

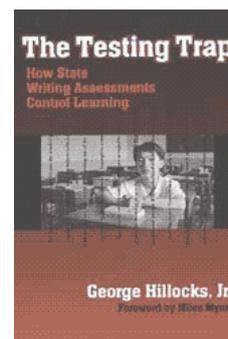
Book Review:

## The Testing Trap: How State Writing Assessments Control Learning

by George Hillocks, Jr.

New York: Teachers College Press, (2002)

ISBN 0-8077-4229-5



What effect do high-stakes L1 writing tests have on writing ability? This book attempts to answer that question in K-12 contexts by examining the mandated testing programs of five U.S. states. It illustrates how government mandated tests can adversely impact writing instruction and underscores some of the problems of relying on high-stakes tests to drive curricula. Hillocks, a professor emeritus of writing at the University of Chicago, denounces the *No Child Left Behind* policies of Bush for promoting the superficial performance of explicit test features at the expense of in-depth learning.

### Chapter 1

This book starts off by examining how political factors impact assessment policies. Outlining the rise of standards-based education in the United States, the author questions the rhetoric espoused by politicians seeking to initiate school reform through testing. Noting how state-mandated testing programs tend to increase drill and memorization practice and the likelihood of non-performing students dropping out, Hillocks rhetorically inquires:

... should we expect tests, in and of themselves, to bring about higher achievement in schools? Do we believe testing will bring about changes in what happens in classrooms, change that will in turn bring about different and perhaps higher levels of student achievement? Do we believe that simply the threat of sanctions will encourage students and teachers to work harder? Why should we believe that? (p. 12)

The first sixteen pages of this text offer a montage of what happens when political imperatives collide with constructionist notions of learning. The aftermath is not only a compromise of test content, but also a sort of *educational McDonaldisation* (Hayes and Wynyard, 2002). One thing that means is that extended essay writing which is not easily tested gets sidelined.

Hillocks notes how testing pervades our culture and our general ambivalence towards testing. Though many accept the myth that "tests indicate achievement, intelligence, or aptitude, or all of these" (p. 14), others complain about the human toll involved in testing and our almost morbid obsession with test scores. Hillocks echoes Ken Jones (2000) in stating "... the prerequisites of high stakes have lead inexorably to standardization and disempowerment."

The author encourages people to consider whether tests measure the skills they purport to. Most writing tests, in Hillocks' view, amount to little more than superficial pre-writing exercises, since examinees are not given enough time to critically develop their essays. Schuster (2004) echoes this view, describing most standard writing tests as "tests of drafting" rather than "tests of writing" due to time constraints and the pervasive use of the five-paragraph essay format. Hillocks describes most timed student compositions as nothing but "organized blather" [i.e. blather] (p. 80) and questions whether the tendency to superficialize education is purely accidental by suggesting:

Part of the problem in this country is that systematic thinking about difficult problems seems to be confined to an elite group. Most Americans are not willing to think much beyond their own desires and perspectives, which often determines what they think is right and appropriate. Perhaps most people cannot think carefully about complex problems. (p. 6,7)

Would these same comments apply to Japan's educational system? If Baker and LeTendre (2000, p. 356) are correct, the situation is by no means exclusive to the USA.

## Chapter 2

Hillocks directs his rhetorical cannons at the writing assessment tests in five U.S. states. Noting how "millions of dollars, thousands of teacher hours, and hundreds of thousands of classroom student hours" are spent on mandated writing assessments, the author concludes it's mostly a wasted investment. He also explores the connections between assessment methodology and current traditional rhetoric (CTR). Rooted in the notion of an objective, external truth, CTR is juxtaposed to the New rhetoric of Perelman (1969) and Berlin (1982). The author's sympathy for the latter and disdain of the former is not concealed. At times Hillocks' censure of opposing views borders is almost too doctrinaire, as in this example:

Contemporary educational theory and practice argue that effective learning must be constructivist in nature, and that students learn best and perhaps only when they can construct knowledge for themselves within the framework of their existing knowledge. (p. 22)

This type of argument presupposes that there is only one sort of learning. However theorists such as Bloom (1956) and Gardner (1983) have suggested that there are in fact multiple types of learning. Rote learning might be appropriate for certain tasks, but certainly not for the full scope of education.

## Chapters 3-4

Differences in the assessment standards of several U.S. states are quickly overviewed. Whereas some states have minimal writing competency requirements for high school graduation, others don't. Moreover, whereas some states include student portfolios in their assessments, other states rely exclusively on timed in-class exercises. Instead of thinking of "writing tests" as uniform, the author points out how writing is operationalized quite differently by various testing agencies.

## Chapter 5-6

Chapter 5 is a lengthy lambaste of the Texas Assessment of Academic Skills, which was in use from 1991 to 2003 (Wikipedia, 2006). Hillocks contends that the narrow focus of this test fosters writing instruction on surface mechanics rather than deep content. Noting how a composition ". . .

*"Most writing tests, in Hillocks' view, amount to little more than superficial pre-writing exercises, since examinees are not given enough time to critically develop their essays."*

may be grammatical, well-organized, and coherent without being well-reasoned, thoughtfully developed, or effective in any way", (p. 35) he alludes to the need for teaching more critical thinking skills in composition classes.

Hillocks also criticizes the way the passing scores can be politically manipulated - instead of investing money to raise test standards, it's often easier to simply lower test standards so that a greater number of students will pass.

The chapter that follows claims to examine teacher responses to the Texas test. Unfortunately, the author doesn't mention how many teachers were interviewed or anything about his data collection process. In a book that is critiquing testing methodology, this type of error is perplexing. All of Hillocks' information is reported in percentile figures without any reference to sample sizes. One could, perhaps, regard this chapter as a qualitative case study. Even by those standards, however, not enough background information about the study to actually interpret it. In short, we are left with nothing but rhetoric and anecdotes.

## Chapter 7-8

The attack then shifts to Illinois. Hillocks contends that the benchmark papers and detailed prompts for the Illinois Goals Assessment Program (IGAP) writing exam leads to a "homogenization of writing" (p. 110) which is predictably banal. Hillock ascribes this condition to two factors: government restrictions and the fact that the tests aren't solidly grounded in what he believes to be sound writing theory. Since only 15 school days per academic year in Illinois may be used for testing, the state writing exams are limited to 40 minutes. In Hillock's view, the theories of Kinneavy (1971) provide the best grounding for writing tests. Kinneavy creatively fused classical rhetoric with communication theory and underscored the importance of textual modes (description, narration, evaluation, and classification). His ideas have shaped the writing exams in many U.S. states, but apparently not Illinois. The Illinois test, according to Hillocks, focuses too much on the traditional five-paragraph essay and not enough on rhetorical principles.

Teacher responses to this test were described as "ambivalent" (p. 124). Once again, not enough methodological information is provided to critically interpret the data.

### Chapter 9

The next chapter starts out with high praise for the target objectives of the New York State Regents exams, then an close look at a benchmark essay for that exam. Gradually, the author voices concerns about whether or not the test is being graded in a way that measures its avowed standards. Hillocks contends that what is purportedly "in depth analysis" often amounts to little more than a blithe regurgitation of previously stated facts.

### Chapters 10 and 11

Shifting to Kentucky and Oregon, the author quickly delineates how writing is assessed in both states. Both states employ some type of portfolio assessment as well as in class timed-writing. Predictably, the author tears apart the ways that tests are graded, contending that the benchmark sample papers often do not fit the standards specified in the test rubrics. This chapter concludes by offering some praise for the Kentucky writing portfolio assessment system. Hillocks says it is one of the few writing tests he examined that strives for some semblance of a real audience and a broader writing purpose (p. 46).

*"the Kentucky writing portfolio assessment system . . . is one of the few writing tests . . . that strives for some semblance of a real audience and a broader writing purpose."*

### Chapter 12

The book concludes with a recap of the main themes of the text. In fact, readers could skip the first eleven chapters of this book and glean the pivotal points from this final chapter. After briefly overviewing the malaise afflicting all five states examined in this text, the author turns his attention to possible solutions.

The main need, in Hillocks view, is not more testing but better teacher training:

At the center of the K-12 testing fury is the myth that testing alone is able to raise standards and the rate of learning. Certainly, testing assures what is tested is taught, but tests cannot assure that things are taught well. If states want teaching to improve they will have to intervene at the level of teaching. Teachers need opportunities to learn more effective procedures for teaching writing. Tests of writing cannot teach that.  
(p. 204)

In addition to providing teacher training, Hillocks feels it is essential that teachers have adequate time to teach writing well. Devoting classroom time to tests takes away from time that could be spent on writing. And since most American writing textbooks geared for the K-12

market are stuck in CTR-mode, Hillocks believes teachers also need time to develop their own writing materials. However, with an average American K-12 writing teacher dealing with 130 students a day, can teachers actually develop their own materials and also give detailed feedback on the essays they receive? The fact that formulaic CTR essays are much easier to grade results in a sort of "production line" mentality among many public school teachers and there is an economics at work that favors McDonaldised instruction. Unless public pressure is increased on law makers and state officials, we may well see a two-tiered educational system: with elite schools for the wealthy few and mediocre schools for all others. Hillocks ends exhorting readers to get away from "vacuous thinking and writing" and make a commitment to "high literacy".

### Conclusion

At times this book could be criticized for being too narrowly focused on the high school writing programs of merely five US states during the 1990s. However, it also raises perennial questions relevant to language teachers around world .

One issue concerns *content validity*. Although many of the writing tests examined in this book have technical reliability, most lack content validity. Writing is a skill which defies superficial, facile assessment: the communication of deep ideas is hard to translate readily into numerical variables.

Another issue concerns the *optimal mode of instruction*. In high stakes testing environments, Hillocks contends that "instruction" often degenerates into mere examination-preparation drilling. He argues against the traditional lecture/recitation modes of learning and emphasizes the need for

*"Writing is a skill which defies superficial, facile assessment: the communication of deep ideas is hard to translate readily into numerical variables."*

more in-depth student interactions. He also complains about the excessive teacher talk in most classes, noting that when students do speak it is often to answer teacher questions through a single word or phrase. In other words, classroom discourse too often is disconnected and

fragmented - ironically similar to the information on many tests. Hillocks echoes Hargreaves and Fink (2003) in criticizing the breadth-without-depth nature of most school interactions, which in turn is mirrored in the shallowness of most tests. Stressful high-stakes testing environments, in Hillocks view, effectively prevent deep interaction and augment the sense of alienation reverberating through many classrooms.

A third issue concerns *test washback* in general. Perhaps a bit too simplistically, Hillocks views state-mandated exams as significantly impacting teacher behavior. Cheng (2004, p. 148) suggests washback is multifaceted and complex and Stoll (cited in Haney, 2000) adds, "It's immensely hard to get a critical mass of teachers within a school, let alone a district, to significantly change their practice." Even as new laws and new tests are enacted, teachers are often remarkably inure to any change.

Are the 224 pages of this book actually worth reading? For American K-13 school administrators, perhaps. For teachers of EFL, or indeed any foreign language with an interest in assessment issues, however, I can't help but wish this volume would be condensed into a single chapter of a larger work which deals with test impact in terms of all four language skills from a wide range of international contexts. Such a book would begin to cover the scope suggested by its title.

- reviewed by Tim Newfields

### References

Baker, D. P., & LeTendre, G. K. (2000). Comparative sociology of classroom processes, school organization, and achievement In M. T. Hillinan (Ed.). *Handbook of the sociology of education*. Berlin: Springer. 345-364.

Bloom B. S. (1956). *Taxonomy of educational objectives, Handbook I: The cognitive domain*. New York: David McKay Company Inc. (republished in 1984 by Pearson Education).

Cheng, L. & Watanabe, Y. (2004). *Washback in language testing: Research contexts and methods*. Mahwah, NJ: Laurence Erlbaum & Associates.

Gardner, H. (1983). *Frames of mind*. New York: Basic Book Inc.

Hargreaves, A. & Fink, D. (2003, September). *Educational reform and school leadership in 3-D perspective*. Nottingham: National College for School Leadership. Accessed December 4, 2006 at <http://www.ncsl.org.uk/media/415/E5/educational-reform-and-school-leadership-in-3d-perspective.pdf>.

Haney, W. (2000, August 19). The myth of the Texas miracle in education, Part 8: Summary and lessons learned. *Education Policy Analysis Archives*, 8 (41). Accessed December 4, 2006 at <http://epaa.asu.edu/epaa/v8n41/part8.htm>.

Hayes, D. & Wynyard, R. (2002). *The McDonaldization of higher education*. Westport, CT & London: Bergin & Garvey.

Jones, K. (2000). High stakes vs. democracy. *FairTest Examiner*. Accessed December 1, 2006 at <http://www.fairtest.org/examarts/Fall%2000/Jones.html>.

Kinneavy, J. L. (1971). *A theory of discourse*. Englewood-Cliffs, NJ: Prentice-Hall.

Newkirk, T. (1999, December 1). *Teaching to the test means 'dumbing down' the curriculum*. Accessed December 19, 2006 at [http://www.nexthorizon.unh.edu/news/news\\_releases/1999/december/tm\\_19991201curriculum.html](http://www.nexthorizon.unh.edu/news/news_releases/1999/december/tm_19991201curriculum.html)

Schuster, E. H. (2004, January). *National and state writing tests: The writing process betrayed*. Accessed December 1, 2006 at <http://www.pdkintl.org/kaplan/k0401sch.htm>.

Wiggins, G. P. (1998). *Educative assessment: Designing assessments to inform and improve student performance*. San Francisco: Jossey-Bass.

Wikipedia. (2006). *Texas Assessment of Academic Skills*. Accessed December 6, 2006 at [http://en.wikipedia.org/wiki/Texas\\_Assessment\\_of\\_Academic\\_Skills](http://en.wikipedia.org/wiki/Texas_Assessment_of_Academic_Skills).

**HTML:** [http://jalt.org/test/new\\_6.htm](http://jalt.org/test/new_6.htm) / **PDF:** <http://jalt.org/test/PDF/Newfields6.pdf>