Journal of Writing Assessment Vol. 2, No. 1, pp. 149-154

REVIEW

Testing the Tests

SUSAN CALLAHAN
Northern Illinois University

George Hillocks. *The Testing Trap: How State Writing Assessments Control Learning.* New York: Teachers College Press, 2002.

The Testing Trap is more than a catchy alliterative title; it is an apt metaphor for the situation currently experienced by America's K-12 teachers who find themselves pressured in both subtle and painfully obvious ways to teach to the expectations of state writing tests. In this meticulously crafted volume, George Hillocks, Jr. analyzes the writing tests used in Texas, Illinois, New York, Kentucky, and Oregon in order to demonstrate their influences on classroom instruction. By choosing states with widely differing writing tests, he is able to reveal how each test shapes the concept of "good writing" within the state that uses it. Although Hillocks' analysis of the evidence from these five states strongly suggests that most current state writing tests promote neither good writing nor good writing instruction, the real value of this volume lies in his thorough exploration of the reasons for these failures.

Before turning to detailed examinations of the five state tests, Hillocks uses the first four chapters of *The Testing Trap* to provide basic information about the tests, to lay out the theoretical basis for his analysis, and to describe his research methodology. He begins with a discussion of the problem as he sees it: America has chosen to focus attention and tax dollars on detailed, written standards and elaborate

Susan Callahan is an associate professor in the English Department of Northern Illinois University where she teaches courses in composition, rhetoric, and English education. She is the past chair of the NCTE Standing Committee on Testing and Evaluation and has conducted a long-term study of the Kentucky writing portfolio assessment. Her work has appeared in *Research in the Teaching of English*, *English Education*, and *Educational Assessment*.

Direct all correspondence to: Susan Callahan, English Department, Northern Illinois University, DeKalb, IL 60115

testing systems rather than on improved teaching. Because writing tests are intended, in large part, to measure how successful students are in reaching high standards for critical thinking, Hillocks believes it is important to examine exactly what these tests actually test. At this point, he introduces the interpretative framework he will use throughout the book: that good writing entails good thinking and that good thinking develops from good teaching.

Chapter 2 is primarily an explanation of the research methodology used by Hillocks and his associates. The five states examined in *The Testing Trap* were chosen to represent diverse geographical locations as well as to demonstrate the types of tests currently in use and the consequences each type of test can have for teachers and students. He shows that state writing tests range in complexity from the single prompt, forty minute timed writing used in Illinois to the elaborate portfolio system used in Kentucky, with many interesting variations in between. Some tests, such as the one in Texas, have high stakes for students who must pass them in order to graduate, whereas others, such as the one in Kentucky, have high stakes for individual schools. The Illinois test represents those that purport to have no stakes for either students or teachers.

Hillocks explains that for each state test, he analyzed the writing theories cited by the test designers; the resulting standards, prompts, and rubrics; the scoring procedures; and the benchmark papers used to illustrate score categories. To determine the tests' effects on teaching, he used the results of coded interviews with state department officials and with teachers and administrators in six school districts of various sizes and demographics in each state. Although the interviews were as comprehensive as possible, they were somewhat limited by time constraints. Hillocks notes that he also wanted to conduct classroom observations, but funding for such work was not available. Data collection was further hampered by the fact that the Oregon and New York tests were not yet well established. Consequently, the classroom impact of these two tests could not be assessed to the same degree as was possible for the other three. Finally, Hillocks does not specify the processes used to select schools for visitation and individual teachers for interview, leaving open the possibility that some state and district officials may have steered these selections in ways that compromised the researchers' perceptions of the tests' effects. Despite these remaining questions, *The Testing Trap* contains more than enough evidence to support Hillocks' contention that most state writing tests have detrimental effects on teaching and learning.

Chapter 3 is provides an overview of the legislative and theoretical foundations of state writing tests. Hillocks examines the often impressive-sounding state standards and official explanations for the content of writing tests to expose the ways that the language in such documents does not match the reality of the tests themselves. He shows that even those states claiming to base their tests on a rich theory of writing usually narrow their tests to a few generic forms. Some states, represented in this volume by New York and Oregon, do not base their tests on any writing theory at all, instead relying on what they believe students should be able to do.

This discussion leads neatly into Chapter 4, the heart of Hillock's textual research. Here he introduces the variables that exist within state tests and shows how these variables affect how writing is understood and taught in different states. The variables he examines are the nature of the prompt or prompts; what counts as a good response; the amount of time allowed for writing; the number of pieces of writing required; the scoring criteria, personnel, and procedures used; the grade levels tested; the kinds of writing tested; the conditions under which the writing is done; the access students have to information that can be used to shape responses; and the feedback teachers and students receive. He then turns to a close examination of the three variables he feels have the greatest impact on instruction: the writing prompts, the stated evaluation criteria, and the benchmark papers used to illustrate how the criteria are applied. He finds that although most prompts are clearly written, they often ask students to do things that are impossible given the constraints imposed by other testing variables. For instance, students in Illinois are asked to write persuasive essays but are given neither the time nor the access to information needed to support persuasive claims. But banal or misleading prompts are not the biggest difficulty that he sees. Far more important are the ways student responses are judged in relation to the stated writing tasks. The evaluation criteria are often ambiguous—for instance, asking evaluators to distinguish between papers that "develop some ideas more fully than others" and those that only "develop ideas briefly." Moreover, these criteria are often interpreted in ways that reward stunningly bad writing.

In the reminder of the volume, Hillocks patiently strips away the rigorous-sounding claims made by the test designers in Texas, Illinois, New York, Kentucky, and Oregon to reveal how little the tests actually demand of the test-takers. He then shows how these minimal expectations negatively influence the way writing is taught. He begins with four chapters on Texas and Illinois, the states with the most potentially harmful tests. For each state, one chapter addresses the flaws of the various components of these single-sitting tests, and the other describes some of the damaging effects the tests have on administrators, teachers, and students. He draws clear connections between the demands of the tests and the instructional strategies of the teachers, most of whom use didactic, presentational teaching methods to enforce the importance of essay structure and give little attention to the content the structures are supposed to enhance. In these states, students are rewarded for "blethering," a term Hillocks uses to describe the nonsense that results when students are forced to write about a subject they neither care about nor really understand. He also describes the insidious effects of supplementary teaching materials marketed to help teachers drill students on the required forms, the narrowing of the writing curriculum especially in schools with many children living in poverty, and the pressure all teachers feel to conform to the tests' standards for writing. He concludes that in Texas and Illinois, the vacuous five-paragraph theme has become the standard of good writing, a standard that some teachers accept and willingly promote.

New York's unusual, literature-based writing test is analyzed in Chapter 9. New York is the only state in the nation to require four pieces of writing in response to challenging prompts asking students to critique both familiar and unfamiliar pieces

of literature. The test is given over a 2-day period, and the evaluation criteria appear to set high expectations. When Hillocks turns to the benchmark papers used to exemplify varying degrees of writing success, however, he is once again able to document how the criteria are interpreted in ways that reward writing lacking any real substance. He speculates that the reason for this gap between stated expectations and reality is the fact that no state wants a high failure rate. He believes all states, not just New York, use benchmark papers as a way to ensure that a relatively low number of students will fail to meet their widely touted high standards. Although New York avoids a stultifying focus on the five-paragraph theme, teachers still are not encouraged to create the engaging classrooms that research indicates would best prepare students for exams requiring true critical thinking. Instead, like teachers in Texas and Illinois, they rely primarily on presentational teaching methods.

Kentucky and Oregon are discussed together in Chapters 10 and 11. Kentucky's writing assessment is a portfolio containing samples of several genres gathered over several years, including a piece from a class other than English. Oregon's students receive three prompts over a 3-day period. Each prompt specifies discourse mode and audience but gives students a choice of topic and extended time in which to respond. The state also requires, but does not evaluate, a "work sample" consisting of four pieces of writing, one each in the narrative, imaginative, expository, and persuasive modes. Unlike the other states, both Kentucky and Oregon value voice in writing as well as the students' ability to select and develop their own topics. Although Oregon's test has not been in operation long enough to gauge its impact on instruction, teacher interviews suggest it is unlikely to encourage formulaic writing even though the test writing, unlike the work sample, is not evaluated for voice in deference to legal opinion that voice is too subjective to be scored.

Hillocks finds Kentucky's portfolio test to be the strongest of the state tests he examined, calling it a "model for imitation" (p. 188). He praises the breadth of writing it demands, which includes imaginative and reflective writing, and the professional development that the state provides for teachers. He finds far more teachers eschewing presentational teaching in favor of writing workshops and inquiry-based methodology. He also finds teachers in all subject areas assigning writing in their classrooms. He does, however, note that Kentucky, like Illinois and Texas, accepts "support" for ideas in place of factual evidence and logical reasoning, thus encouraging teachers and students to settle for "persuasive" writing samples that do not present well-argued positions. He does not discuss how well or how poorly the imaginative and reflective components of the portfolio are assessed, focusing his critique solely on the rubric's weakness in the areas of expository and persuasive writing.

In Chapter 12, Hillocks reinforces the conclusions drawn in other chapters and returns to his thesis: that good writing—writing that demonstrates and stimulates good thinking—grows from good teaching. Although he does include imaginative thinking in this definition of good writing and laments the fact that imaginative writing is seldom taught because most states do not test it, he clearly believes that good writing is best examined in terms of an argumentative essay for which students have adequate time and access to appropriate data and/or texts. To support his belief in the importance of teaching reasoned argument, he provides useful sug-

gestions for teachers who want to employ an inquiry-based methodology to foster critical thinking. He does not, however, follow this discussion with ideas for using inquiry-based teaching in conjunction with other methods intended to encourage the development of student voice, student control over topic selection, and student experimentation with writing in multiple genres for various audiences and purposes—elements of the Kentucky portfolio assessment he shows to have had positive effects on pedagogy in that state. Furthermore, although Hillocks never suggests that state writing assessments should consist solely of a persuasive essay, his intense focus on this genre certainly has the effect of giving it priority and may have the unintended effect of encouraging states such as Illinois and Texas to follow his suggestions for changing their prompts, their single writing session limitations, and their evaluation criteria without making the more substantive changes he calls for.

The larger changes in education that Hillocks suggests—more money spent on professional development, smaller classes, and increased class time for writing instruction—are not new ones. In fact, they are discouragingly familiar. They are not, however, impossible to attain. Hillocks shows that Kentucky has taken a dramatic step in that direction by designing a writing test that encourages teachers to provide thoughtful, sustained writing instruction and by supporting the test's writing goals with adequate funding for professional development. Although the Kentucky writing test is similar to the other more limited and limiting tests Hillocks studied in that its standards for persuasive writing are not rigorously written or rigorously enforced, it does demonstrate that writing tests do not need to narrow the concept of good writing to the mastery of one or two formulaic essay forms. Unfortunately, Hillocks' praise for the Kentucky writing test also may be used as support for the position that state tests should continue to be used to drive changes in instruction, a position that is hotly contested by many in education who feel that even a well-designed trap is a trap.

In *The Testing Trap* George Hillocks demonstrates that the current focus on using writing tests to improve pedagogy and curriculum simply is not working as intended. State tests do, indeed, influence classroom practices, but most of these influences are not positive ones. Instead of encouraging a rich writing curriculum, most state tests promote a narrow range of writing activities. Instead of promoting critical thinking and strong writing voices, state tests settle for empty prose in cookie-cutter formats. Furthermore, because state writing requirements can be met by teaching students to reach minimal standards, teachers often limit themselves to those standards. Sadly, many teachers do not realize what they and their students are missing. Having had little preparation to teach writing and few meaningful professional development opportunities, they accept the writing models supplied by the state and the prepackaged teaching materials they are offered as valuable additions to their classrooms; and working with impossibly large numbers of students, many are grateful that writing is not expected to play a larger role in their classrooms. In fact, the most disturbing conclusion to emerge from this study may well be that teaching to the test is cheating so many teachers and students out of discovering what good writing really is.

No one who reads *The Testing Trap* can continue to believe that state writing tests are merely assessments of student progress. George Hillocks has shown that

their influences on education are profound and far-reaching. Moreover, he has provided a workable methodology for evaluating state writing tests and the educational goals they promote. His methodology is one that educational leaders could well employ to study the design, implementation, and consequences of the writing tests they currently require. Such an analysis, unlike test score reports, would provide truly useful information about student writing. The next step, of course, would be deciding what to do with this information. An obvious possibility would be to improve state writing tests, but readers of *The Testing Trap* may well decide it is time to dismantle the trap and turn our efforts and resources toward actually improving writing instruction.