

THE EFFECTS OF PRAISEWORTHY GRADING ON STUDENTS AND TEACHERS

SAM DRAGGA

Praiseworthy grading is a method of evaluating students' writing in which the instructor comments only on the praiseworthy characteristics of a given essay (Dragga). In addition, following receipt of the graded essay, each student is given a listing of appropriately individualized and correctly exemplified writing objectives or recommendations—things to focus on while revising the current essay or writing a subsequent essay. Guidance on the *writing process* is thus disassociated from the instructor's evaluation of the *written product*. And the risk of "appropriation of the [student's] text by the teacher" (Somers, 150) is minimized.

Praiseworthy grading is, however, quite different from traditional grading. Traditional essay grading requires the teacher to examine the text for its various strengths and weaknesses, to identify these strengths and weaknesses for the student through marginal comments and textual markings, and to summarize these strengths and weaknesses in a final comment and a letter or numerical grade. This is the ideal version of traditional grading typically prescribed to writing teachers (e.g., Najimy). However, as Lil Brannon and C. H. Knoblauch explain:

The teacher-reader assumes, often correctly, that student writers have not yet earned the authority that ordinarily compels readers to listen seriously to what writers have to say. Indeed, teachers view themselves as the authorities, intellectually maturer, rhetorically more experienced, technically more expert than their apprentice writers (158).

It is difficult for us as writing teachers to view the strengths of our students' writing and, as a consequence, we correct far more often than we commend. In doing so, we also obscure a crucial distinction between editors of writing and teachers of writing: i.e., editors focus on the written product and teachers on the writing process. While the editor's objective is to improve a given piece of writing, the teacher's objective is to improve a given writer's writing process, thereby enabling that writer to compose improved written products. Similar to editors, however, we sincerely believe that applauding success motivates the writer, but identifying failure instructs the writer and improves his or her writing.

Empirical examination of praiseworthy grading is therefore essential if it is to receive widespread adoption by writing teachers. This investigation assesses praise of the product and corrective diagnosis of the process (praiseworthy grading) versus praise and correction/condemnation of product and process (traditional grading). In addition, this inquiry examines the effects of praiseworthy grading on writing teachers as well as their students.

METHOD

I conducted my investigation during the Spring, 1985 semester at Texas A&M University, randomly selecting eight writing teachers and randomly assigning four to the experimental group and four to the control group. I demonstrated the method of praiseworthy grading for the teachers of the experimental sections and instructed them to evaluate all their students' writing in this manner during the course of the semester. Teachers in the control sections employed traditional grading on all assignments.

To neutralize the variable of teaching experience, I chose these eight instructors from a group of new graduate teaching assistants, all with a single semester of college-level teaching experience; in addition, all were teaching English 104 (the second course in a two-semester sequence of introductory composition) for the first time, all using the same textbooks and following a standard syllabus of five referential and persuasive essay assignments.

Unknown to the instructors and their students, I randomly selected one of the two sections each instructor was teaching, and from that section randomly selected ten students (i.e., forty students from the praiseworthy grading sections and forty students from

the traditional grading sections.) Though this was sufficient to assure the "equality" of the eighty students (Campbell and Stanley, 25), their verbal scores on the Scholastic Aptitude Test (SAT) verified the equalizing effect of randomization, indicating no significant difference between the forty praiseworthy grading students (mean verbal score = 462) and the forty traditional grading students (mean verbal score = 486).

From all students in all the praiseworthy and traditional grading sections, I collected one typewritten sample of out-of-class writing which had been graded by the student's instructor (a persuasive essay, the fourth of five assignments, written during the tenth week of the semester). I then located those essays written by the eighty randomly selected students. Examination of these eighty graded essays would indicate the level of praising versus corrective comments on students' essays in the praiseworthy and traditional grading sections.

In the same manner, I obtained two typewritten samples of out-of-class writing, prior to grading by the student's instructor, from each of these same eighty students for subjection to holistic grading by independent raters (a referential essay and a persuasive essay, the second and fourth of five assignments, written during the fourth and tenth weeks of the semester, respectively). Evaluation of these samples would indicate the effect of praiseworthy grading on the caliber of the students' writing.

In addition, instructors of the experimental sections solicited anonymous written opinions of their students on praiseworthy grading. And I interviewed the instructors to determine their impressions.

I considered this combination of quantitative and qualitative evaluations sufficient to demonstrate praiseworthy grading as a valid alternative to traditional grading. The hypotheses of my investigation were the following:

- 1) Teachers in the traditional grading sections would be considerably more error-oriented than teachers in the praiseworthy grading sections in marking and commenting on students' essays.

- 2) Essays of students in the praiseworthy grading sections would be judged superior to those of students in the traditional grading sections.

- 3) Students would prefer praiseworthy grading to traditional grading.

RESULTS

Level of Praising versus Corrective Commentary

One sample of writing from each of the eighty students was examined after instructors had graded the essays to determine the level of praising versus corrective comments. For the traditional grading sections, this examination involved counting every marginal comment, every end comment, and every textual marking. For the praiseworthy grading sections, this examination involved counting every marginal comment, every end comment, every textual marking (all praising comments) and every writing recommendation listed on a separate sheet of paper (the praiseworthy equivalent of corrective commentary).

This examination showed that the four instructors of the traditional grading sections wrote a total of 864 comments on these forty essays, 813 corrective comments or 94%, and 51 praising comments or 6%. This averages to 20.325 corrective comments and 1.275 praising comments per essay. Individual instructors ranged from a low of no praising comments to a high of 23% praising comments. Instructors in the praiseworthy grading sections wrote a total of 414 comments on the 40 essays, 160 corrective recommendations or 39%, and 254 praising comments or 61%. This averages to 4 corrective recommendations and 6.35 praising comments per essay. Individual instructors ranged from a low of 43% praising comments to a high of 69% praising comments. (See Table 1.)

It is important to note that while instructors of the praiseworthy grading sections wrote fewer total comments, their comments were generally longer: except for underlining of praiseworthy passages, coded marking of the text did not occur.

CALIBER OF WRITING

Two samples of writing from each of forty students in the praiseworthy grading sections and forty students in the traditional grading sections were holistically scored on a traditional grading scale of 1-100 by four independent raters. If two raters disagreed in their evaluations of a given essay by more than ten points, the essay was submitted to a third rater. (A third rater was required for 23 of the 160 essays, or 14% of the cases.) The two closest

scores were averaged to compute the holistic evaluation of each essay. Inter-rater reliability was determined according to Pearson Product Moment Correlation of the 160 essays; inter-rater reliability was .69, a score which Diederich describes as “typical” (33), especially given the sensitivity of the holistic grading scale.¹

One-way analysis of variance shows that for both samples of writing, essays of the praiseworthy grading sections were judged slightly superior to those of the traditional grading sections. On the first writing sample, the mean score was 71.57 for the traditionally graded essays and 75.26 for the praiseworthy graded essays—a statistically significant difference. On the second writing sample, the mean score was 68.96 for the traditionally graded essays and 71.75 for the praiseworthy graded essays—also a statistically significant difference. (See Table 2.) Because the discourse aims of each student’s two writing samples differ (i.e., referential versus persuasive), interpretation of the scores as pre-test and post-test indicators is invalid; the evaluation of two writing samples for each student is designed only to verify the effect of praiseworthy grading on the quality of the students’ writing.²

STUDENTS’ ATTITUDES

Anonymous written surveys of students in the experimental sections reveal a favorable attitude toward praiseworthy grading, especially among better writers. A number of students, however, expressed a preference for traditional grading. The following comments exemplify the students’ attitudes:

The praiseworthy grading system is better than other methods because it doesn’t leave the students feeling stupid if they don’t receive an A. It builds up confidence for the next paper, but it still lets us know what we need to work on.

I find that it is easier and much more fun and exciting to work in “praiseworthy” rather than critical conditions. This method encourages me to look for my mistakes, rather than just read and forget what my teacher counted me wrong on.

This system builds confidence and encourages the student to try new approaches and make his writing more interesting. Other systems tear down confidence and can often leave the student very discouraged and afraid to try new approaches.

I was sick of receiving papers with red ink splattered everywhere. (Even if it was a B+ paper, there were always those red scratches!). The praiseworthy grading system makes me feel good about my papers. This, in turn, makes me want to try harder—to improve my positive qualities, rather than dread the negative ones.

I don't feel that the positive way of grading helps the student. I myself like knowing exactly what I did incorrectly so that I can correct errors. With this new method, I only know what I did right. The best way to grade would be to join the positive and negative together.

TEACHERS' ATTITUDES

In personal interviews, all four instructors of the experimental sections noted their difficulty in labeling and explaining the desirable characteristics of their students' writing. This, coupled with the analysis of errors necessary for the listing of writing recommendations, initially slowed their grading process.

DISCUSSION

Comparison of the levels of praising versus corrective commentary on students' essays between the experimental and control sections shows considerably more corrective commentary resulting from traditional grading. While traditional grading theoretically focuses on both strengths and weaknesses in students' essays, it practically ignores success and emphasizes failure, as though only the identification and correction of errors possessed instructional merit. Praiseworthy grading, though designed to emphasize success, comes closer to a balance between praising and corrective commentary than does traditional grading: i.e., while traditional grading averaged 94% corrective commentary and 6% praising commentary, praiseworthy grading averaged 39% corrective commentary and 61% praising commentary. In addition, the corrective commentary of praiseworthy grading is more narrowly focused and therefore probably more instructive than the dispersed corrective commentary of traditional grading: praiseworthy grading averages 4 corrective comments per essay, while traditional grading averages 20 comments per essay.

This instructive merit of praiseworthy grading is demonstrated by the superior quality of the essays written in the experimental sections. With both samples of writing obtained, the essays of the praiseworthy grading sections were judged better than the essays of the traditional grading sections. In addition, students themselves attested to the merits of praiseworthy grading. Asked to compare this method with traditional grading, students indicated they felt free to be creative in composing their essays and felt better about their writing since grading was based on things done well. Skilled writers especially believed the teacher provided them with much more individualized treatment with praiseworthy grading.

Less skilled writers, however, were likelier to object to praiseworthy grading. Conditioned by their previous twelve years of schooling to the traditional method of grading and possibly uncertain of their writing skills, these students found praiseworthy grading a worrisome departure from the "normal" method of grading: the change required them to make a difficult adjustment in their ideas about writing and to take responsibility for the editing and revision of their essays. Such students continued to ask the teacher to mark every mistake for them, relying on the teacher's identification of errors to determine the revisions they would make, seeing the eradication of these errors as the single way to improve their writing. Ironically, these students continued to believe in the traditional system of grading even though it had previously failed to help them improve their writing.

In addition, teachers of the praiseworthy grading sections claimed the experimental method required them to make difficult, but ultimately rewarding adjustments in their grading process. Trained to examine the writing of their students for its various failures, these instructors relied on an extensive vocabulary and coding system for commenting on the errors in a piece of writing: e.g., logical errors (logic), mistakes in subject-verb agreement (agr), omission of a necessary transition (trans), or peculiar sentences (awk). Asked to do praiseworthy grading, the instructors quickly discovered themselves without similar explicit criteria for identifying what was good in a student's essay: their training simply never focused on this. The instructors realized it was rarely enough just to write "good" in the margin: students often failed to understand why a particular passage was good, and the teachers had to struggle with putting into words their implicit and intuitive appreciation of

the students' writing: e.g., "I like your repetition of the word *beautiful* in this paragraph. You use it so often that by the end of the paragraph it has taken on the sarcastic meaning that I think you intend it to have. Good job!" Such comments required more thought and more time than the easy identification and circling of errors. However, because praiseworthy grading focused the teachers' and the students' attention on what worked in a piece of writing, it provided a better understanding of what good writing must *include* as opposed to what it must simply *exclude*.

CONCLUSIONS

Quantitative and qualitative evaluations of praiseworthy grading thus indicate that it is superior to traditional essay grading. With praiseworthy grading, teachers give balanced and focused commentary on the strength and weaknesses in students' writing.

Good writers especially seem to prefer praiseworthy grading to traditional grading, gaining more individual attention from the teacher and more commentary on their writing than traditional essay grading typically allows. Poorer writers, however, might still prefer traditional essay grading, possibly because it is easier for these students if the writing teacher acts as their editor and dictates the scope of their revising process. Such students might benefit from additional guidance on improving their editing and revision strategies.

Similarly, teachers in the experimental sections believed that traditional essay grading, with its extensive vocabulary for errors and its numerous correction symbols, was probably easier and therefore more seductive than praiseworthy grading: unless the grading method itself requires identification of the desirable characteristics of the student's writing, the teacher is likely to emphasize errors because the traditional language of grading is the language of editing. The absence of vocabulary with which to discuss the praiseworthy characteristics of a student's writing thus itself contributes to the teacher's corrective orientation to writing and the propensity to serve as the student's editor. Without appropriate words to verbalize the rhetorical efficacy of a given passage or stylistic trait, the teacher is disposed to ignore the merits of the student's writing and to perceive only the characteristics easiest to explain—the errors.

Emphasis in teacher training on explicit criteria for praiseworthy writing, as a consequence, is essential. (Required reading, for example, might be Coles and Vopat's unique *What Makes Writing Good*, offering forty-eight superior student essays, each chosen by a prominent writing teacher and coupled with his or her extensive evaluation of its merits.) The absence of a coding system for the praiseworthy characteristics of writing, however, seems clearly beneficial, requiring teachers to limit and focus their commentary instead of bewildering their students with excessive textual markings.

I am grateful to Joanna Barnett, Karen Birkner, Nelda Bravo, Joann Cain, Clay Daniel, Susan Dollar, Phyllis Miller, Greg Moses, and Terena Kleckner of the Department of English of Texas A&M University and to their students for their participation in this research. I also appreciate the diligence of Larry Beason, Kim Brewer, Eleanor Mader, and Pat Repka in their holistic grading of essays. I am indebted to Paul Meyer and Roger Pace for their guidance in the statistical analysis, and to Gwendolyn Gong, Director of Freshman English Studies, for her cooperation and encouragement.

Sam Dragga is Assistant Professor of English at Texas A&M University and Coordinator of the Writing Specialization, a writing-across-the-curriculum program.

NOTES

¹See White, *Teaching and Assessing Writing*: "When reliability is reported in terms of the degree of difference in scoring by several raters of the same student paper, it is well to be suspicious of almost absolute agreement. One way to ensure agreement, for example, is to use a scoring scale with very few points on it; a two point pass/fail scale, or even a four point scale, will induce much more agreement than a six-point or a fifteen-point scale. . . . Scrupulous researchers may well trade the very high reliability they know they can achieve on such a compressed scale for the more refined distinctions that emerge from a scale that offers raters more decision points" (178).

²See White: "[W]e should attempt to measure anything we call 'writing ability' by more than one writing sample and in more than one writing mode. . . . If financial or time limitations make it impossible to test in more than one mode, we ought to be particularly alert to the validity problem contained in a single-mode test" (118).

WORKS CITED

- Brannon, Lil and C.H. Knoblauch. "On Students' Rights to Their Own Texts: A Model of Teacher Response." *College Composition and Communication*, 33 (1982): 157-166.
- Campbell, Donald T. and Julian C. Stanley. *Experimental and Quasi-Experimental Designs for Research*. Chicago: Rand McNally, 1966.

- Coles, William E., Jr., and James Vopat. *What Makes Writing Good: A Multiperspective*. Lexington, MA: D.C. Heath, 1985.
- Diederich, Paul. *Measuring Growth in English*. Urbana, IL: NCTE, 1974.
- Dragga, Sam. "Praiseworthy Grading." *Journal of Teaching Writing*, 4 (1985): 264-268.
- Najimy, Norman C., ed. *Measure for Measure: A Guidebook for Evaluating Students' Expository Writing*. Urbana, IL: NCTE, 1981.
- Somers, Nancy. "Responding to Student Writing." *College Composition and Communication*, 33 (1982): 148-156.
- White, Edward M. *Teaching and Assessing Writing*. San Francisco: Jossey-Bass, 1985.

Table 1
Praising versus Corrective Commentary on Students' Essays

	<u>Traditional Grading</u> (n ₁ = 40)			<u>Praiseworthy Grading</u> (n ₂ = 40)		
	<u>number</u>	<u>percent</u>	<u>mean</u>	<u>number</u>	<u>percent</u>	<u>mean</u>
Praising Comments	51	6%	1.28	254	69%	6.35
Corrective Comments	813	94%	20.33	160	31%	4.00
Total Comments	864	100%	21.61	414	100%	10.35

Table 2
Holistic Evaluations of Students' Essays

	<u>Traditional</u> <u>Grading</u> (n ₁ = 40)	<u>Praiseworthy</u> <u>Grading</u> (n ₂ = 40)	<u>F</u>
	<u>mean</u>	<u>mean</u>	
1st sample	71.57	75.26	7.368*
2nd sample	68.96	71.75	4.195**

*p < .01
**p < .05