Herrington, Anne and Sarah Stanley. "*Criterion*^{SM}: Promoting the Standard." Paper delivered at the Conference on College Composition and Communication, San Francisco, CA, March 2008.

Criterion^{SM}: Promoting the Standard

Anne Herrington and Sarah Stanley

Automated assessment of writing has moved into the classroom. Instead of just being used for some standardized testing programs, including two popular college placement programs (the College Board's ACCUPLACER test using WritePlacer *Plus*, 2009, and ACT's COMPASS test using e-Write, 2009), automated assessment programs are now being sold for instructional purposes. One of the more popular of these is *Criterion*, a product of the Educational Testing Service (ETS), our focus for this essay (1). A web-based program, *Criterion* is marketed as an aid to both teachers and students for the fast computer evaluation it can provide to student writing and to teachers. As the lead sentence on the ETS *Criterion* Online Writing Evaluation homepage proclaims, "The *Criterion* service evaluates a student's writing skills and within seconds provides score reporting and diagnostic feedback to both writing instructors and students" (2008). With many teachers facing the pressures of large teaching loads and larger classes, students with a range of home language and dialects, and accountability demands, it is not surprising that *Criterion* could seem attractive.

To date, while more has been written in Composition Studies about automated programs in general (Ericsson & Haswell, 2006; Herrington & Moran, 2001 and 2006; Hesse, 2005, Williamson, 2009), there have been few independent reviews published on instructional programs designed to teach writing, specifically *Criterion* (Broad,2006; Chen & Cheng, 2008, Herrington & Moran, in press; Hutchinson, 2007). All point to the

limited capability of the program to do more than identify structural features. For

example, on the basis of their study of *Criterion* in three EFL courses for English majors

in a Taiwan university, Chen and Cheng (2008) conclude, "if the goal is to communicate

the writer's thoughts effectively to real audiences and demonstrate the writer's creativity

and originality, using AWE [automated writing evaluation] is probably not a good

choice" (p. 109). While pointing to the theory of writing programmed into *Criterion*, not

even the Chen and Cheng study considers also the ideology of language that drives the

program. Our purpose, then, is to review *Criterion* in this light, asking the key questions

we, as teachers, need to ask of any text book or instructional technology, recognizing it as

promoting particularly ideological orientations (Kemp, 1992; Huot, 2002): what beliefs

about writing, language, and learning is it programmed to follow as it evaluates student

writing? what message does *Criterion* send to teachers and students alike about dialect

and language difference? As Kemp (1992) reminds us, "instructional software is not

merely a set of instructions that drive computers (and sometimes users) but a subtle

platform of belief which can . . . carry implicit messages regarding . . . viewpoints which

remain invisible even to writing professionals trained to recognize ideology in written

texts" (p. 12).

What we will show in our review is that *Criterion* enacts an arhetorical view of

writing, a homogeneous view of English as a single standard dialect, and an error-focused

approach to teaching and learning. In saying this, we acknowledge that *Criterion*

operates on values and practices of some of us as English teachers. In other words, it

mirrors some portion of our profession, but not in a way that prompts self-examination.

Who are we in writing this essay? Two, white teachers of two generations, both of whom grew up in homes where standardized American English was spoken, the dialect of middle and upper class white Americans, the dialect of *Criterion*.   We have some knowledge of linguistics; we strive to enact the 4C's policy statements on language diversity in our teaching; and, to that end, we profit from theoretical and pedagogical scholarship on ways to do so.

Recognizing and Honoring Multiple Varieties of English

The chapters in the collection *Language Diversity in the Classroom: From Intention to Practice* (Smitherman & Villanueva, 2003), challenge us, as Smitherman and Villanueva put it in their introduction "to think about how we enact our belief in the multiplicity of language, of English, in our classrooms" (p. 1). *In our classrooms*: Regardless of where we teach, some multiplicity of languages and varieties of English are used by us and our students. In the 1970's, the Conference on College Composition and Communication's policy statement, "Students' Right to Their Own Language" (1974), called us to recognize and honor as viable the multiple dialects of American English; the same organization's 1988 National Language Policy Statement called us to recognize and honor the variety of languages present among us in our classrooms. On-going scholarship on dialects of English (e.g., Smitherman,1993; Richardson, 2003a) and more recent scholarship on World Englishes (e.g. Canagarajah, 2006) and the myth of linguistic homogeneity in U.S. schools (Matsuda, 2006; Horner & Trimbur, 2002; Lu, Matsuda, & Horner, 2006) make the same call, arguing as well that students' language capabilities in whatever variety of English they speak are resources for learning to write and that a classroom policy of monolingualism ill prepares students for a world of

increasing inter- and intra-mixing of languages. This scholarly work and our policy statements rest on what linguistics and rhetorical analysis demonstrate: all varieties have conventions—and standards—for discourse patterns, style, and grammar. While language scholars continue to assert that *linguistically* speaking, no dialect is any more "standard" than another, it remains the case that one dialect of English remains socially dominant, the dialect of American English taught in our schools and used in institutions-- the dialect most associated with middle and upper class white Americans; what we will refer to as "as standardized American English" (SAE) to mark it as a particular dialect that has become standard*ized*, and thus, dominant. Recognizing the dominance of SAE does not require, however, teaching students to follow it as the *only* language standard. As Ball (1999) has pointed out, though, in reference to African American Vernacular English (AAVE), many educators still fail to recognize it as a viable language:

> Although linguists and anthropologists have assured teachers that AAVE is a logical language with systematic patterns of expressions, many educators have trouble seeing and appreciating these patterns. Instead of patterns, they see only "mistakes"; instead of 'efficacies'—powerful resources that are part of an oral tradition that students can use to produce an effect—they see only "errors." (p. 226)

Criterion enforces the same misconceptions about all varieties of English other than SAE.

Marketing Standardization: Overview of Criterion

The ETS (2008) *Criterion* Online Writing Evaluation homepage markets *Criterion* as designed for instruction in K-12 and college classrooms and for an ethnically diverse group of learners. The page features photos of three learners for each of three categories—a blonde haired, blue-eyed white girl with the banner K-12; a straight haired,

brown-eyed woman of color with the banner Higher Education; and a dark haired, dark-eyed man of color with the banner English Language Learning.   The implication of the images is that *Criterion* is designed to help all learners, including those learning English who are visually represented as racially as well as linguistically different from the assumed White norm. What isn't stated is that *Criterion* recognizes only one dialect of English. In other words, helping all learners becomes, in practice, erasing other languages or dialects.

The homepage for *Criterion* also prominently displays the seals of three technology awards that *Criterion* has received, including the 2007 Platinum Level Award from the IMS Global Learning Consortium for "high-impact use of technology in learning."   Even more persuasive, other pages for each of the three main categories contain testimonials from teachers and administrators who use *Criterion* and praise it for saving time for instructors so they can devote more time to individual students, providing pre- and post-testing and placement capabilities, and supporting students in revising their writing (2). What these awards and testimonials do not speak to are the ideologies of writing and language programmed into *Criterion*.

To examine these more closely, we move now to the section for Higher Education. While we focus on this section, much of the marketing language is the same for other sections as well. According to the "*Criterion* Service Overview" for Higher Education, *Criterion* is "a reliable assessment tool" "allowing instructors to benchmark writing, to make placement decisions, adjust instruction and track progress." Not only is it useful for assessment and benchmarking, "the *Criterion* service is a learning tool that does not increase workload and adds value to writing instruction by providing a teachable

moment." As these two quotes illustrate, ETS draws on the discourses of both accountability (benchmarking, value added) and progressive pedagogy (teachable moment) to market the product. But what sort of teachable moment is it and for whom?

The page *Criterion Details* for Higher Education includes a Frequently Asked Questions section, the lead question and answer being the following

> How can the ***Criterion*** service help students?
>
> Students get a response to their writing while it is fresh in their minds. They find out immediately how their work compares to a standard and what they should do to improve it. . . .

Note that here "a standard" is invoked as the benchmark against which students' work is evaluated. The nature of that standard is not identified here, however. The language of writing pedagogy is evoked as well, "directed writing practice," "writing and revision,"— with the qualification that it is by being "capable and motivated" that one can use, and presumably profit from, *Criterion.*

Another of the Frequently Asked Questions links the question of a "standard" to students and their languages. The question is posed thusly:

> Does the ***Criterion*** service discriminate against students who may be bright but who may not have mastered standard English — for example minorities and ESL students?

At first glance, it seems a strangely worded question: Why insert "who may be bright"? Can you imagine a question on a program for learning Spanish that asks whether the program discriminates against "students who may be bright but who may not have mastered Spanish yet"? Actually, through this wording, an ideology linking dialect with

race, but not class or region, is implicitly acknowledged: that is, "whites" are not included in the "for example," implying that if you are white and bright, you have mastered the "standard": it is your white dialect, implying also only one dialect for white Americans, regardless of class or region. It is "bright" minorities and ESL students, those who may have mastered another dialect or language, who need to "master" the standard.   This distinction between the bright and "not bright," is also at odds with the attitude of "reflective optimism" that Ball and Lardner (2005) call for, a "belief in our students' ability to succeed" (p. 17). With the question worded this way, the answer, of course, is "No." The answer in full reads as follows:

> No. The *Criterion* service is incapable of discriminating on the basis of race, sex, national origin or student's history because these factors do not exist in its analysis. The program simply measures features in a piece of writing and compares them to features in previously scored essays used to define the rubric. If the collection of sample essays includes essays that use non-standard English and also earn high scores, then the *Criterion* service will assign a high score to other essays with the same features.

As Valerie Balester (1993) reminds us, "Judgments of style are to a large extent subjective and culturally conditioned" (p. 88). However, the first sentence of the answer tries to decouple cultural and historical understandings of difference from how they relate to language use. Race, national origin, and histories of power and discrimination exist in the evaluation of essays because they are embedded in the language norms by which those judgments are made. Still, as the answer goes on, don't blame *Criterion*. It is "simply" doing what it is programmed to do, and it is programmed against the norms of

human readers. Those readers, of course, are ones trained by ETS to evaluate standardized tests. For those readers of standardized tests as well as for some classroom teachers, there is only one standard. *Criterion* only perpetuates that view. The response also implies that since the *intent* was not to discriminate, the producers of *Criterion* are not responsible for any racist *effects* of the product. Notice, too, that "standard English" is referred to as a single language variety, set against "non-standard English," implicitly also a single language variety that encompasses, we suppose, all other varieties of English. There is not even a mention that this standard might be different from the standards that some are familiar with and that knowledge of these other standards can be resources for learning standardized academic English. Further, *Criterion* presents an impoverished construction of standardized academic English, focusing as it does on a deficit view of language with limited prompts for writing that do not reflect the range of genres called for in college courses.

As is evident by the section on ELL, *Criterion* does recognize multiplicity of languages and language users and, in general, is better designed for L2 speakers than for speakers of nonmainstream dialects of American English. For example, the online *Writer's Handbook* comes in an ELL version and a Spanish/English bilingual version. It seems that more is done to recognize the validity of multiple languages, than multiple dialects. This difference in valuation is consistent with the finding of the 1996-98 CCCC sponsored study of high school and college language arts professionals that found more acceptance for using languages other than English than for using nonmainstream dialects (Richardson, 2003b, p. 52), implying less acceptance of U.S dialects of English as viable

language varieties and pointing to the complex intermingling of dialect, race, and prejudice in the U.S.

Using Criterion

Presently, *Criterion* is capable of providing holistic evaluation of expository or persuasive essays. To be specific, *e-rater*, the program used for the holistic evaluation, is normed against impromptu essays written for timed ETS tests that use expository and persuasive prompts. As explained in one of the Frequently Asked Questions, the essay prompts come from "ETS testing programs like NAEP[®], the English Placement Test designed for California State University, *Praxis*[®], GRE[®], and TOEFL[®]." In other words, *Criterion* is designed to evaluate test-situation writing, not, for example, a researched argument, an interpretation of a short story or essay, an analysis and response to published text, an oral history interview, a science laboratory report, a website, a digital story, the list goes on. In short: *Criterion* does *not* present anything like the range of rhetorical genres that students would be asked to write in their college courses, nor engage them in the kind of critical thinking valued in college and professional work.

As teachers, we would have the option of selecting one of the topics *Criterion* provides or creating one of our own. In order to use the holistic rating feature, though, our "teacher-designed" prompt would have to be of the same type as the *Criterion* ones. Further, if one wishes to use the holistic feedback rating, *The Criterion Teaching Guide* advises setting a time limit for writing: "Note that it's very important to assign the *same amount of time* for writing the essay from instructor-generated and modified prompts as is required by the standard higher education prompts" (ETS, 2007, p. 28). "The

standard" is evoked again, here equating timed, impromptu writing with the standard for writing assignments in college classrooms.

Once receiving the prompt, students could first use the "Make a Plan" option (e.g., outlining, free-writing). The demonstration video shows that if the timed writing option has been chosen, students are reminded "SPEND NO MORE THAN 10 MINUTES MAKING YOUR PLAN." They would then draft their essays and submit them for evaluation. Here's where the real focus of *Criterion* lies. *Criterion* is designed to provide two kinds of evaluation: Holistic, using the scoring program *e-rater*, and Trait Feedback, using the scoring program *Critique* (Burstein, Chodorow, & Leacock, 2004, p. 27). On the basis of the holistic score evaluation, the draft would receive a holistic score on a scale of 1 to 6, determined by comparing this draft to the bank of essays scored for ETS tests. Figure 1 (on the next page) shows the holistic explanation for a rating of 6. Noteworthy is that the description for a rating of 6 explicitly acknowledges the language standard that is being used: "American English usage." Here, "American English usage" is presented as a singular, homogenous *American* usage—the "national" standard, not even as the usage of a *particular* American dialect of English. Another implicit standard programmed into the holistic evaluation is the standardized American school essay structure: the five paragraph essay (Burstein, Chodorow, & Leacock, 2004; Attali, 2004).

Figure 1: Attributes of a Holistic Score of 6 for a Persuasive Prompt, Criterion

Holistic Score of 6
   You have put together a convincing argument. Here are some of the strengths
   evident in your writing:

   Your essay:

   • Looks at the topic from a number of angles and responds to all aspects of
     what you were asked to do
   • Responds thoughtfully and insightfully to the issues in the topic
   • Develops with a superior structure and apt reasons or examples (each one
     adding significantly to the reader's understanding of your view)
   • Uses sentence styles and language that have impact and energy and keep the
     reader with you
   • Demonstrates that you know the mechanics of correct sentence structure, and
     American English usage—virtually free of errors

The Trait Feedback is identified on the online demo as "the key reason why *Criterion* is such a valuable remediation tool," —again, the deficit focus. The trait feedback includes five general traits: Organization and Development, Grammar, Usage, Mechanics, and Style. As with the holistic feedback, the trait feedback is of questionable accuracy, implicitly organized around a single standard, and excessively form and error focused. We base this conclusion on our testing of *Criterion's* evaluation of three student-written essays to *Criterion* and also the four versions of the "African American History" essay in Ball (1992). The essays in Ball model four distinct discourse patterns: standard school essay and three patterns that she identifies with African American Vernacular English. Two of the student essays were written by our students. The other one is an essay that Ball and Lardner print in full in *African American Literacies* (2005, pp. 37-49), identifying it as written by Lisa, an "AAVE-speaking student." We report on *Criterion's* evaluation of Lisa's essay in the next section.

Criterion Fails the Test: Misreading Lisa's Essay

Lisa's essay was written in response to a Joyce Carol Oates story, "Theft." (See excerpt in Figure 2). It's a thirteen paragraph essay written in a pattern that Ball and Lardner (2005) identify as "narrative interspersion," where narrative is inserted within the structure of an expository text in the manner of African American oral tradition (p. 47). Now, if Lisa's essay were to be evaluated by *Criterion*, what feedback would she receive to assist her in revising? First of all, her essay couldn't receive a holistic evaluation because the assignment and essay do not conform to the more simplistic prompts required for *Criterion*'s holistic analysis. It did receive Trait Feedback, however, and we use the feedback the essay received to illustrate each of the trait feedback categories and their operating criteria. We will also make reference to the other essays as relevant.

Figure 2: Excerpts from Lisa's Essay

"Theft," by Joyce Carol Oates tells the story of a girl name Marya who has something taken from her. In the beginning Marya had her favorite pen and wallet stolen from her drawer. She did eventually get her wallet but her pen she didn't see anymore. Imogene who's another girl in this story has nothing taken from her. Yet, she becomes Marya's friend and eventually tries to steal her image.

The meaning of "Theft" in this story can be explained in two different ways. One was when Marya had her wallet and pen taken from her. The other was when Imogene her supposed to be friend, tries to take away Marya's image. It's one thing to steal someone's belongings. It's another when you steal someone's image, identity or reputation. And this story "Theft" is a good example of both.

Imogene was the most popular around the campus. Marya was not. Imogene had nice clothes and was very pretty. Marya on the other hand was not. Marya did not have fancy clothes and she wasn't as pretty as Imogene was. What Marya did want was the kind of things Imogene possessed. And there's nothing wrong with that. I would want to have those nice things myself. But what seemed to be the problem was that Imogene pretended to be Marya's friend.

Organization and Development: *Criterion* recognizes only structural features. Further, it is programmed to read for only one culturally specific structure, that of the standardized American school essay. As the analysis of Lisa's essay illustrates, *Criterion* is incapable of accurately identifying the structure of essays that deviate from this pattern, thus leading to *mis*identification of aspects of the essay.   *Criterion* erroneously tells Lisa that she has no introduction, failing to recognize the first paragraph as likely functioning as such. It also erroneously identifies the first sentence as the thesis, when the thesis seems to appear in the second paragraph. Typical of its other stock comments, *Criterion* hedges a bit in the comment: "Is this sentence really part of your thesis? Remember that a thesis controls the whole content of your essay. You need to strengthen this thesis so that you clearly state the main point you will be making. Look in the Writer's Handbook for tips on doing this."

For identification of main ideas, the program underlines the first sentence of each paragraph—except for the first, final, and one following a quote—with everything following first sentences marked as supporting ideas: for example, the first sentences of the second and third paragraphs of Lisa's essay are underlined as main ideas, although that is misleading also.   (With all the other essays, this pattern was also followed—first sentences as main ideas; the following as supporting ideas—whether accurate or not. One student's essay that followed a more inductive pattern was also erroneously flagged as having no thesis. Reflecting *Criterion's* five-paragraph bias, another essay identified as having only two main ideas was flagged "because a good essay will contain at least three main ideas.")   While misleading Lisa with inaccurate information, *Criterion* also

fails to provide useful feedback on the actual structure of her essay, feedback that could serve to either reconfirm what she aimed to do or help her discern that structure.

Criterion's Grammar, Usage, and Mechanics analyses are all error-focused, programmed to read for errors in standardized American English, represented as "general English grammar," the corpus used for identifying these "violations of general English grammar" being 30 million words of newspaper text (Burstein, Chodorow, & Leacock, 2004, p. 28). Figure 3 shows the types of errors identified. For the Grammar trait, "Proofread this!" we see a human personality being invoked, although what types of "errors" evoke this response is not evident, neither is how the program knows whether it is appropriate for a given paper, say, one written by an English language learner or one written by someone using African American English or Spanglish.

Figure 3: Types of Grammar, Usage, and Mechanics Errors Identified by Criterion

Grammar Errors
fragment or missing comma	pronoun errors
run-on sentences	possessive errors
garbled sentences	wrong or missing word
subject-verb agreement	"Proofread this!"
ill-formed verbs	

Usage Errors
wrong article	faulty comparisons
missing or extra article	preposition error
confused words	nonstandard verb or word form
wrong form of word	

Mechanics Errors
misspellings	compound-word errors
missing commas	

As with the feedback on Organization and Development, *Criterion*'s identification of these three categories of errors is often misleading, and sometimes

inaccurate. For example, for Lisa's essay, *Criterion* identified 33 grammar, usage, and mechanics "errors," 23 of which were not errors, reflecting an accuracy rate of only 30%. In other words, at best, this feedback would be distracting, requiring the writer to identify what feedback is correct and what not and in terms of which standard, which presumes the writer has the knowledge to recognize that.

The flagging of Grammar errors shows the *Criterion*'s SAE bias and its limited capability to identify features accurately or helpfully. *Criterion* identified 17 grammar errors in Lisa's essay: 12 fragments or missing comma; 2 run-on sentences, 2 subject-verb agreement, and 1 Proofread this!. The 12 fragment or missing comma "errors," include 6 instances of page references—e.g. "pg. 452"—that *Criterion* did not know how to read, 2 missing commas after an introductory subordinate clause, and 4 fragments, two of them within a quotation. The subject-verb "error"—"if you're the kind of <u>person </u>that basically <u>stay</u> to yourself."—Ball and Lardner (2005) identify as AAVE, which *Criterion* fails to recognize as a viable dialect. The two "Proofread This!" errors were simply additional places where a comma was omitted (e.g., "When you think about <u>it Marya</u> and Imogene. . . ").   The 3 Usage errors detected in Lisa's essay were again errors on *Criterion*'s part, reflecting its inability to detect nuances of language: for example, mistakenly flagging a "wrong article" in the phrase "<u>a</u> girl name Marya" and mistakenly flagging "friend" in the phrase "supposed to be friend" as missing an article. (See paragraph 2, sentence 3 in Figure 2) For Lisa's essays, *Criterion* flagged 13 Mechanics errors, but only 2 were correctly identified, the failure to use a hyphen in the phrase "so-called," which also evidently misled *Criterion* into thinking an article was needed for "so

called friend." Had Lisa followed *Criterion*'s advice, she would have revised to write "so-called a friend."

For Style, the trait report summarizes not "errors," but the number of "comments" relating to the following subcategories: repetition of words, inappropriate words or phrases, sentences beginning with coordinating conjunctions, too many short sentences, too many long sentences, passive voice. Obviously, again, *Criterion* presents a reductive and still "error" based view of style, in contrast to viewing style as rhetorically and creatively motivated uses of language and syntax. Before submitting Lisa's essay to *Criterion*, we conjectured that it would flag the "oral-based idioms" (e.g., "See a person like Imogene is phony to me.") as inappropriate words or phrases and the repetitions that structure the entire essay as repetition. Well, it did pick up on repetition, but not of whole phrases. Instead, *Criterion* identified 178 "repetition of words," specifically Lisa's repeated and understandable uses of the names of the story's main characters (Marya and Imogene), and of the personal pronouns "her," "she," and "you." The use of "you" reflects Lisa's oral style as does her use of coordinating conjunctions to begin sentences (*Criterion* identifies 13 such occurances) What *Criterion* is unable to do is identify this as an aspect of a style. *Criterion* also identifies "too many short sentences," flagging 16 of them, with the injunction that reflects the value-judgment of the comment: "Your essay contains too many short sentences. It will be stronger if you combine some of these sentences with others and vary the length of your sentences more."

To appreciate the limitations of *Criterion* and the error-focus, instead of rhetorical resource focus of the feedback, consider paragraph three in Figure 2. To our minds, Lisa effectively uses short sentences in setting up the contrast between Imogene and Marya

although *Criterion* flags both the second and fourth sentences as "too short," along with the third to the last. The third to the last and the last sentence are also flagged for "beginning with a coordinating conjunction." Finally, every use of "Marya," "Imogene," "did," "she," and "I" are flagged for "repetition of words," with the injunction "You have repeated these words several times in your essay. Your essay will be stronger if you vary your word choice and substitute some other words instead. Ask your instructor for advice." By focusing on these isolated categories—flagged as problems—*Criterion* misses what is happening rhetorically and, in turn, directs the writer's attention to these isolated traits. What doesn't get highlighted is Lisa's rhetorically effective parallel structure—only to be noted if a problem—coupled with her alternation of relatively longer and shorter sentences. Neither are the cues to an oral register noted in the essay overall as that register alternates with a more academic register. In short, the 207 style comments—178 being repetition of words—are distracting and unhelpful.

    Criterion could highlight Lisa's "errors" for her a number of ways. She could call up each category, one by one, and review her draft with each type highlighted: e.g., first view all grammar errors, then view all mechanics errors. Roll-overs would explain the errors in a very limited way, and she could also call up an online *Writer's Handbook* explaining the error. For each category, a bar graph could also be displayed showing how many of each type of error occurred, giving a further sense of objectivity and mathematical power to the analysis. The task for revision would be to reduce these errors. The underlying conception of revision as error elimination is also evident in a paper by an ETS researcher, "Exploring the Feedback and Revision Features of the *Criterion* Service." (Attali, 2004).

We recognize that in being programmed to recognize error and enforce a single language standard, *Criterion* is programmed to respond as some teachers do. For instance, Ball and Lardner (2005) cite some of the following comments from teachers who have been shown Lisa's essay in workshops "The writer cannot write properly." "You shouldn't use informal language." "Proofread." (p. 41). Their point is that if teachers were more knowledgeable about AAVE and read not to find mistakes, but to discern patterns, they would see in Lisa's paper a rhetorically effective use of repetition as a patterning device for the essay and an effective "intermingling of discourses" (p. 49). In short, they would see that "the choices are purposeful." This is something that *Criterion* cannot recognize either. What's different about our human response is that we can choose to respond differently, especially when we learn more about our students and language difference.

Beyond the Bar Graph: Choosing a Different Direction for Teaching and Learning

To suggest that Lisa made *choices* in composing her essay points to a different direction for pedagogy: one that respects the multiplicity of languages and dialects of English; one that aims to teach students to write in a variety of rhetorical genres and contexts, including ones valued in college; one that aims to help them develop the language and rhetorical skills to make choices as they compose based on their intentions in a given rhetorical situation. (See, for example, Lu, 1994; Lovejoy, 2003;Okawa, 2003; Stanley 2009). The *Criterion* materials imply that *Criterion* can be used for such a pedagogy. For example, *The Criterion Teaching* Guide invokes Chickering and Gamson's "Seven Principles for Good Practice in Undergraduate Education, " including active learning,

prompt feedback, and respect for diversity of talents and ways of learning" (ETS, 2007, p. 3). However, the automated assessment features of *Criterion* are <u>not</u> central to any of these lessons; instead, what are used are features for posting and sharing drafts with one's teacher and/or peers—features that many other web-based platforms offer, without the additional cost of *Criterion*. Tellingly, there are not examples in the *Teaching Guide* of lessons that would make visible to students the language values programmed into *Criterion* or involve students in critically examining those values and standards. Such lessons would be at odds with the ideology of language promoted by *Criterion*. Yes, *Criterion* does enable "active learning" but using a narrow and limited range of prompts and no more so than any writing course that asks students to write; it provides prompt feedback, but enacting a single standard and often either inaccurate or misleading feedback; fundamentally, while it aims to "respect diversity of talents," it does not respect "the multiplicity of language, of English, in our classrooms."

If the findings of the 1996-98 CCCC survey of English teachers are still valid—and we suspect they are—*Criterion* mirrors beliefs that some of us hold about a single appropriate standard of English, beliefs that are at odds with our own organization's position statements and scholarship on language, rhetoric, and composition. We should look at the mirror self-critically, then, for what it reflects about ourselves and our values. That is the anti-racist work that Keith Gilyard (1999) calls us to in "Higher Learning: Composition's Racialized Reflection." What unacknowledged ideologies are implicit in our beliefs about language and our students, and what ideologies do we promote using the technology of *Criterion*—not only ideologies about language, but also about the function of response to student writing?

We can also learn by reading scholarship on language, race, rhetoric, and ethnicity and by studying our students' writing: an automated assessment program with bar graphs of errors will not help us discern language and rhetorical patterns in our students' writing that may reflect standards other than the program's implicit "standard" and that are resources for their writing and learning, not hindrances.

Notes

1. *My Access* and *Holt Online Essay Scoring* are marketed for K-12 instruction. A new web-based program, *Sentenceworks*, includes an "automated grammar tutor" that is being marketed as "like having a dedicated writing coach always available at your fingertips."

2. The *Criterion* website provides a partial list of "clients," listing 43, including, for example, the Air War College, AL; Argosy Education Group, IL; Dallas County Community College, TX; East Texas Baptist University; Miami Dade College, FL; Texas A & M; UC Irvine, CA: Wayne State, MI; and Westwood College Online, CO (ETS, 2008). Of the 38, seven are in Texas and five California. The extent of the uses at all of these schools seems to range from a single teacher to full programs, making it difficult to infer just how pervasive the use of *Criterion* is.

References

ACT. (2009). COMPASS Writing Essay Test (e-Write). Retrieved from
http://www.act.org/compass/tests/writingessay.html

Attali, Y. (2004). Exploring the feedback and revision features of *Criterion*. Retrieved
from http://www.ets.org/Media/Research/pdf/erater_NCME_2004_Attali_B.pdf

Ball, A. (1992). Cultural preference and the expository writing of African-American adolescents. *Written Communication*, 9, 501-532.

Ball, A. (1999). Evaluating the writing of culturally and linguistically diverse students: The case of the African American English speaker. In C. Cooper and L. Odell (Eds.), *Evaluating writing: The role of teachers' knowledge about text, learning, and culture* (225-248). Urbana, IL: NCTE.

Ball, A., & Lardner, T. (2005). *African American literacies unleashed: Vernacular English and the composition classroom*. Carbondale: Southern Illinois UP.

Balester, V. (1993). *Cultural divide: A study of African-American college level writers.* Portsmouth, NH: Heinemann-Boynton/Cook.

Broad, B. (2006). More work for teacher? Possible futures of teaching writing in the age of computerized writing assessment. In Ericsson and Haswell, 221-233.

Burstein, J., Chodorow, M., & Leacock, C. (2004). Automated essay evaluation: The *Criterion* Online Writing Service [Electronic version]. *AI Magazine,* 25(3). 27-36.

Canagarajah, S. (2006). The place of World Englishes in composition: Pluralization continued." *College Composition and Communication*, 57, 586-619.

Chen, C.-F.E., & Cheng, W.-Y. E. (2008). Beyond the design of automated writing evaluation: Pedagogical practices and perceived learning effectiveness in EFL writing classes. *Language, Learning & Technology*, 12(2), 94-112.

College Board. (2009). WritePlacer: ACCUPLACER's writing assessment. Retrieved from http://professionals.collegeboard.com/higher-ed/placement/accuplacer/write

Conference on College Composition and Communication. (1974). *Students' right to their own language.* Special issue of *College Composition and Communication* 25.

Conference on College Composition and Communication. (1988). *National language policy*. Retrieved from http://www.ncte.org/cccc/resources/positions/nationallangpolicy.

Ericsson, P. F. and Haswell, R. (Eds.). (2006). *Machine scoring of student essays: Truth and consequences*. Logan: Utah State UP.

ETS. (2008). *Criterion*[SM] Online Writing Evaluation. Retrieved from http://www.ets.org/portal/site/ets/menuitem.435c0b5cc7bd0ae7015d9510c3921509/?vgnextoid=b47d253b164f4010VgnVCM10000022f95190RCRD.

ETS (2007). *The Criterion teaching guide: Using the Criterion Online Evaluation Service for differentiated instruction in the college classroom: A guide for faculty and administrators.* Retrieved from http://www.ets.org/Media/Resources_For/Higher_Education/pdf/*Criterion*_Teacher_Guide_web_6487.pdf

Gilyard, K. (1999). Higher learning: Composition's racialized reflection. In K Gilyard (Ed.), *Race, rhetoric, and composition* (44-52). Portsmouth, NH: Boynton/Cook.

Herrington, A., & Moran, C. (2001). What happens when machines read our students' writing? *College English*, 63, 480-499.

Herrington, A., & Moran, C. (2006). WritePlacer *Plus* in place: An exploratory case study. In Ericsson and Haswell, 114-129.

Herrington, A., & Moran, C. In press. Writing, assessment, and new technologies. In

Paretti, M., & Powell, K. (Eds). *Assessment of writing*. Tallahassee FL: Association of Institutional Researchers.

Hesse, D. (2005). Who owns writing? *College Composition and Communication*, 57, 335-57.

Holt, Rinehart, and Winston. (2009). *Holt Online Essay Scoring*. Retrieved from http://my.hrw.com/hosdemo/hos_entry.jsp

Horner, B., & Trimbur, J. (2002). English Only and U.S. college composition. *College Composition and Communication*, 53, 594-630.

Hutchison, D. (2007). An evaluation of computerized essay marking for national curriculum assessment in the UK for 11-year-olds." *British Journal of Educational Technology*, 38(6), 977-989.

Kemp, F. (1992). Who programmed this? Examining the instructional attitudes of writing support software. *Computers and Composition*, 10, 9-24.

Lovejoy, K.B. (2003). Practical pedagogy for composition." In Smitherman and Villanueva, 89-108.

Lu, M-Z. (1994). "Professing multiculturalism: The politics of style in the contact zone." *College Composition and Communication*, 45, 442-458.

Lu, M.-Z., Matsuda, P.K., & Horner, B. (Eds.). (2006) Special issue: Cross-language relations in composition. *College English* 68.6.

Matsuda, P. K. (2006). The myth of linguistic homogeneity in U.S. college composition. *College English*, 68, 637-51.

Okawa, G. (2003). "Resurfacing roots": Developing a pedagogy of language awareness from two views." In Smitherman and Villanueva, 109-133.

Richardson, E. (2003a). *African American literacies*. New York: Routledge.

Richardson, E. (2003b). Race, class(es), gender, and age: The making of knowledge

about language diversity. In Smitherman and Villanueva, 40-66.

Sentenceworks. (2009). Retrieved at http://www.sentenceworks.com/

Smitherman, G. (1993). "The blacker the berry, the sweeter the juice": African

American student writers and the National Assessment of Educational Progress.

National Council of Teachers of English Conference, Pittsburgh, PA, November

17-22, 1993. ERIC Document Reproduction Service, ED 366 944.

Smitherman, G., & Villanueva, V. (2003). *Language diversity in the classroom: From

intention to practice*. Carbondale: Southern Illinois UP.

Stanley, S. Writing a path toward social change: A critical pedagogy of grammatical

choice." Conference on College Composition and Communication. San Francisco.

12 March 2009. unpublished ms.

Vantage Learning. (2009). *My Access*. Retrieved from

http://www.vantagelearning.com/school/products/myaccess/

Williamson, M. (2009). Validity of automated scoring: Prologue for a continuing

discussion of machine scoring student writing.   In B. Huot & O'Neill, P. (Eds.),

Assessing writing: A criticalsSourcebook (435-55). Boston: Bedford/St.

Martin's.