem human and real without gushing over how wonderful or heroic they were and without dwelling on the tragedy." As Freyhof saw it, "This kind of control identifies a mature writer who can distance himself from a situation but still share the feelings with the reader."

But there were just as many readers who were turned off by what they perceived as the essay's insincerity. Although Marilee Pallant was impressed by the writer's insight, she "was not convinced that the piece was honest." The details "did not feel authentic" to her. Teresa McGowan reacted in much the same way; she "felt that there was something in the tone . . . that smacked of insincerity." For Dick Arthur, too much of the essay is "sentimental," for Tina Meinke the essay is "stilted," and for DJ Hammond "some of the sentiments were (rotten to say about someone's brother's death) cliches." Phyllis DeMarco speaks for most of the readers who sensed something artificial in the essay when she says, "I hear control—what the writer believes she *should* say and feel. What I

don't hear is real emotion and real wrestling with experience."

Between those who were touched by an essay they found sincere and those untouched by an essay they found artificial lies a group of readers who had trouble deciding whether they had in fact been moved by the essay and, if so, whether they *should* have been. For example, Linda Balmos explained that she "had to bypass the 'emotional pull' of the writer's experience . . . to remain objective in my rating." Lori Newton makes much the same point in saying that, because it is "an emotion-provoking paper, I had a tendency to score too high because of what *happened*, not because of the writing." Virtually the same response came from Lee Ann Adams: "I vacillated back and forth on this (4 or 5?). It's hard to separate the emotional reaction you have to the subject from the actual content of the writing." Diane Rawlings' reaction was similar: "The experience of a brother's death evokes emotional responses from me which can cloud my judgment when evaluating the writing performance. In other words, I tend to respond to the experience, rather than the effectiveness of her writing." As Jean Gilles cautions, "you have to be able to divorce yourself from granting a grade just because the paper touched the strings of your heart." But all five of these raters apparently succeeded in resisting the essay's "emotional pull": they each gave it the holistic score of "4."

What analysis of the rater questionnaires makes clear, then, is the wide continuum of emotional response to the central event of the student essay: there were readers who were deeply affected by the brother's death, there were readers who felt nothing at all, and there were readers at every stage in between. Although such a range of responses will not surprise teachers who have listened carefully to the reactions of their students—and colleagues—to the death of Arthur Dimmesdale or Edna Pontellier or Captain Ahab, it must give pause to those of us committed to holistic evaluation. It must be especially disconcerting when analysis of the rater questionnaires suggests a strong connection between depth of emotional response and holistic score. The more a reader is moved by the brother's death, the higher the holistic score she or he assigns.

No doubt there are complex reasons why readers differed in their emotional response to the student essay and in their holistic score. But surely Ahab is right in asserting that what we see is in some sense a reflection of what we are. For Ahab, the dubloon

"is but the image of the rounder globe, which, like a magician's glass, to each and every man in turn but mirrors back his own mysterious self." In fact, several readers made much the same point on their rater questionnaire. In response to the question "What would you say if you knew that several raters gave this paper a score at least two points different from your own score?" Marilee Pallant wrote, "Some teachers would like the grammatical correctness. Others would be turned off by the 'constipated' emotional state of affairs in the piece." In answer to the same question, Phyllis DeMarco ventured that "The organized, proper person who wants all people to 'line up in order' will love it. The disorganized searcher like me will be saddened. Where did the nosy wonder of childhood go?" Like Starbuck, Stubb, and Queequeg, English teachers sometimes project their "own mysterious self" upon the object they are viewing.

IMPLICATIONS FOR HOLISTIC TESTING

Knowing that what we value is colored by what we are, those responsible for holistic scoring sessions have a special obligation for ensuring fairness to our students. Of course, the writing of every student essay should be evaluated by more than a single rater. But, as the scoring of the student essay above demonstrates, multiple scoring in itself is no guarantee of fairness—especially if the student has chosen to write about an emotionally-charged issue or event. The student who wrote of her brother's death might as easily have received the summed score of "7" or "11" from a pair of readers: the first score would have identified her paper as average in quality; the second as superior.

Like multiple scorings, multiple writing opportunities are a second means of helping to achieve fairness in holistic evaluation. White is absolutely right in asserting that no single test—or no single piece of writing—"is sufficiently reliable to be depended upon, by itself, for a major decision about students" (43). But results may not be reliable even in a testing situation when each student writes two separate essays and when each essay is scored twice: these were precisely the conditions under which results "changed radically" for one out of every twelve students during the rescoring of a California State University English Equivalency Examination (180).

Because it now seems clear that certain essays invite discrepant readings, it seems equally clear that we need to build an appeals

procedure into any holistic evaluation program. The more consequential the test—a proficiency test required for graduation is obviously more important than a placement examination—the more students are entitled to appeal its result. In deciding whether to appeal, students will of course need access both to their test essays and their test scores—as well as to scoring procedures and criteria.

That a student essay occasionally elicits responses as widely divergent as Ahab's dubloon should not be taken as grounds for eliminating the direct assessment of student writing. Direct assessment, despite its difficulty and complexity, remains the most valid measure of writing quality. Indeed, if we accept Lee Odell's definition of writing competence as "the ability to discover what one wishes to say and convey one's message through language, syntax, and content that are appropriate for one's audience and purpose" (103), then Davida Charney is right in arguing that no indirect assessment of writing—no so-called "objective" or "standardized" test—can adequately measure competence. Whatever its limitations, the direct assessment of writing through holistic evaluation—either general impression or primary trait scoring—continues to be the profession's best hope.

IMPLICATIONS FOR CLASSROOM TEACHING

There is every reason to believe that the discrepancies which occur during holistic scoring sessions occasionally occur in our classrooms as well. In either setting, we will evaluate certain pieces of writing in substantially different ways than one or more of our colleagues. Indeed, discrepancies in grading seem even more likely when we know who wrote the paper and when students freely choose their topics—not to mention the absence of shared criteria and anchor papers. When my disappointed student implied that her essay on The Talking Heads would have fared better than "C+" with another instructor, odds are that several of my colleagues here in Bachelor Hall would confirm her judgment.

Should those of us who teach and grade writing simply accept the truth that, even within the same institution, evaluation criteria and their application differ widely from instructor to instructor? Perhaps so. It just seems a lot harder for a student to accept such a truth. It may be, as I remember hearing long ago in my Roundtable for College Composition Instructors, that "*Fair* is an

unrewarding concept in education." But I can't help believing that we shouldn't at least be moving in the direction of fairness. Indeed, I have recently become convinced that writing instructors need to establish the same kind of appeals procedure that should be part of any larger program in writing assessment. Here, in brief, are three different appeals procedures that can be made an integral part of a college composition course. They may be used separately or in conjunction with each other.

1. Make an informal contract with a colleague whose mind and values you especially admire. You will each evaluate papers whose grades have been appealed by the other's students. Set up whatever conditions you choose. Must the student write a statement to accompany the paper to be reevaluated? Are there any limitations or deadlines? How will the final grade be determined—an average? the higher score? There will be fewer papers to grade than you think, but class morale will improve remarkably.
2. Establish a team grading program at your institution. At Miami University,instructors of freshman English voluntarily meet twice a semester—one Saturday at midterm, one Saturday just before finals week—to evaluate student papers from sections other than their own. Before scoring begins, we train ourselves for an hour or more in holistic evaluation, principally by reading and discussing sample essays. Every paper submitted is scored twice on the standard letter scale, and instructors are free to use the results in any way they choose.
3. If you trust your students more than your colleagues—after all, they've been in your class for at least several weeks—let them re-evaluate any paper whose grade has been appealed. Set up three-person student evaluation committees on an *ad hoc* basis: that way you can select the committee members who are likely to be least biased and most competent for any given paper. Ask each committee member to submit a signed, written statement (the original to you and a copy to the appellant) along with the grade. In return, be sure to reward the committee members in some way for their time and work: you can bet that they've

learned a good deal from the evaluation task they've performed.

* * *

The action of "The Dubloon" in *Moby-Dick* involves more than a crew member looking at the coin, interpreting it in his own private way, and then retreating so that another sailor can repeat the process. Within this pattern is another: several crew members remain to watch and overhear the responses of their fellows. Stubb sees Ahab and Starbuck as they leave the dubloon, and then he stays behind, hidden by the try-works, to listen to Flask, the Manxman, Queequeg, Fedallah, and Pip. By listening to his shipmates, Stubb gains insight both into the others—he realizes for example, that Pip is "crazy-witty"—and into himself: he acknowledges his guilt feelings—"I could go hang myself"—for having earlier abandoned Pip during a whale chase. Pip too has observed the others, Stubb included, and for him this experience has driven home a central truth: all human beings, and especially blacks, are separated from each other in fundamental ways. As the novel's narrator, Ishmael has of course watched over all, and as readers, we stand one step behind Ishmael, looking over his shoulder, as it were. For Ishmael, and presumably for the reader as well, there is a gain in wisdom in witnessing the contrasting responses to the dubloon. From the wealth of different responses we glimpse the uniqueness of every human being, and we begin to place greater value on what is different and unusual and strange. In this way we gain new perspectives—that is one of the reasons Ishmael goes to sea in the first place—and we are educated in the virtues of tolerance and open-mindedness.

By the same token, the student essay as dubloon is more than a problem to be solved by the holistic scorer or the classroom teacher, although on one level it is surely that. The paper that consistently invites discrepant responses is no less a reminder than Ahab's gold coin that none of us responds to any text in exactly the same way and that "calibration" is not always the highest goal in human affairs. Without at least an occasional student essay to invite our most wildly checkered responses, we would experience less of what Fitzgerald's Nick Carraway, simultaneously repelled and enchanted by the world about him, calls "the inexhaustible variety of life."

Donald A. Daiker is Professor of English at Miami University and Secretary-Treasurer of the National Council of Writing Program Administrators. He is co-author of *The Writer's Options* and *Literature Options for Reading Writing*. He has been reader and table leader for the Educational Testing Service, and he is a consultant for The College Board.

Notes

[1]I would like to thank Ohio Writing Project Directors Mary Fuller and Max Morenberg and Assistant Director Janet Ziegler for their extensive help in conducting the study. I would also like to thank the teachers who permitted me to quote from their responses to the rater questionnaire: Diane Rawlings of Ansonio H.S.; Teresa McGowan of Badin; Thelma Swihart of Eaton; Bob Dizney and Nancy Simmons of Fairfield; Dick Arthur and Carol Schiavone of Harrison; Gary Gilmer of Kings; Jean Gilles and Bob Kiracoff of Lima Shawnee; Debbie Meyer of Live Oak JVS; Winifred Cairns of Loveland; DJ Hammond of Madeira; Sue Sherman of Mt. Healthy; Susan Freyhof of Norwood; Carole Knight of Oak Hills; Lee Ann Adams, formerly of Richmond; Sister Marie Irene Schneider and Sister Mary Dolores Schneider of Seton; Tally Harwood of Talawanda; Phyllis DeMarco and Marilee Pallant of Tippecanoe; Tina Meinke of Tipp City; Brenda Pansing of Tri-Village; Lori Newton of Tri-County North; Linda Balmos of Turpin; and Cindy Briggs of Wyoming.

[2]Of the four analytic categories, readers gave the lowest score to "Supporting Details." Here are the averages on the six-point scale for each category: Ideas = 4.44; Supporting Details = 4.13; Unity and Organization = 4.23; and Style = 4.36. The average holistic score was 4.41.

WORKS CITED

Charney, Davida. "The Validity of Using Holistic Scoring to Evaluate Writing: A Critical Overview." *Research in the Teaching of Writing* 18 (1984): 65-81.

Cooper, Charles R. "Holistic Evaluation of Writing." *Evaluating Writing: Describing, Writing, Judging*. Ed. Charles R. Cooper and Lee Odell. Urbana: NCTE, 1977.

Diederich, Paul. *Measuring Growth in English*. Urbana: NCTE, 1974.

Fitzgerald, F. Scott. *The Great Gatsby*. New York: Scribners, 1925.

Melville, Herman. *Moby-Dick; or, The Whale*. Ed. Harrison Hayford and Hershel Parker. New York: Norton, 1967.

Odell, Lee. "Defining and Assessing Competence in Writing." *The Nature and Measure of Competency in English*. Ed. Charles R. Cooper. Urbana: NCTE, 1981.

White, Edward. *Teaching and Assessing Writing*. San Francisco: Jossey-Bass, 1985.