

Kathryn R. Fitzgerald

COMPUTERIZED SCORING? A QUESTION OF THEORY AND PRACTICE

ABSTRACT: *This article examines the implications of computerized scoring of placement essays for the theory and practice of writing assessment, particularly for the complexly interrelated issues of economics, universality, and validity; and then considers its broader implications for the theory and practice of writing instruction. It argues that the very feature that would make computerized scoring inexpensive—its generalizability for widespread use—undermines its validity. The ultimate criterion for measuring the validity of any placement instrument is whether the instrument matches the specific local conditions, in this case, the purposes and content of the courses in which the students begin writing instruction. A generalized scoring system by definition cannot meet that criterion. More importantly, a computerized scoring system undermines the hard-won though still insecure recognition that writing is a communicative act inseparable from audiences, purposes, and contexts.*

If, as Emil Roy suggested in his article "Computerized Scoring Placement Exams: A Validation" (*Journal of Basic Writing*, Fall 1993), computer programs can accurately and inexpensively assess samples of student writing for placement purposes, it appears that computer technology might reform holistic scoring, frequently criticized and often rejected for being too costly and too subjective. Terming the computer a "new and authoritative tool," Roy seeks to appropriate its efficiency to

*Kathryn R. Fitzgerald coordinates the placement program for the University of Utah's Writing Program. She is also past coordinator of Utah's basic writing program and currently coordinates standard freshman writing. Coauthor with Jamie McBeth of the basic writing textbook, *The Student Writer* (HarperCollins, 1991), she has also published and conducted workshops on holistic placement techniques.*

© *Journal of Basic Writing*, Vol. 13, No. 2, 1994

reduce expenses and its authority to legitimate writing evaluation processes whose reliability still remains questionable. I wish to be clear from the outset that the objections to these assertions that I will raise are a particular response to Roy's arguments insofar as they are based on his use of data supplied from the placement files of University of Utah's Writing Program. But more important, Roy's article provides an excellent opportunity to reconsider the complex practical and theoretical issues currently surrounding writing assessment. Before embracing computerized scoring technology, I wish to examine its implications for the theory and practice of writing assessment, particularly for the complexly interrelated issues of economics, generalizability, and validity, and then consider its broader implications for the theory and practice of writing instruction. My concern is that this "new and authoritative tool" will authorize views of writing that run counter to composition's most generative theories and its theory-based pedagogical practices.

Such results certainly are not Roy's intention. On the contrary, his research is focused on demonstrating the validity of a computerized scoring instrument according to current theory. It is important to note the entirely proper shift from a focus on reliability to validity at this point. In this connection, Brian Huot (1990) points out that the researchers who developed holistic evaluation procedures simply assumed the validity of direct writing assessment. However, from the beginning they were forced to focus on challenges to the reliability (replicability or consistency in scoring from rater to rater and session to session) of holistic scoring. Huot argues that it is now time to turn to the question of validity in writing assessment—whether the instrument measures what it claims to. The proposal for a computerized scoring system propels this issue front and center. By its construction reliable, a computerized system is more vulnerable to challenges on the grounds of validity. Roy summarizes four types of validity to be met: *predictive, concurrent, face, and construct validity* (43).

Ed White (1985) points out that testing theory typically calls for establishing the validity of a new test by demonstrating *concurrent validity*, that is, by demonstrating that the new test will produce the same results as an existing test whose validity in other regards has, presumably, already been demonstrated (185). When he began his project, Roy contacted the University of Utah among many other institutions to request copies of holistic placement criteria and exemplary placement essays.

Concurrent validity would be achieved by coordinating quantitative measures of selected stylistic features with the results obtained by means of human readings based on apparently sound holistic criteria. Upon Roy's decision to use Utah's criteria as the standard for validating his scoring program, I provided another set of 46 randomly chosen placement essays written in the summer of 1990 and the writers' grades in writing courses and ACT scores when available.

The stylistic features to be quantified for Roy's Structured Decision System (SDS) were selected on the basis of the findings of prior research into the correlations between verbal features and essay quality and of his stylistic analysis of the four anchor essays provided by the University Writing Program (UWP). Features initially identified as salient to placement included essay length, "demonstrat[ing] development within paragraphs, structural completeness, and scribal fluency" (42); high syllable average per word, indicating a "mature lexicon" (42-43); and low percentages of unique words, a negative correlation based on the reasoning that competent writers use repetition and reuse transitional words to achieve coherence (42-43). As he expected, Roy found that the majority of computerized ratings by these three stylistic features (65%) agreed with the essay's holistic ranking (45). He sought to adjudicate the remaining discrepancies by triangulating the computer's placement level with grades in the writing course when available and with ACT scores, adjusting cutoff points in the quantitative measures or adding new categories when they seemed warranted. Following standard principles for concurrent validation of a new testing instrument, Roy searched for and in most cases located quantitative features that could be adjusted to align his SDS rankings with holistic ratings. These procedures resulted in aligning the SDS ranking with the holistic scoring of nine of the original disagreements, determining five more to be misplaced by holistic reading, and leaving two unresolvable by any combination of available information (48). The bottom line is, Roy claims, that the computerized scoring has the potential of producing "an accuracy rate of 95.66%" (48), surpassing holistic ranking, which produced an accuracy rate of only 85% (48).

Roy's procedures, based on current testing theory, seem to be implemented responsibly. I have no quarrel with his procedures, but I am concerned about implications for composition theory and practice of this or any similarly conceived program. One implication emerges from the economics of marketing such

a program. The principle that would allow a computerized program to be produced inexpensively is that it could be generalized for widespread use. And fundamental to broad general use is the premise that a universal ideal of good writing exists which varies only as a function of cognitive development. However, theorists informed by post-structuralist notions of linguistic difference, especially as located in the speech and literacy acts of discourse communities, no longer accept the possibility of a single standard of successful written or oral performance.

That Roy, unfortunately, accepts that possibility is evident in his reference to White's regret that "our profession has no agreed-upon standard of proficiency, and certainly as a consequence, no agreed-upon definitions for proficiencies at various levels of schooling" (42). But White, with many others, now advocates the acknowledgement of discursive difference (White, 1990), while Roy's desire to "limit the ambiguities of holistic grading as applied to impromptu placement exams" (42) reveals his adherence to the expectation of such a standard.

The assumption of a universal standard for good writing becomes an issue when attempting to establish *construct validity*. Roy posits an ideal that has become traditional in composition studies since the early eighties, the ideal of the "professional writer." He asks of his model, "Does the SDS measure the essential skills and abilities that comprise the writing competence of professional writers, establishing its *construct validity*?" Huot defines *construct validity* by stating that it ensures "the theoretical soundness of an assessment procedure" (206) or "the extent to which the test may be said to measure a theoretical construct or trait" (206). Ed White explains a *construct* as "an unobservable trait that is hypothesized to understand and account for observable behavior" (1985, 188). Thus, in the case of writing assessment, something called "writing ability" or "essential skills and abilities" in Roy's terms are hypothesized to account for writing performance.

But to propose the performance of professional writers as the ideal by which other performances are measured is again to posit a single, universal standard for writing competence. This standard fails to acknowledge differences in the expectations for various purposes, contexts, and genres within professional writing, to say nothing of differences between professional writing and good student writing. Clearly, criteria for judging performance by professional writers varies with contextual factors: the standards by which an advertising copywriter's texts would

be judged vary from those by which a legal clerk's would be judged and those by which a software manual writer's, a journalist's, or a novelist's would be judged. The underlying skills and competences explaining excellent performance in these areas would as a consequence also vary.

Not only are there crucial differences among criteria for good writing in the multiple genres of professional writing, there are equally crucial distinctions between professional and student writing. To achieve construct validity for assessment of college-level writing competence, the *construct* must be a description of rhetorical and stylistic expectations for good student writing, not for some outside genre. It is essential to note that school writing is itself a unique rhetorical situation in which writers write to demonstrate their content knowledge, reasoning ability, and rhetorical competence to readers whose competence in all three areas, as well as authority to evaluate, surpasses the writer's. Such a situation requires quite different, and in many ways more complex, competences than professional writing (Fitzgerald, 1988).

Here I am directly raising the deferred question of the validity of applying quantitative measures to writing assessment. By addressing the rhetorical situation of the entering student writer in the ways that they do, the University of Utah's placement criteria confound the results obtained by purely quantitative measures of stylistic features. Of course, any attempt to describe "the" rhetorical expectations for student writing will be met with the same critique I have made of Roy's criteria—that student writing genres, bounded by discipline-specific purposes and expectations, call for a variety of competences which cannot be summed up in a single set of rating criteria. Granting that view, yet required to develop a placement instrument within certain institutional constraints, we at Utah have attempted to distill rhetorical as well as stylistic features for which construct validity in the entering student's rhetorical situation might be claimed.

Ten years of administering the placement essay to over 30,000 students at Utah have resulted in our recognition of the significance of evidence in the writing of two specific cognitive moves to accurate placement of students within the rhetorical context of the University of Utah's Writing Program. They are 1) evidence of an inductive move from specific to general, and 2) evidence of viewing an issue from more than one perspective. The inductive move represents the fundamental logic of em-

pirical investigation that is the basis, at least ideally, of the scientific method undergirding a major portion of university research. In its inverse—thesis followed by evidential support—it is still the predominant format for student writing. All of Utah's placement essay prompts invite students to demonstrate their written competence in managing the inductive move. The essays Roy used responded to a prompt asking students to describe a disturbing situation, explain desired changes, and draw conclusions about how people can or should respond to such situations. Here are two other similar prompts:

Describe an experience that was educational for you and tell what you learned from it. Then go on to consider education in more general terms by discussing what your experience indicated about how people learn.

We all experience peer pressure at one time or another in our lives. We usually think of peer pressure in negative terms, but it can actually have either positive or negative effects. Write an essay in which you first describe a time when you experienced peer pressure. Then, on the basis of your experience, discuss how people respond to peer pressure, and what we can learn about ourselves, about others, and/or about how to make choices from experiences with peer pressure.

Like all of our prompts, these ask students to provide details about a personal experience and then to draw logical conclusions from their experience—mimicking the inductive move. Because we view the ability to make the inductive move a better predictor of writing performance than background knowledge of any particular subject matter, the prompt always asks students to draw upon specifics from personal experience.

The second characteristic that determines accurate placement is evidence of the capacity to view an issue from more than one perspective. As a developmental issue, this capacity has been viewed as an indication of maturity, and admittedly, that rationale originally supported its inclusion in Utah's criteria. However, recent interrogation of developmental universals as cultural and/or ideological constructs has opened an alternative space for construing the question of multiple perspectives as a rhetorical issue. It is relevant to student writing because it reflects the culturally constructed rhetorical ethos of the academic, which requires him or her to hear and consider all reasoned views. In other words, consideration of multiple per-

spectives "works" as a placement criterion because it is a feature of the expectations of the particular rhetorical context constituted by the University.

Not convinced that incorporating language expressing this expectation into the prompt itself would be helpful to students, we point it out in the evaluation criteria they read before writing:

The essay readers will be reading to see:

. . . . How you make your answer relevant to the broader concerns and perspectives that the essay question suggests.

Attention to evidence of the two features of an inductive move and multiple perspectives undermines paper length as a placement criterion, though paper length is viewed as "the most reliable measure of the quality of impromptu writing exams" by several researchers and by Roy (42). Previous researchers have noted that paper length is an indirect result of students' ability to generate detail and the explicit structural language of cohesion. Indeed, students able to use vivid detail, to consider complex interrelationships between details and possible generalizations, and additionally to consider experience from a variety of perspectives will often write longer papers. But the crucial point is that length does not signal whether the student's text demonstrates the inductive move and/or multiple perspectives. And this discrepancy is where a human reader's discrimination is crucial.

Some students, especially those whose narration of an experience is lengthy, may generate long essays without making any inductive move, while others may demonstrate an ability to make efficient logical connections among several layers of abstraction in relatively short essays. Students whose writing remains at a single level of abstraction are almost invariably placed into one of two levels of preparatory writing no matter how long their essay. Those whose writing moves among levels of abstraction, even if the logical connections are not tight, are usually placed in Utah's standard freshman course, while students who, to some extent, manage both valid logical connections between specifics and generalizations and multiple perspectives on the topic are placed in the advanced freshman course.

I have selected the following essays to illustrate this point from a file of sample essays used for norming essay raters. Both

were timed forty-five-minute essays written to what is probably our most successful prompt, the educational experience question quoted above. The first example, 489 words, demonstrates a typical level 2 placement (the second quarter of preparatory writing). The second, placed at level 4 (advanced), is 430 words long. I quote the level 2 essay in its entirety because one cannot omit material in demonstrating a lack.¹

Level 2 Essay (placement in second quarter preparatory writing)

Recently I had an opportunity to visit Europe. While we were there we went into East Berlin. I have never had such a dramatic learning experience as the one I received in here. People sure do take the freedom we have in America for granted, and so did I. After going into East Berlin and witnessing the lifestyle, I sure became very honored that I was able to live in America. We sure are blessed with the freedom we have, and we should never take that for granted.

As we rode the train into East Berlin, we were greeted with dark, unfriendly looks. I even felt as if the sky seemed to get grayer. We hardly saw any cars, but the ones we did see were very small with the color either black or dark gray. The roads we saw were made out of cobblestones. Many of the stones were missing making the road seem very bumpy and rough. People over there seem so unhappy. I never saw anyone smile or laugh. I sure am not saying that they never do smile or laugh, but they don't seem to as much as you or I. And of course, that is quite explainable.

My mother has been writing a pen pal who lives over in East Berlin. That was one of the major reasons for going there. We wanted to meet her and also talk with her. It was very hard to find and communicate with her. In fact, we actually never did talk directly to her. We always had to relay messages through other people. Finally we arranged a time and a place to meet her. The exciting moment came where we were able to meet face to face. We had so much to talk about. While we were conversing back and forth, some policemen came and broke us up. We were not allowed to stand and speak like that. So we had to walk and talk very quietly. She told us she knew she was not free, but she knew that America was. Also we learned that at the age of sixty, you have the

choice whether to leave the country or not. But at that age, the people usually don't have the money to do so, and besides, who wants to leave their family. Well, the time had come for us to leave. Tears fell from everyone's eyes. We wanted to take her back with us, and I knew she wanted to come.

I think when one goes into a communist country and experiences their lifestyle, one becomes so much more appreciative for the word freedom. I know I did. I will never forget this great learning experience I was able to receive. I wish that everyone who takes freedom for granted could experience what I did. Freedom is much more than a word, it is the best place to live. We should all be grateful that we live in America.

Much could be said about this essay, but what is pertinent to this discussion is its lack of an inductive move and its reliance on a single perspective. On an imaginary continuum from the very specific to the very general, this essay is written in a very narrow range slightly to the specific end of center. It recounts an experience in some detail, but until the final paragraph it does not stray from a single level of specificity. Moreover, the attempt at generalization in the final paragraph does not develop out of the material in the essay—it does not elaborate what the experience demonstrated about how people learn. Instead, the essay falls back on a commonplace construction of cold-war experience. No explanation appears to link observed incidents to generalizations: the observed unhappiness was “of course” easily understood.

Except for the caveat, “I . . . am not saying that they never do smile or laugh,” the writing remains within a single point of view. The text does not attempt to account for experience in Berlin in any terms other than those of cold-war rhetoric. Though the topic is now dated, this essay is still an excellent example of the logical and rhetorical limitations of a typical level 2 essay. Length is irrelevant: the salient feature for placement is whether the essay moves between specific detail and more generalized statement and among two or more perspectives. (This is not to say that other stylistic features are irrelevant; syntactical complexity and lexical sophistication are components of our rating criteria.)

The second essay is also about a trip. I must again quote the entire essay, this time to demonstrate the presence of significant logical and rhetorical moves.

Level 4 Essay (placement in advanced freshman writing)

I spent my summer break last year working on the floor of the New York Stock Exchange. I had always been fascinated by the interactions and reactions of the buzz and selling of stocks and when I learned that I could participate in this distillation of the American culture, I was exuberant. It taught me not only how bids and offers affect the stock prices, but also how people interact. The floor was not at all how I expected it. Granted I knew that people were screaming at all hours of the working day, but it was what they were screaming that had a profound impact on me. I had expected people to be at least a little courteous when it came to making a deal, but instead, the traders were exactly the opposite. They could be standing, acting completely friendly toward each other and suddenly when trading began they would start calling each other (and their co-workers in the booths) every name in the book (and quite a few that can't even be printed in *that* book). I learned through this experience that in our society, at the heart of capitalism, every man is for himself. There is no second prize, and compassion is only for those who wish to lose (or can put it in an ad campaign). This in and of itself may not be any great revelation to a thinking person, however this experience wiped any and all of the haze that people use to tarnish that stark reality.

My experience was fairly unique (especially for a sixteen year-old), however the way I learned from my experience, I feel, is universal. It was not as if someone pulled me aside and said "Can you see that? It means each and every man for himself." I learned it by absorption, bit by bit, day by day I picked up more of a feel for what it was really like on the floor. I would assume that this is the way the traders learned to be cutthroat, they little by little found that if they were nice, some other guy bought his family a new apartment while he struggled to keep his that he has been in for the last five years. This process of rationalizing what one observes may well be the oldest form of learning, dating back before communication was even possible (of course I can't say for sure, I've never read anything on precommunicative thought processes). I think that people learn by rationalizing what

they observe, the same method by which I made my "amazing discovery."

Clearly, this student's available repertoire of syntactical and lexical alternatives is much greater than the previous student's, and clearly these features, which are quantifiable, will affect the readers' judgment. But these were not the primary criteria for placing this essay in advanced writing. Its exceptional logical and rhetorical sophistication were.

The writer's use of the inductive move is impressive. The essay elaborates the logical connection between the experience and *how* the learning occurred on at least two levels. First, it generalizes the specific process of learning to apply to human behavior at large: "My experience was fairly unique (especially for a sixteen year-old), however the way I learned from my experience, I feel, is universal." In detail, the text traces the writer's own rational experience and then, incorporating parallel syntax to reinforce the observed similarity between self and others, generalizes to the limited group of other traders. From that incremental hop, the logic leaps to humankind as a whole, including "humans" who could not communicate verbally. But apparently aware of the danger of the broad generalization's sliding into the ridiculous, the writer steps back by acknowledging other possible views—"of course, I can't say for sure, I've never read anything on precommunicative thought processes."

At a second level, the learning process the writer describes, though termed only "rational," is indeed the inductive process. The text explicates it in some detail: "It was not as if someone pulled me aside and said 'Can you see that? It means each and every man for himself.'" The writer did not learn by direct instruction. Instead, "I learned it by absorption, bit by bit, day by day I picked up more of a feel for what it was really like on the floor." The writer learned by the accumulation of observed detail until it could be generalized to a pattern of behavior, in other words, by inductive logic.

As impressive as is this essay's use of inductive logic are its multiple perspectives. On the lighter side he gives us, "they would start calling each other (and their co-workers in the booths) every name in the book (and quite a few that can't even be printed in *that* book)" with its humorous nod to more than one scale of "dirty" words. Capitalists, a group from whom the writer is distanced enough to judge, evoke two differentiated views about compassion: it is either for losers or for exploita-

tion. Worth noting is that this sentence about compassion—"There is no second prize, and compassion is only for those who wish to lose (or can put it in an ad campaign)" —is scrawled in the margins of the original test booklet, an afterthought which indicates that the writer kept working on the capitalist's perspective even while composing another sentence voicing alternative positions. The next sentence reflects on past naiveté (without taking ownership of it) and acknowledges more mature perspectives, perhaps that of the placement reader: "This in and of itself may not be any great revelation to a thinking person, however this experience wiped any and all of the haze that people use to tarnish that stark reality." This mature/naive duality reappears in the final sentence of the essay in quotation marks around a phrase no longer the writer's: "I think that people learn by rationalizing what they observe, the same method by which I made my 'amazing discovery.'"

These two essays are quite common in Utah's placement experience. Both were among the first ten I reviewed when looking for examples for this article. Though Roy's sample group did not seem to have any examples of level 2 essays longer than level 4s, several of the sixteen discrepancies between the SDS and holistic ratings were based on word length, including all six that were not borderline cases. Roy attempts to explain these anomalies quantitatively by adding other features to the equation, seemingly on an ad hoc basis. For instance, in the case of three essays that were too long by SDS criteria for their level 2 holistic rating, Roy found low counts of syllables per word, and so, added that feature to the criteria for placement in preparatory writing, reasoning that "short, simple words probably overrode the favorable impression created by paper length" (46). Indeed, short simple words would have affected placement, but probably to verify a judgment made on the basis of the absence of inductive reasoning and multiple perspectives. In fact, far from positively influencing raters schooled to look for evidence of these logical and rhetorical features, long essays that fail to demonstrate such moves tend to try their patience.

In another case where a short essay placed its writer in advanced freshman writing, Roy noted that sentences were unusually long. He consequently added this criterion to level 4 placement, hypothesizing that "sentence lengths averaging \Rightarrow 23.25 words apparently override modest word production" (47). Again, sentence length, or more likely syntactical complexity, probably do influence placement. But more salient

would have been evidence of inductive reasoning and multiple viewpoints. On the basis of other anomalies, Roy also adds to the placement equation percentage of unique words, the Flesch-Kincaid readability level of the essay, and percentage of prepositions.

Making these ad hoc adjustments, Roy is addressing a problem that human readers constantly confront—that few essays perfectly fit the profile of a single placement category, even when described holistically. Human readers must constantly weigh the importance of one feature of a text against another, a process that a computerized system is forced to quantify. This is the necessity that leads to complex and inelegant ad hoc additions to a computerized scoring system.

It might appear, then, that logical and rhetorical features of texts are likely to invalidate any quantitative system feasible in the near future. That is not the argument I wish to make. It may be that if the resources were available for further research, and Roy points out the difficulties of that situation (50), a theoretically sound quantitative scoring system could be devised for which construct validity could be demonstrated. To achieve construct validity, the researcher would need to return to a critical examination of the writing features hypothesized as relevant to expectations for texts in the rhetorical context of a student entering the university. On the basis of Utah's experience, I would predict that such research would result, for starters, in the elimination of length as a criterion but the inclusion of syllable-per-word count, and, if valid measures could be devised, of syntactical complexity.

However, even if such a computerized program were to be devised and found to be less expensive and more accurate than holistic placement, I would still be concerned about using it to the exclusion of human readers. My first concern would have to do with the very feature that would make such a program inexpensive—its general acceptance for widespread use. In this regard, Roy's research provides an opportunity to discuss the theoretical issue of the universality of writing standards in the context of a particular practical application. It is a good example of theory, misguided theory in this case, meeting practice. I noted above my reservations about the claim that it is possible to identify a single set of criteria even for a genre apparently so limited as "student writing." Utah's holistic criteria have been effective for placing students into freshman level writing courses developed in tandem with the placement

essay. The essay prompts and criteria are designed to elicit writing to demonstrate whether students as yet handle competently the features of writing taught at each course level. If a student questions a level 2 placement, for instance, I can point out that the essay failed to demonstrate the ability to generalize from specifics and assure her or him that the level 2 course focuses on a series of writing assignments that will all, in a variety of ways, give students practice in making this inductive move. The ultimate criterion for measuring the validity of any placement instrument is whether the instrument matches the local conditions, specifically in this case, the purposes and content of the courses in which students begin writing instruction. The danger is that cost-conscious administrators would be tempted to adopt a computerized program without regard to the instrument's theoretical and contextual premises, which determine its appropriateness to any local situation. On the other hand, the reliability requirement that forces human readers continually to discuss their rationales for rating also works to ensure content validity. Local concerns will, willy-nilly, figure into the discussions.

There are additional reasons related to the material impact of college testing, well-documented since the initiation of Harvard's composition test for admission over a century ago, to question the use of computer scoring for student writing. Placement procedures that avowedly ignore the content and discursive moves in a text convey an old, product-centered message about writing to a public that matters for writing instruction—high school English teachers, administrators, school boards, college instructors, and college administrators. The message is that writing consists merely of discrete stylistic components that operate independently of communicative contexts, that is, of audiences, purposes, and genres. The implications are frightening. A text's reception by readers could again be ignored, and school boards and administrators would not be required to pay (or, at least, not required to feel guilty for not paying) the costs of small class size for English teachers. Teachers could revert to workbook exercises in vocabulary and complex sentences, saving themselves the immense time commitment to read student papers. The message would condone the reinstatement of Harvard's Subject A and its correctness-oriented descendants.

The question is, finally, of greater import than the validity of a testing instrument. It goes to the core of the identity of practice and theory. As we make decisions about practices—whether

they be the practices of large-scale assessment, of placement, of pedagogy, of textbook selection or authoring, of teacher training—it behooves us to consider how one practice implicates another and how practices determine strategic articulations of theory. Compositionists have fought against the view of writing instruction described above for twenty-five years, especially in regard to basic writing at the college level. The recognition that “writing” is a rhetorical act inseparable from its content, contexts, and purposes has been hard-won. It would be a mistake now to undo these admittedly partial victories in the name of seemingly “new” but, no doubt, “authoritative” technologies.

Note

¹I have corrected some mechanical errors to foreground the rhetorical and logical differences between the two.

Works Cited

- Fitzgerald, Kathryn R. “Rhetorical Implications of School Discourse for Writing Placement.” *Journal of Basic Writing* 7 (1988): 61-72.
- Huot, Brian. “Reliability, Validity, and Holistic Scoring: What We Know and What We Need to Know.” *College Composition and Communication* 41 (1990): 201-13.
- Roy, Emil L. “Computerized Scoring of Placement Exams: A Validation.” *Journal of Basic Writing* 12 (1993): 41-54.
- White, Edward M. *Teaching and Assessing Writing*. San Francisco: Jossey-Bass, 1985.
- . “Language and Reality in Writing Assessment.” *College Composition and Communication* 40.2 (May 1990): 187-200.