



## What Teachers Say About Different Kinds of Mandated State Writing Tests

PEGGY O'NEILL

*Loyola College in Maryland*

SANDRA MURPHY

*University of California-Davis*

BRIAN HUOT

*Kent State University*

MICHAEL M. WILLIAMSON

*Indiana University of Pennsylvania*

This article reports on a study of the impact of the form of high-stakes, state-mandated writing tests on high school curricula and teaching practices. Through surveys and focus group interviews of high school English teachers, we addressed two main

---

**Peggy O'Neill** is associate professor of writing, directs the composition program, and teaches writing in the Department of Writing at Loyola College in Maryland. Her scholarship focuses on writing assessment, pedagogy, and writing program administration and disciplinarity. She is currently at work, with Brian Huot and Cindy Moore, on the *Handbook of College Writing Assessment*.

---

**Sandra Murphy** is a professor who teaches graduate-level courses on research on writing, reading, and assessment in the School of Education at the University of California-Davis. She has written several articles on the teaching and learning of writing and reading and has co-authored several books on writing assessment. Her research interests include writing assessment and its impact on teachers and curriculum, reading comprehension, and critical perspectives on literacy.

---

**Brian Huot** is professor of English and Writing Program Coordinator at Kent State University. His work on writing assessment has appeared in various articles, book chapters, and in the monograph *(Re)Articulating Writing Assessment for Teaching and Learning*. He is co-founder and editor of the *Journal of Writing Assessment*.

---

**Michael M. Williamson** is professor of English at Indiana University of Pennsylvania. His research interests include writing assessment theory and practice and empirical research methodologies. His work has appeared in various journals and anthologies devoted to the study of writing. He is, along with Peggy O'Neill, co-editor of the Hampton book series on the study of writing.

---

Direct all correspondence to: Peggy O'Neill, Writing Department, Loyola College, 4501 N. Charles St., Baltimore, MD 21210, [poneill1@loyola.edu](mailto:poneill1@loyola.edu)

questions: How do different kinds of high-stakes, statewide assessments impact writing curriculum and instruction in secondary schools? and What are teachers' views of the impact of different kinds of high-stakes tests? We conducted our study in three states—California, Georgia, and Kentucky—each with a different type of writing test at the time of the study: multiple choice, timed impromptu, and portfolio, respectively. The survey results contribute to the growing body of research that indicates the forms of writing tests influence what teachers teach and how they teach it. This influence was complex with significant differences across the three states in the types of assignments, the length of assignments, number of drafts, and the amount of time allowed for assignments. Our results also indicated that the form of high-stakes writing tests also impacts teacher morale and attitudes.

In the wake of *A Nation at Risk* (National Commission on Excellence in Education, 1983), state-mandated testing in K–12 spread throughout the United States during the 1990s as part of accountability and educational reform efforts. Policymakers and some educational experts championed testing as a way to improve public education. Advocates of assessment-based reforms argued that well-developed assessments may be the most cost-effective approach to improving the quality of education (Popham, 1987) and that tests worth teaching to would encourage effective teaching, improve learning, ensure standards, and hold schools accountable (Wiggins, 1993). Other educational researchers and scholars offered more cautionary tales, warning against the dangers of using testing mandates as a means for ensuring high standards. They argued, among other things, that testing can narrow and trivialize curriculum, discourage higher order learning, and/or undermine teacher professionalism and expertise (Bracey, 1987; Madaus, 1988; Madaus & Kellaghan, 1993; Pearson & Valencia, 1987; Smith, 1991).

In the midst of these debates, a variety of different assessment formats continue to be developed and used in state-mandated testing programs, including multiple-choice tests of skills thought to be associated with writing, assessments of single samples of writing collected under impromptu, timed conditions, and multigenre collections of student writing collected in natural classroom settings. The various forms of these tests reflect competing—and evolving—conceptions of writing: Writing characterized as a set of discrete skills, as a set of cognitive and linguistic processes that can be demonstrated on demand, and more recently, as a collection of writing, involving some reflection on the production of various texts for different audiences, purposes and genres.

These different conceptions of the nature of writing ability have implications for the impact of assessment on curricula and teaching practices. Using testing as a means of achieving reform rests on the assumption that testing will influence curricular content and allocation of resources as well as teachers and students. And in fact, a growing body of literature shows that districts, as well as teachers, alter their curriculum to reflect the form and content of tests (Almasi, Afflerbach, Guthrie, & Schafer, 1995; Center for Education Policy, 2003; Corbett & Wilson, 1991; Door-Bremme & Herman, 1986; Firestone & Mayrowetz, 2000; Grant, 2000, 2001; Haertel, 1989; Haney, 1991; Kortez, Mitchell, Barron, & Keith, 1996; Koretz, Stecher, Klein, & McCaffrey, 1994a, 1994b; Koretz, Stecher, Klein, McCaffrey, &

Deibert, 1993; Madaus, 1988; Smith, 1991). More to the point for the current study, recent results of case studies conducted as part of a multi-year national study of state exit exam policies show that high school exit exams influence teachers' focus on ensuring that students are prepared to take and pass the exams (Center for Education Policy, 2005).

But although the impact of assessment on classroom practice seems assured, it is no means consistent (Cimbricz, 2002). A number of different variables can influence the impact of an assessment, including the level of stakes attached to the test as well as local, contextual variables such as teacher understandings of the test content and purpose, professional support at the district/county and school level for teachers to facilitate change, financial support for new materials, typical levels of student performance (low-performing schools respond differently from high-performing schools). Another important variable is the form of the assessment. To put it another way, different kinds of assessment appear to influence the curriculum in different ways.

### Multiple-Choice Tests

The validity for making important decisions based on multiple-choice tests of writing has been challenged because such tests require passive recognition of error and selection of best examples as opposed to active generation of text and on the grounds that they adversely affect the educational environment. Evidence suggests that large-scale, high-stakes multiple-choice tests affect writing curriculum in two ways: (a) actual writing begins to disappear from the curriculum and (b) the curriculum begins to take the form of the test. In an early study relevant to this issue, Smith (1991) observed that teachers shifted from a writing process curriculum to "worksheets covering grammar, capitalization, punctuation, and usage" when their district's test date neared because those activities were better aligned with the test. In a later study, Murphy (2003) found that when a direct assessment was changed to an indirect multiple-choice format, teachers spent less time teaching writing, more time teaching grammar and usage, and emphasized grammar and usage more heavily in their comments on student work. The concern, of course, was not that teachers were teaching more grammar, but that they were teaching less writing. Such findings are troubling, especially because the most common format used in large-scale accountability systems is the multiple-choice format (Quality Counts, 2002, cited in Hamilton, 2003).

### Timed, Impromptu Direct Tests of Writing.

Impromptu tests of writing have been challenged because they are not well aligned with contemporary views of effective writing instruction and because they fail to provide information about students' ability to manage other, more extended kinds of tasks (Camp, 1983; Elbow, 1997; Purves, 1995). Evidence of their consequences on teaching and learning is mixed. Some research suggests that teachers are likely to increase the time students spend writing when an assessment includes one or more writing components (Koretz et al., 1996; Koretz & Hamilton,

2003; Stecher, Barron, Kaganoff, & Goodwin, 1998). However, other studies demonstrate a negative effect on student attitudes (Ketter & Poole 2001; Loofbourrow, 1994), a narrowing effect on the curriculum (O'Neill, Murphy, Huot, & Williamson, 2004; Scherff & Piazza 2005; Wallace, 2002;), and a turn toward formulaic teaching (Hillocks, 2002; Johnson, Smagorinsky, Thompson, & Fry, 2003; Loofbourrow, 1994). Hillocks (2002) concluded that "when states provided for writing over more than one session . . . persuasive writing is not so likely to be formulaic" (p. 201).

Truncating the time allowed can impact validity issues in yet other ways. When time is a serious factor for most of the test population, or for particular groups within that population, bias is introduced, and any decisions made on such test results would have to be seriously limited. In such cases, what one learns from the results is not so much who is capable of performing the task, but who can perform it within the allotted time. Several studies support the view that increased time for writing may provide a more valid picture of English as a second language students' writing abilities (Cho 2003; Hilgers, 1992; Polio, Fleck & Leder, 1998). In a comparison of timed writing samples and portfolio collections, Simmons (1992) found that the weakest writers and writers from the poorest schools were disadvantaged by the timed test.

In addition to the time allowed for writing, the number and kinds of impromptu samples collected are also at issue in direct assessment. For example, single-sample assessments may not represent the variety of types of writing that examinees will be expected to employ in the context for which the assessment is designed. Research has demonstrated that students in college are assigned a wide variety of writing tasks, that they vary on a number of rhetorical and practical dimensions, and that their frequency varies across disciplines and graduate and undergraduate levels (e.g., Bridgeman & Carlson, 1983; Hale et al., 1996). Yet many placement tests sample a single type of writing, one that may not align in important ways with the kinds of tasks that students in college may be asked to perform.

Impromptu state writing tests also influence educators' morale and attitudes, although the cause of the impact is more likely the policies associated with the test rather than the test format per se. For instance, in Texas, where schools can be taken over by the state if they fail to improve, Hillocks (2002) reported that many schools "have a tension-filled environment" (p. 87). One principal explained that jobs are on the line if improvement is not made (Wallace, 2002). Similarly, Ketter and Poole (2001) reported in their qualitative study of three teachers and their students in a rural Maryland high school that teachers "were so focused on students' passing the test, they acted against their own beliefs" in teaching and designing a curriculum guide (p. 384). The students appeared to be disengaged from the writing instruction and "appeared to perceive the MWT (Maryland Writing Test) as a meaningless but threateningly difficult hurdle" (p. 383).

## Portfolio Assessments

As an assessment method, portfolios appear to address many of the concerns discussed above about the validity of using timed, impromptu, single-sample assessments to assess writing ability and make important decisions. Advocates argue that time and support for writing gives students a better chance to do their best, that good instruction can be mirrored in the assessment because writing can be treated as a recursive process and revisited for revision. Writing samples are collected under more natural and authentic conditions that can be directly linked to instruction. Portfolios also offer opportunities to broaden the assessment construct by sampling a range of genres and to engage students more directly in the assessment process in ways that give them some responsibility for evaluating their own learning.

The evidence on the impact of portfolio assessments on the educational environment, however, like other forms of assessment, is mixed. Several studies show positive effects. For example, nearly 75% of the principals interviewed in a study of Vermont's portfolio assessment program reported positive changes resulting from the program, changes such as "an increased emphasis on higher order thinking skills . . . lessened reliance on textbooks and worksheets; an increase in writing overall and more integration of writing with other subjects; more work in cooperative groups" (Koretz et al., 1994b, p. 31). Locally developed classroom and school portfolio assessments provide other evidence that teachers develop higher expectations for students and put more emphasis on individual growth and development when they use portfolios in the classroom (see, e.g., Graves & Sunstein, 1992; Jennings, 2002; Shepard, 1995). Still other studies show that participation in scoring sessions for curriculum-embedded assessments such as portfolios contributes to teachers' knowledge and expertise (Gearhart & Wolf, 1994; Shay, 1997; Sheingold, Heller, & Paulukonis, 1995; Storms, Sheingold, Nunez, & Heller, 1998).

Portfolio assessments, however, have not exerted uniformly positive effects. The policies that surround portfolios, like other forms of assessment, influence their impact. For instance, although Kentucky, with its statewide portfolio system, fared much better overall than other states in Hillock's (2002) study of the impact of statewide assessments on curriculum, other research has revealed several problems. When studying the perceptions of first-year students who completed the University of Kentucky's compulsory 12th-grade portfolios, Spaulding and Cummins (1998) found that "two-thirds of the students stated that compiling the portfolio was not a useful activity" (p. 191). Callahan (1999) reported that high school English teachers in her study saw the portfolios "primarily as a stressful administrative task" . . . "imposed from outside, introduced as a high-stakes accountability task, and embedded in a massive top down reform effort" (pp. 34-35). Callahan also reported that the pressure of the assessment situation encouraged "a form of dishonesty among both teachers and students" when portfolios were accepted, despite questions about the origin of some of the texts they contained. Taken together, the findings of these studies suggest that the policies that surround the assessment, as well as the form of the assessment itself, play a critical role in the ultimate impact of any assessment on curriculum and teachers.

In recent years, policymakers in the United States have been turning more frequently to using tests as levers for reforming curriculum, for defining, in effect, what should be taught. For instance, a survey of statewide assessment practices in 1997 showed that 46 of the 50 states had some kind of statewide assessment. The purpose most frequently identified by respondents (mentioned by 43 states) was the “improvement of instruction” (Roerber et al., 1997, cited in Mehrens, 1998). Our study was designed to contribute to the developing research base in this area. Although several studies have examined the impact of individual assessments, none have explicitly compared how teachers respond to different kinds of tests. In this study, our goal was to examine the impact of the form of the test on curriculum and teaching practices and to gather information about teachers’ views of the tests. More specifically, we asked the following questions:

1. How do different kinds of high-stakes, statewide assessments impact curriculum and instruction in writing in secondary schools?
2. What are teachers’ views of the impact of different kinds of high-stakes tests?

We selected three different states, California, Kentucky, and Georgia, as sites for our research. Each state had a different form of high school writing assessment.

To address the research questions just presented, we collected several sources of data during 2001 using surveys of high school teachers and focus group interviews with high school teachers. Because this study was part of a larger study investigating the impact of high school assessments on students’ preparation for the demands of college composition, we also collected focus group interviews with first-year writing instructors, focus group interviews with first-year college students, and state and institutional documents and policies. In this article, we focus primarily on the results from the survey of high school teachers as we address the two research questions.

### The State Tests

**A**s noted previously, each of the states had a state-wide mandated writing assessment but the form and stakes varied across the states (see Table 1).

### The Survey

**T**he survey was modeled on an earlier survey that had been developed by Cooper and Murphy (1989) for the Center for the Study of Writing. The earlier survey had been used to assess the impact of the California Assessment Program writing test on teaching and curriculum in that state. After updating and revising the survey to address the current assessment systems, we piloted it with teachers in each state to make sure it was aligned with the individual contexts. The survey contained both multiple-choice and open-ended items. We asked questions about various aspects of teachers’ experiences related to testing including influences on their teaching, their classroom practices, and faculty development opportunities.

Table 1: State Tests

	Test Format	Test Population	Stake	Background
<b>California</b> Stanford9/SAT9	Multiple choice; no writing required at all	Students in Grades 2-11 (along with other tests)	Schools: Scores used to calculate schools accountability performance (API) and track school performance; schools could be reconstituted based on API; no individual stake for students <b>Individual students:</b> required to pass for graduation from a public high school, results reported as part of the	Achievement test part of the Standardized Testing and Reporting (STAR) Program; California High School Exit Exam, which contains a writing prompt in development
<b>Georgia</b> Georgia High School Graduation Test (GHSCT) Writing Assessment	<b>Persuasive prompt:</b> on-demand writing scored in four domains—content/organization, style, conventions, sentence formation—by trained readers with a common rubric	High school juniors (similar tests given in earlier grades); retaken until passed		Initiated as part of the 1991 education reform; 1994 was first administration to count for students
<b>Kentucky</b> Kentucky Instructional Results Informational System (KIRIS)	<b>Portfolio:</b> produced as part of classroom instruction; contains seven selections including letter to the reviewer; personal narrative memoir, vignette, or essay; short story, poem, or play, and two expository pieces; scored by teachers with a rubric that focuses on six domains: purpose/audience; development/support; organization; sentences, language; correctness	Collected mid-March of senior year (portfolios also collected and scored in Grades 4 and 8)	<b>Schools:</b> required, high stakes for schools—financial rewards and sanctions, no state-mandated consequence to student; results part of the schools' accountability index	Initiated as part of 1990 Kentucky Education Reform Act; implemented in AY 1991-1992

During Spring 2001, surveys were sent to the principals of the individual university's feeder high schools. Principals were given specific directions for randomly selecting one English teacher from their school to complete the survey. After completing the survey anonymously, teachers returned it directly to us. We sent reminder postcards to schools who had not replied 4-6 weeks after the surveys were mailed. Table 2 details the distribution and return rates of the surveys. Note that the number of surveys sent depended on the number of feeder high schools at one particular state university so that the actual numbers sent varied across states. In California the surveys were sent to 770 high schools, 34% of the 2,269 high schools in the state. The returned surveys (419) represented 18% of the high schools in the state. In Georgia, 434 high schools were sent surveys, representing 60% of the 729 secondary schools in the state. The returned surveys (129) represented 18% of Georgia high schools. The Kentucky survey was sent to 196 secondary schools, representing 77% of the 254 high schools in the state. The returned surveys (63) represented 25% of Kentucky's high schools.

We expected differences across the three states in several areas. Because the form of the statewide test was different in each of the three states, we expected differences in the kinds of writing assignments that teachers typically gave and in the conditions accompanying the assignment (e.g., length of writing assigned, or time for writing). In California, the "writing" test was a multiple-choice test. In Georgia, the test asked for a single sample of writing produced under controlled, timed conditions. The focus there was on producing writing that fits the form assessed by the state, a persuasive essay, in timed conditions. In Kentucky, the major writing "test" asked for portfolios of writing produced under normal classroom conditions with the focus on "writing for authentic audiences" and "situations" (Kentucky Department of Education, 1999). According to the 1999 Portfolio Handbook, students at grades 4, 7, and 12 were required to collect pieces in three categories: personal expressive writing, literary writing, and transactive writing. Personal expressive writing included personal narratives (works that recount a single incident in the student's life), memoirs (works focused on the significance of the relationship between the student and another person), and personal essays (an essay that focuses on a central idea about the writer or the writer's life). Literary writing included poems (compositions in verse), short stories (plot, setting, character, theme, point of view), and scripts/plays. Transactive writing included "writing written from the perspective of an informed writer to a less

**Table 2: Survey Return Rate**

	<b>Sent</b>	<b>Returned</b>	<b>Response Rate</b>
California	770	419	54%
Georgia	434	129	30%
Kentucky	196	63	32%



informed reader...; writing [that] is produced to get something done" (p. 14). Kentucky also included an impromptu essay as part of the KIRIS testing system but that was not included as part of the portfolio.

Because the educational systems in the three states were very different, we also expected to see differences in other areas. In addition to questions about the teaching of writing, the survey included questions about the kinds of professional, programmatic work teachers engaged in at school, and the amounts of time they typically spent on different language arts activities. The survey also included a specific question about the influence of tests on teaching. One question concerned teachers' perceptions of the degree to which a variety of factors such as in-service activities, district and state curriculum guidelines, new textbooks, professional publications, and the statewide test influenced their teaching.

Because of the number of items on the questionnaire and the number of open-ended items, purely quantitative items were selected that seemed to provide the basis for demonstrating the similarities and differences between the approaches used by teachers in response to the differing types of tests their students were facing. Subsequently, the quantitative items were examined using SPSS-PC. First, item frequencies were used to discard those items that had too many missing responses. Subsequently, 11 items were examined using cross-tabulations. The cell frequencies were tested for statistical significance using chi-square analysis.

Although the number of responses to each item provided sufficient statistical power to use a stringent level for each test, the nature of the data themselves suggested that a less stringent level be employed. All tests were run with the expectation that a probability value of less than or equal to .05 would exclude variables that would be of no interest while preserving some variables that might be important in future research.

There are two interpretive issues with cross tabulations of the data. The first is the number of cells in some tables with frequencies less than 5. The second was the differences in the number of responses to each item based on the return from each state. Table 2 shows that the number of teachers in California asked to complete the questionnaire was considerably larger than the numbers of teachers in Georgia and Kentucky. Furthermore, Table 2 shows that nearly twice the percentage of teachers in California returned the questionnaire than did teachers in Georgia and Kentucky. The return rates from Georgia and Kentucky are typical for a mailed survey, although they should be considered very good for a questionnaire as complex and lengthy as used in this study. Under any circumstances, the return rate from California is unusually high. One interpretation of these differences is that teachers in California were facing the movement from the type of assessment used in Georgia to the SAT9/STAR system and wanted to respond. Because the other two systems of assessment had been around for some time, teachers may have been less concerned.

The responses patterns in the cross tabulation tables suggest some consistent differences among teachers that are consistent with the hypothesis that the three different assessment systems influenced teachers' approaches to instruction depending on the kind of testing their students were facing. Results also indicated differences across the three states in teachers' perceptions of the degree the tests influ-

enced curriculums. Responses to an open-ended question asking teachers for general comments about the impact of the tests provided a further basis confirming the quantitative results while they also provided a basis for meaningful interpretation of the basis for the differences observed.

## Results and Discussion

### The Influence of Tests on Teaching

Although most teachers in Kentucky, and many in Georgia, indicated that the test “strongly influenced” their curricula, few teachers in California acknowledged such influence (Table 3). Seventy-six percent of the teachers in Kentucky and 46% of the teachers in Georgia indicated that their statewide test “strongly influenced” their curriculum, but only 15% did so in California.

Table 3: The Influence of Statewide Tests on Teaching

	CA		GA		KY	
Strongly influenced	60	15%	58	46%	48	76%
Somewhat influenced	141	34%	37	29%	7	11%
Slightly influenced	113	27%	9	7%	1	2%
Did not influence	99	24%	23	18%	7	11%
	413	100%	127	100%	63	100%

$df = 6, \chi^2 = 141.12, p < .005$

Questions on the survey were designed to gather information about the impact of the different assessments on the teaching of writing. There were no significant differences in the ways teachers in the three states responded to questions about the *amount of time* they spent on various language arts activities (vocabulary, grammar, literature, writing, oral communication). However, results of the survey indicated significant differences in the *kinds of writing assignments* given to students across the three states. These included differences in the *types of writing assigned*, in the *typical length* of writing assignments, and in the *time allowed for writing*.

*Types of Writing.* One of the questions asked teachers which type of writing they assigned most frequently. Teachers selected from a menu of genres but were allowed to include types not listed under by selecting “other” and writing in the genre. Table 4 reports on the responses to this question. The percentages suggest that students in English classes in Kentucky and Georgia, are exposed to a broader variety of writing types than students in California. It is easier to see this trend when the categories of types are collapsed as they are in Table 5. In all three states, most teachers indicated they assigned response to literature most frequently. But in California, 72% of the teachers reported “response to literature” as the writing they assign most frequently and 28% reported some other type of writing. In con-

**Table 4: Most Frequently Assigned Writing**

	CA		GA		KY	
Short story	3	1%	0	0%	11	18%
Summary	26	7%	17	14%	1	2%
Argument	13	3%	5	4%	1	2%
Response to literature	287	72%	63	50%	32	52%
Problem/solution	1	0%	4	3%	0	0%
Persuasive	10	3%	12	10%	5	8%
Autobiographical narrative	13	3%	2	2%	8	13%
Reflective essay	18	5%	13	10%	3	5%
Observational report	1	0%	0	0%	0	0%
Reading report	5	1%	2	2%	0	0%
Other	23	6%	7	6%	0	0%
	400	100%	125	100%	61	100%

**Table 5: Most Frequently Assigned Writing**

	CA		GA		KY	
Response to literature	287	72%	63	50%	32	52%
Other	113	28%	62	50%	29	48%

df = 2,  $\chi^2 = 23.99, p < .005$

trast, 52% of the teachers in Kentucky reported “response to literature” and 48% some other type of writing. In Georgia, there was an even 50/50 split.

In addition to asking teachers what type of writing they assigned most frequently, we asked teachers *how frequently* they assigned particular types of writing. Significant differences were found in how often teachers indicated they assigned five types of writing: response to literature, problem solution, short story, persuasive writing, and autobiographical narrative (Table 6).

**Table 6: Response to Literature: Assignment Frequency**

	CA		GA		KY	
Never	2	0%	6	5%	1	2%
Once	19	5%	12	9%	2	3%
two or three times	104	25%	34	27%	14	22%
At Least 4 Times	290	70%	76	59%	46	73%
	415	100%	128	100%	63	100%

df = 6,  $\chi^2 = 18.62, p < .005$

Consistent with their responses about their most frequent assignment, teachers in California and Kentucky reported that they assigned response to literature more often than teachers in Georgia. Seventy percent of the teachers in California and 73% of the teachers in Kentucky reported that they assigned response to literature at least four times during the school year, compared with 59% of the teachers in Georgia.

The differences in responses across the three states may be due in part to differences in the form of the respective tests. The Georgia High School Graduation Test Writing Assessment (GHSGT) is a persuasive prompt that does *not* rely on a response to literature. Literature is a significant part of the GHSGT assessment of English/language arts (47–49%), but the test is in a multiple-choice format. Differences among the states were also found in teachers responses to questions about autobiographical narrative. Although in general, teachers in the three states reported assigning autobiographical narrative less frequently than some of the other types of writing, proportionately fewer teachers in Georgia assigned it more than once than in California and Kentucky. In Georgia, 29% of the teachers reported assigning it more than once, compared with 39% in California and 38% in Kentucky (Table 7).

**Table 7. Autobiographical Narrative: Assignment Frequency**

	CA		GA		KY	
Never	68	16%	31	24%	6	10%
Once	183	44%	59	46%	33	52%
Two or three times	115	28%	34	27%	17	27%
At least four times	47	11%	3	2%	7	11%
	413	100%	127	100%	63	100%
df = 6, $\chi^2 = 15.45, p < .025$						

Recall that Georgia's writing assessment is a persuasive prompt. The test does *not* call for autobiographical narrative writing. Teachers in Georgia are *not* assigning response to literature and autobiographical narrative as frequently as teachers in other states, but they did report assigning problem-solution papers more often than teachers in California or Kentucky. Thirty-six percent of the teachers in Georgia reported assigning this kind of writing more than once, compared with 23% of the teachers in California and 25% of the teachers in Kentucky (Table 8). Proportionally more teachers in Georgia also reported assigning persuasive essays more than once: 60% of the teachers in Georgia reported assigning persuasive essays more than once compared with 41% in California and 44% in Kentucky (Table 9). In Georgia, the writing assessment is a persuasive task and students are allowed 90 minutes for writing. The Department of Education (2001) test guide explains the task as follows:

**Table 8. Problem Solution: Assignment Frequency**

	CA		GA		KY	
Never	154	38%	40	32%	19	31%
Once	158	39%	40	32%	28	45%
Two or three times	66	16%	39	31%	14	23%
At least four times	30	7%	6	5%	1	2%
	408	100%	125	100%	62	100%

df = 6,  $\chi^2 = 17.86, p < .01$

**Table 9. Persuasive Essay: Assignment Frequency**

	CA		GA		KY	
Never	58	14%	12	9%	5	8%
Once	185	45%	39	31%	29	48%
Two or three times	109	27%	55	43%	21	34%
At least four times	59	14%	21	17%	6	10%
	411	100%	127	100%	61	100%

df = 6,  $\chi^2 = 18.27, p < .01$

In persuasion, the writer assumes a position on an issue and uses language to influence the reader. The purpose is to express a writer's opinion on a subject either explicitly or implicitly. Through the support provided, the writer presents a convincing point of view. Support for the writer's position should include evidence such as logical appeals, emotional appeals, facts (which may or may not be accurate), personal experiences, extended narratives, etc. (p. 2)

Both problem-solution and persuasive writing call on similar skills. For instance, when writing problem-solution papers, writers need to convince readers of the value of the solution they propose. Taken together, these results support the interpretation that the form of the test influences what is taught. Teachers in Georgia tended to assign writing that matched the form of their state test more often than teachers in other states. Other results also support the idea that the form of the test influences what teachers teach. For example, teachers in Kentucky reported assigning short story writing more often than teachers in California or Kentucky. Seventy-nine percent of the teachers in Kentucky reported assigning short story writing at least once, compared 63% in California and 58% in Georgia (Table 10). Recall that the Kentucky portfolio requires literary as well as samples of personal and transactional writing. It is likely that short stories are assigned more frequently in Kentucky than in the other states because they fulfill one of the portfolio requirements.

Table 10. Short Story: Assignment Frequency

	CA		GA		KY	
Never	150	36%	52	41%	13	21%
Once	167	40%	46	36%	40	63%
2 or 3 Times	79	19%	22	17%	10	16%
At Least 4 Times	18	4%	7	6%	0	0%
	414	100%	127	100%	63	100%

df = 6,  $\chi^2 = 16.89, p < .01$

The teachers' responses to the question about how often they assigned short story writing are congruent with their responses to the question about the type of writing they assigned most frequently. Of the teachers in Kentucky, 18% reported that short story (a literary form) is the writing they assign most frequently, but none of the teachers in Georgia, and only 1% of the teachers in California assigned this type of writing most frequently.

In summary, teachers in California assign response to literature writing most frequently, perhaps because that form of writing aligns most closely with the typical content of English courses—the study of literature. Teachers in Kentucky assign response to literature writing most frequently, but they also assign a variety of other kinds of writing, including short story writing and autobiographical narrative, perhaps because in Kentucky the assessment system calls on students to choose pieces that address a variety of audiences for a variety of purposes, using a variety of forms. More specifically, the portfolio should include a letter to the portfolio reader as well as personal, literary, and transactional writing. These breadth requirements might, in part, explain why the responses of teachers in Kentucky differ from the responses of teachers in California. Teachers in Georgia also assign response to literature writing frequently, but they also frequently assign other types of writing that align with the type of writing on the state test. Taken together, these results suggest that students in the three states are getting very different sorts of preparation for college from their English teachers, in terms of the types of writing they are being prepared to write.

*Features of Writing Assignments.* Teachers were asked how often they assigned writing of different lengths. Interesting differences appeared across the states. More teachers in Georgia than in California or Kentucky reported assigning papers of 75 words some or most of the time, a pattern that suggests some adaptation to the timed nature of the Georgia test. Recall that students in Georgia have 90 minutes to write in the test. Of the teachers in Georgia, 47% reported assigning papers of 75 words some or most of the time compared with 39% of the teachers in California and 35% of the teachers in Kentucky (Table 11).

**Table 11. Length of Writing Assignments: 75 Words**

	CA		GA		KY	
Most of the time	39	11%	22	19%	3	6%
Some of the time	92	27%	32	28%	15	29%
Only rarely	108	31%	39	34%	22	42%
Never	105	31%	21	18%	12	23%
	344	100%	114	100%	52	100%

df = 6,  $\chi^2 = 13.20, p < .05$

The pattern across the states was nearly reversed when teachers were asked about how frequently they assigned papers of 450 words. More teachers in California and Kentucky than in Georgia reported assigning papers of 450 words some or most of the time. Of the teachers in Georgia, 44% reported assigning papers of 450 words some or most of the time, compared with 63% of the teachers in California and 58% of the teachers in Kentucky (Table 12).

**Table 12: Length of Writing Assignments: 450 Words**

	CA		GA		KY	
Most of the time	126	33%	26	22%	11	20%
Some of the time	115	30%	25	22%	21	38%
Only rarely	101	27%	41	35%	14	25%
Never	37	10%	24	21%	9	16%
	379	100%	116	100%	55	100%

df = 6,  $\chi^2 = 20.95, p < .005$

There were no significant differences between the states on questions about papers of 150, 300, or 750 words or more. Although not statistically significant, there was some indication that teachers in Georgia require fewer drafts than teachers in California or Kentucky. Of the teachers in Georgia, 39% typically required only one draft, compared with 29% of the teachers in California and 26% of the teachers in Kentucky (Table 13).

The pattern of teachers' responses to a question about the time allowed for writing was consistent with their reports about the number of required drafts. More teachers in California and Kentucky, than in Georgia, reported allowing 2 or more days before the final draft was due. In California, 84% of the teachers allowed 2 or more days, in Kentucky, 71%, and in Georgia, 48% (Table 14). Proportionately

Table 13. Number of Drafts Required

	CA		GA		KY	
One	116	29%	48	39%	16	26%
Two	196	48%	60	49%	32	52%
Three	85	21%	11	9%	12	20%
Four	6	1%	3	2%	1	2%
	403	100%	122	100%	61	100%

$df = 6, \chi^2 = 12, p < .10$

Table 14. Time Allowed for Writing

	CA		GA		KY	
Same day	32	8%	29	23%	8	13%
Next day	33	8%	37	29%	10	16%
Two days	55	13%	27	21%	11	17%
Three or more days	289	71%	34	27%	34	54%
	409	100%	127	100%	63	100%

$df = 6, \chi^2 = 87.39, p < .005$

more teachers in Georgia required the writing assignment to be turned in on the same day that it was assigned (23% compared with 8% in California and 13% in Kentucky).

To summarize, the results reported here support the interpretation that the conditions for writing associated with a test will influence how teachers teach. In Georgia, teachers were more likely than teachers in California and Kentucky to assign single-draft, short pieces of writing and to require that the writing be turned in the same day. This is consistent with the conditions for writing of the 90-minute Georgia state test. In California and Kentucky in contrast, teachers were more likely to allow 3 or more days for writing, to require three or more drafts, and to require assignments of greater length than did teachers in Georgia.

Although it seems likely that the form of the writing assessment influenced teachers in these states, other factors no doubt played a role and the survey asked about some of these. Teachers were asked about professional development activities. For example, significantly more teachers in Kentucky and California than in Georgia reported that they had participated in in-service on the teaching of writing during the previous year (Table 15). In California, 72% participated and in Kentucky 82% participated, compared with only 53% in Georgia. Although the



**Table 15. Participation in Inservice on the Teaching of Writing**

	CA		GA		KY	
Participated	295	72%	68	53%	51	82%
Did not participate	116	28%	60	47%	11	18%
	411	100%	128	100%	62	100%

df = 2,  $\chi^2 = 21.61, p < .005$

survey only asked about in-service activity for the most recent year, teachers may have been involved in professional development activities earlier or outside of in-service programs. For example, the Kentucky Writing Program developed a wide ranging and sustained approach to professional development, including regional writing consultants, writing project summer institutes, grants to schools, and a long list of televised professional development programs on topics such as writing in the science classroom, poetry, and high school journalism, as part of the Kentucky Educational Reform Act of 1990 (Kentucky Department of Education, 1999). In fact, professional development opportunities were a cornerstone of the Kentucky system as evidenced by the Kentucky Writing Program's support of eight National Writing Project affiliates and its many other projects. California, where the National Writing Project started in 1974, has a strong history of professional development in the teaching of writing with 16 sites of the California Writing Project. It is likely then that in both California and Kentucky, many teachers are knowledgeable about and practice process approaches to the teaching of writing, approaches that typically involve more than one draft and in which writing is completed over a period of days instead of minutes. Georgia, on the other hand, also has several sites of the National Writing Project, but many were only founded in the mid-to-late 1990s and have not enjoyed widespread support as have those in Kentucky and California.

Open-ended comments in response to questions about the test helped to interpret the teachers' responses about the influence of the test. Again, there was a marked difference in the pattern of comments from the teachers in different states as evinced in Table 16, which summarizes teachers open-ended comments. Comments were coded positive, negative, and neutral as indicated by the tone and content of the teachers' response. Comments that merely reported information were coded as neutral; comments that conveyed teachers' approval were coded as positive; and comments that conveyed stress, coercion, fear, or disapproval were coded as negative.

Table 16. Open-Ended Comments on the State Writing Assessment

	CA		GA		KY	
Positive	18	13%	8	16%	16	41%
Negative	91	64%	12	24%	4	10%
Neutral	33	23%	30	60%	19	50%
Totals	142	100%	50	100%	39	101% <sup>a</sup>

<sup>a</sup>Percents were rounded to the nearest whole number.

### Responses in California

In California, where few teachers acknowledged that the test strongly influenced their teaching, the open-ended responses indicated that many teachers were dissatisfied with the state test. At the end of the California survey, teachers were asked: “Has the SAT9/STAR assessment had any impact on your school that we haven’t already asked about?” Responding to this item were 142 teachers. Eighteen of the comments were positive in tone (13%), but as the examples illustrate, the comments tended to focus on successful school performance or rewards received instead of the quality of the test or its impact on curriculum and students.

Positive comments included the following:

- We improved enough to get the bonuses!
- We were one of six schools in LAUSD to meet our API goal—this has brought positive attention.
- SAT 9/Star assessment is a big focus at [name deleted]. We are the number 1 most improved high school in LA County. This has really helped to motivate both teachers and students.
- We’re proud of our scores.

Thirty-three (23%) of the comments were about changes made or actions taken and were more or less neutral in tone:

- Scheduling changes
- More math classes and class size reduction.
- We arranged for more tutoring.

However, 91 (64%) of the comments were decidedly negative in tone and content. The negative comments were about the test, the test policy, the time it took from instruction, and the negative impact it was having on morale, the curriculum, the school in general, and students in particular. The following comments are representative:

- We are losing 3 days of instructional time. Students are very stressed.
- It has put my English colleagues under such pressure to "perform" on standardized tests that measure things that we do not believe should be the measure of an accomplished student. These tests do not measure true composition skills nor creative thinking. SHAME ON SACRAMENTO!
- Huge time demand. Shortens our classes for 6 days.
- We do not buy breakdown scores from the testing service. To be quite honest our district/teachers assessed several tests 4 years ago and ranked the STAR test dead last. . . . We don't feel that it measures the skills laid out in the State Frameworks.
- Oh yes, it's caused panic. Soon we'll be teaching directly to the test I'm afraid. The ESL [English as a second language] students I teach sit for 2-hour periods and stare at their papers: a form of child abuse.
- Fear.
- Teachers feel sometimes that they are being judged on superficial and sometimes irrelevant matters, things students haven't been taught and shouldn't be taught.
- Our school raised its APO by 14 points last year. Focus has been on maintaining this mundane surge of our students ability to work with the letters A–E. Instead of promoting higher learning, critical thinking, and preparation for the real world, students are focused to be good test-takers.

Survey responses to this open-ended question suggest that many of the California teachers held negative opinions about the SAT9/STAR program. This dissatisfaction with the test and the policies associated with it may explain, in part, why few California teachers were willing to say that the test strongly influenced their curriculum.

#### Georgia Responses

Like teachers in California, only a few teachers in Georgia commented favorably about the state writing assessment. However, far fewer Georgia teachers indicated a negative response to the influence of the writing test than those in California. Of the 129 Georgia teachers who completed the survey, 50 (39%) chose to respond to an open-ended question "What impact if any, has the statewide GHSGT Writing Assessment had on your school that we have not already asked about?" Of these comments, 16% (8) were positive, 24% (12) were negative, and 60% (30) were neutral. The positive and neutral comments, which mentioned changes or influences attributed to the GHSGT writing assessment, may explain why many teachers in Georgia reported that the test had a strong influence on teaching at their school. In both of these categories, teachers commented on the increased attention to writing that the test there had generated. However, like teachers in California and Kentucky, some teachers in Georgia made negative comments about the increased pressure and stress on teachers created by

the test, with some mentioning specific pedagogical impacts (such as a focus on form or minimal skills).

Positive comments included the following:

- The GHSGT has made us decide to emphasize writing across the curriculum as our school focus for school improvement.
- It has made our department more unified as we work toward a common goal.
- More schoolwide concentration on writing and improvement of writing skills.
- We spend more time teaching writing and helping students proofread and revise their work.
- The GHSGT has forced us to provide writing instruction for all levels. For too many years, non-college prep students were allowed to go through with multiple-choice/true-false evaluations, whereas college-prep students wrote essays. We are now doing more of what we should have been doing.

But unlike comments made by California and Kentucky teachers, most of the comments by Georgia teachers were neutral in tone—that is they did not signal approval or disapproval of the changes but merely reported them. Comments in this category reported on programmatic changes, such as remediation efforts, as well as pedagogical requirements, such as teaching the five-paragraph essay.

- Teachers in Grades 9–11 do specific writing assignments for the GHSGT and grade them by the state rubric.
- Scheduling of tech prep students for English during the fall when the GHSGT in writing is given. Workshop/review sessions with students not scheduled for English during test time.
- Teachers focus more on weak areas of writing, especially focusing on the process and development of ideas and coherence.
- We remediate students who need help if they do not pass.
- 99% pass—no impact.
- Counselors meet with every junior in classroom sessions to practice for the writing assessment.
- We stop classes to review for the GHSGT in a round-robin fashion for a solid week. Tutorial programs are ongoing after school hours. Review sessions are held in the summer to "review" those who've failed parts of the test. Much emphasis has been placed on writing a five-paragraph essay in all grades.

Negative comments included the following:

- Intense pressure to prepare students for the test. A great amount of stress among students before test and after results are given out. We now “teach to the test”.

- Places more emphasis on minimal skills rather than higher level performance.
- It forces us to practice the five-paragraph essay in order to better prepare our students for the test.
- They stress the five-paragraph essay which is not real-life writing. The emphasis is on this to the point that practically every other writing form is ignored.

### Kentucky Responses

The picture in Kentucky was somewhat different. Responses to two questions (Questions #22 and 23) were coded on the Kentucky survey because both questions invited a response about the impact of the statewide assessment. Seventeen teachers responded to Question 22 and 19 to Question 23. Two teachers wrote both positive and negative comments in their responses to the same question. As a result, there were 39 responses. Of these, 16 (41%) were positive, 19 (50%) were negative, and 4 (10%) were neutral. Recall that most Kentucky teachers said the test had a strong influence on the curriculum. This is what one might expect given that one component of the assessment involved the construction of a portfolio. The portfolio guidelines dictated what would be taught and the contents of the portfolios were in turn the products of the curriculum. However, other components of the statewide test took different forms, so the teachers' responses were mixed. Although Kentucky teachers said they were spending too much time and energy on testing, many also said that students were learning to write better. The open-ended comments suggested that some, although not all teachers in Kentucky, valued the portfolio component of the assessment as well as the impact it had on the writing curriculum.

Positive comments were as follows:

- Our teachers would not be teaching nearly as much as they are asked to by the state writing assessment.
- Our students write more because of the state writing assessment.
- Portfolio development has also made the level of writing improve. Students are more aware of process, proofreading—all phases of writing.
- Portfolios are the best way for students to grow as writers and be able to see their growth.
- Encourages teachers to use more and different types of genres.

Some negative comments included the following:

- Vast amounts of time are spent on practice testing and actual testing. Obviously this takes away from class time.
- It has helped breed indifference, anxiety, and anger (much of this repressed).
- It makes testing too important.

- State assessment: very stressful on both students and teachers.
- Portfolios are too much work for little direct benefit to the student's learning.

The overall picture these comments give is that teachers in all three states have concerns about testing, the time it takes away from instruction, and the negative impact it can have on morale. But teachers in the three states had different opinions about the impact on curriculum. In California, where the form of the test was *multiple-choice*, teachers made negative comments about the test in this regard, commenting on the emphasis on what one teacher characterized as “superficial and sometimes irrelevant matters, things students haven’t been taught and shouldn’t be taught.” In Georgia, where the form of the test encouraged a *single-sample five-paragraph essay*, teachers acknowledged that there was more teaching of writing, yet they expressed concern about the emphasis on the five-paragraph essay format. In Kentucky, in contrast, several teachers praised the *portfolio* format, although at least one teacher questioned its value in relation to the time it took to create one.

### Implication and Conclusions

The results of the survey support the conclusions of previous research that indicate tests will influence curriculum and teachers (Center for Education Policy 2003; Corbett & Wilson, 1991; Door-Bremme & Herman, 1986; Firestone & Mayrowetz, 2000; Grant, 2000, 2001; Haertel, 1989; Haney, 1991; Koretz et al., 1993, 1994a, 1994b, 1996; Madaus, 1988; Pedulla et al., 2003; Smith, 1991). More specifically, the survey results contribute to the growing body of research that indicate the forms of writing tests influence *what* teachers teach and *how* they teach it (Almasi et al., 1995; Callahan, 1997, 1999; Hillocks, Johnson et al., 2002, 2003; Ketter & Poole, 2001). However, the influence on instruction was complex; significant differences in the types of assignments, the length of assignments, number of drafts, and amount of time allowed for the assignment existed across the three states studied, each of which had different forms of writing assessments. Teacher responses indicated that these differences were attributable, at least in part, to the differences in the form of the test. In Georgia and Kentucky, teachers assigned a broader range of types of writing than teachers in California, where the most frequently assigned writing type was response to literature. In Georgia, teacher comments reinforced the other data that showed that the form of the test—a 90-minute impromptu persuasive essay—influenced writing instruction in their classroom and in their school much like the Illinois and Texas teachers in Hillocks’ (2002) study. In Kentucky, teachers reported more variety in the genres and more multidraft writing than in Georgia. In California, where there was no writing for teachers to model and no state rubric at the time, one interpretation is that teachers simply assigned the type of writing that aligned most closely with their subject matter—literature. In Georgia, the state test prompted teachers to teach the type of writing on the state test (persuasion), so that writing type competed with what we might call the more typical English assignment—response to literature. In Kentucky, the variety of genres included in the portfolio required teachers to teach

more than response to literature. The writing tests may also have prompted English teachers in Georgia and Kentucky to provide students with a better sense of how writing varies with audience and purpose than teachers in California.

Our results, like those reported by Hillocks (2002), Callahan (1997, 1999), Ketter and Poole (2001) and others, indicate that not only does the form of high-stakes writing tests influence instruction, it impacts teacher morale and attitudes. Because high-stakes tests can make teachers feel that they must compromise what they believe is best practice, they can feel pressured to emphasize preparation for the test at the expense of other material, forms, or processes. Even in Kentucky, where portfolios that promoted writing process and rhetorical approaches to teaching were used, some teachers noted that the emphasis on assessment results had detrimental effects.

The implications of these results are not just about instructional materials and methods, but also about students' preparation in writing. Students in the classrooms in these states will have had very different experiences learning to write based on the feedback we received from their teachers. Although there is always variety in individual classrooms, the patterns that emerged from the data suggest that students in California got more experience in writing multidraft responses to literature; Georgia students got more practice in producing short, single-draft persuasive essays; and Kentucky students wrote in more genres with attention to concepts of audience and purpose. In other words, students in these states will leave high school having had very different experiences in writing—and by implication—will have developed different definitions of writing and different understandings of what it means to write. How will these differences impact their readiness for the demands they encounter in college? Although there is not a standard college composition curriculum, the National Council of Writing Program Administrators (NCWPA) have endorsed an Outcomes Statement for first-year college composition, which encourage a rhetorical framework and a process approach among other things. Many college writing programs have articulated their own outcomes or learning aims that share many of the same goals as NCWPA. For example, at the University of California (UC) Davis, the course goals for first-year composition, English 101, demand a rhetorical approach that does not rely on response to literature:

- To improve students' analytical skills in reading and writing; to explore through readings and writing assignments issues and problems both unique and common to particular disciplines and professions.
- To help students understand the rhetorical context of all writing, both academic and professional; to provide instruction in writing for different audiences and purposes.
- To give students an opportunity to explore a variety of nonfiction writing forms including narrative, analysis, explanation, argument, and critique.
- To adapt academic writing skills and modes of expression to the kinds of writing tasks that different professions and careers demand.

- To explore ways in which in variety of different research strategies (including, e.g., field studies, surveys, interviews, and observations) can inform academic, literary, and professional nonfiction writing.
- To help students develop a clear, lively, and forceful prose style, and to adapt that style to different writing situations and audiences.

These goals, as stated on the UC Davis University Writing Program Web site, are in line with what many first-year programs have articulated. It is true that students will be learning these concepts in the course, but exposure and practice to them in high school can only facilitate their success once they arrive in the English 101 classroom. Limited experience, as indicated by the California teachers' report that the most frequently assigned writing task is response to literature, may make students' transition to the demands of college more difficult. Students from Georgia may not be as prepared for longer essays (750–1,250 words) often required in college writing classrooms, and because of the isolated context of the exam, may not have had much practice using writing in conjunction with reading. On the other hand, students in Kentucky may be better prepared for this curriculum, according to their teachers' responses. Because the Kentucky portfolio requires students to produce a variety of texts, writing for a variety of audiences and purposes, these students may at least have been exposed to the concepts and have a rhetorical foundation that the college course can build on. What teachers in our study said about teaching writing and the influence of the test in their own teaching and in their school supports other recent research (Callahan, 1997, 1999; Hillocks, 2002; Johnson et al., 2003; Ketter & Poole, 2001) that indicates that the type of writing assessment mandated by the state will influence the writing instruction that high school students experience.

Although our findings contribute to the research on writing assessments and high-stakes testing, we realize that there are some limitations to our study. One limitation that we faced, although not unique to us, was the changing landscape of educational assessments, specifically in terms of the writing assessments. During the surrounding years of our study, California was in the process of piloting and implementing a new writing assessment, the California High School Exit Exam. In 1998, Kentucky passed the Commonwealth Accountability Testing System that revised the 1990 educational reform and accountability system. The more recent passing of No Child Left Behind promises more changes as tests develop and implement the required assessment programs. For example, during 2003 and 2004 in Georgia, the State Department of Education revised the state's curricular goals, the Georgia Performance Goals, that includes a writing across the curriculum component (Georgia Department of Education, 2003). The Georgia Department of Education also reported that it was "moving forward with a federally approved plan to dramatically revamp, strengthen, and enhance [the] existing high school graduation test." Continual changes in assessments, stakes, or policies, make it difficult for teachers and researchers to be sure not only that they are informed on testing policies and procedures but also that they can discern how particular tests or policies are impacting the classroom. Constant change also makes longitudinal research difficult because of the variability from year to year. Although these are



limits of our work that we recognize, they also reinforce our belief that more ongoing research is needed to ascertain the consequences of high-stakes writing tests on teaching and learning. Only through this type of research can policymakers and educators make more informed decisions about how to revise assessment systems and improve education for all students. This type of research is part of the “ongoing evaluation of intended and unintended effects of high stakes testing” that the American Educational Research Association (2000) considers essential, and which is a component of the validation process that high-stakes testing requires.

### Acknowledgment

This research was supported in part by grants from the National Council of Teachers of English Research Foundation and the University of California Research Center in Fresno, CA.

### References

- Almasi, J., Afflerbach, P., Guthrie, J., & Schafer, W. (1995). *Effects of a statewide performance assessment program on classroom instructional practice in literacy* (Reading Research Report no. 32). College Park, MD: National Reading Research Center, University of Maryland.
- American Educational Research Association. (2000). *AERA position statement concerning high-stakes testing in PreK-12 education*. Washington, DC: Author. Retrieved July 22, 2004 from <http://www.aera.net/about/policy/stakes.htm>
- Bracey, G. (1987). Measurement-driven instruction: Catchy phrase, dangerous practice. *Phi Delta Kappan*, 68, 683-686.
- Bridgeman, B., & Carlson, S. (1983). *Survey of academic writing tasks required of graduate and undergraduate foreign students* (TOEFL Research Report no. 15; ETS Research Report no. 83-18). Princeton, NJ: Educational Testing Service.
- Callahan, S. (1997). Tests worth taking?: Using portfolios for accountability in Kentucky. *Research in the Teaching of English*, 31, 295-336.
- Callahan, S. (1999). All done with best of intentions: One Kentucky high school after six years of state portfolio tests. *Assessing Writing*, 6(1), 5-40.
- Camp, R. (1983, April). *Direct assessment at ETS: What we know and what we need to know*. Paper presented at the National Council on Educational Measurement, Montreal, Canada.
- Center for Education Policy. (2003). *State high school exit exams put to the test*. Washington, DC: Author.
- Center for Education Policy (2005). *How have high school exit exams changed our schools?* Washington, DC: Author.
- Cho, Y. (2003). Assessing writing: Are we bound by only one method? *Assessing Writing*, 8(3), 165-191.
- Cimbricz, S. (2002). State-mandated testing and teachers' beliefs and practice. *Education Policy Analysis Archives*, 10(2). Retrieved May 28, 2004 from <http://epaa.asu.edu/epaa/v10n2.html>.
- Cooper, C., & Murphy, S. (1989). *What we know about teachers' perceptions of CAP and it's impact on instruction*. Paper presented at the California Assessment Program “Beyond the Bubble” Conference, Anaheim, CA.
- Corbett, H. D., & Wilson, B. L. (1991). *Testing, reform, and rebellion*. Norwood, NJ: Ablex.
- Door-Bremme, D., & Herman, J. (1986). *Assessing student achievement: A profile of classroom practices*. Los Angeles, CA: Center for the Study of Evaluation.
- Elbow, P. (1997). Taking time off from grading and evaluating while working in a conventional system. *Assessing Writing*, 4, 5-28.

- Firestone, W. A., & Mayrowetz, D. (2000). Rethinking "high stakes": Lessons from the United States and England and Wales. *Teachers College Record*, 102(4), 724-749.
- Gearhart, M., & Wolf, S. (1994). Engaging teachers in assessment of their students' narrative writing: The role of subject matter knowledge. *Assessing Writing*, 1, 67-90.
- Georgia Department of Education. (2001). *Assessment and instructional guide for the Georgia High School Writing Test*. Atlanta: Author.
- Georgia Department of Education. (2003). Superintendent says Georgia's writing scores show need for improvement. *Newsroom*. Retrieved November 11, 2004 from <http://www.doe.k12.ga.us/doe/media/03/071003>
- Grant, S.G. (2000). Teachers and tests: Exploring teachers' perceptions of changes in the New York State-mandated Testing Program. *Education Policy Analysis Archives* [On-line serial], 8(14). Available: <http://epaa.asu.edu/epaa/v8n14.html>.
- Grant, S.G. (2001). An uncertain lever: Exploring the influence of state-level testing on teaching social studies. *Teachers College Record*, 103(3), 398-426.
- Graves, D., & Sunstein, B. (Eds.). (1992). *Portfolio portraits*. Portsmouth, NH: Heinemann.
- Haertel, E. (1989). Student achievement tests as tools of educational policy: Practices and consequences. In B.R. Gifford (Ed.), *Test policy and test performance: Education, language, and culture* (pp. 25-50). Boston, MA: Kluwer Academic.
- Hale, G., Taylor, C., Bridgeman, B., Carson, J., Kroll, B., & Kantor, R. (1996). *A study of writing tasks assigned in academic degree programs* (TOEFL Research Report no. 54). Princeton, NJ: Educational Testing Service.
- Hamilton, L. (2003). Assessment as a policy tool. In R. Floden (Ed.), *Review of research in education* (Vol. 27, pp. 25-68). Washington, DC: American Educational Research Association.
- Haney, W. (1991). We must take care: Fitting assessments to functions. In V. Perrone (Ed.), *Expanding student assessment* (pp. 142-163). Alexandria, VA: Association of Supervision and Curriculum Development.
- Hilgers, T. (1992). *Improving placement exam equitability, validity, and reliability*. Paper presented at the Conference on College Composition and Communication, Cincinnati, OH.
- Hillocks, G., Jr. (2002). *Testing trap: How states writing assessments control learning*. New York: Teachers College Press.
- Jennings, C. (2002). *Aligning writing instruction in secondary and post secondary institutions. League for innovation in the community college*. Retrieved December 31, 2004 from <http://www.league.org/leaguetc/express/inn0207.htm>
- Johnson, T.S., Smagorinsky, P., Thompson, L., & Fry, P. G. (2003). Learning to teach the five-paragraph theme. *Research in the Teaching of English*, 38, 136-176.
- Kentucky Department of Education. (1999). *Kentucky writing portfolio: Writing portfolio development teacher's handbook*. Frankfurt, KY: Author
- Ketter, J., & Pool, J. (2001). Exploring the impact of a high-stakes direct writing assessment in two high school classrooms. *Research in the Teaching of English*, 35, 344-393.
- Koretz, D., & Hamilton, L.S. (2003). *Teachers' responses to high-stakes testing and the validity of gains: A pilot study* (CSE Tech. Rep. No. 610). Los Angeles: Center for Research on Evaluation, Standards, and Student Testing.
- Koretz, D., Mitchell, K., Barron, S., & Keith, S. (1996). *The perceived effects of the Maryland School Performance Assessment Program* (CSE Tech. Rep. no. 409). Los Angeles: National Center for Research on Evaluation, Standards, and Student Testing, University of California, Los Angeles.
- Koretz, D., Stecher, B., Klein, S., & McCaffrey, D. (1994a). The Vermont portfolio assessment program: Findings and implications. *Educational Measurement: Issues & Practice*, 3(3), 5-16.
- Klein, S., Koretz, D., McCaffrey, D., & Stecher, B. (1994b). *The evolution of a portfolio program: The impact and quality of the Vermont Program in its second year (1992-1993)* (CSE Technical Rep. No. 385). Los Angeles: National Center for Research on Evaluation, Standards, and Student Testing, University of California, Los Angeles.

- Koretz, D., Stecher, B., Klein, S., McCaffrey, D., & Deibert, E. (1993). *Can portfolios assess student performance and influence instruction? The 1991-1992 Vermont Experience* (CSE Tech. Rep. No. 371). Los Angeles: National Center for Research on Evaluation, Standards, and Student Testing, University of California, Los Angeles.
- Loofbourrow, P. (1994). Composition in the context of the CAP: A case study of the interplay between composition assessment and classrooms. *Educational Assessment*, 2(1), 7-49.
- Madaus, G. F. (1988). The influence of testing on the curriculum. In L. Tanner (Ed.), *Critical issues in curriculum, Eighty-seventh Yearbook of the National Society for Study of Education* (pp. 83-121). Chicago: University of Chicago Press.
- Madaus, G. F., & Kellaghan, T. (1993). Testing as a mechanism of public policy: A brief history and description. *Measurement and Evaluation in Counseling and Development*, 26(6), 10-21.
- Mehrens, W. (1998). Consequences of assessment: What is the evidence? *Education Policy Analysis Archives*, 6(13). Retrieved from <http://epaa.asu.edu/epaa/v6n13/>
- Murphy, S. (2003). That was then, this is now: The impact of changing assessment policies on teachers and the teaching of writing in California. *Journal of Writing Assessment*, 1(1), 23-45.
- National Commission on Excellence in Education. (1983). *A nation at risk: The imperative for educational reform*. Washington DC: Author.
- National Council of Writing Program Administrators. (1999). The outcomes statement for first-year composition. *Writing Program Administration* 23, 59-66. (Also available at <http://comppile.tamucc.edu/WPAoutcomes/index.htm>.)
- O'Neill, P., Murphy, S., Huot, B., & Williamson, M. (2004, November). *What high school teachers in three states say about high stakes writing assessments*. Symposium held at the annual conference of the National Council of Teachers of English, Indianapolis.
- Pearson, P.D., & Valencia, S. (1987). Assessment, accountability, and professional prerogative. In J.E. Readence & R.S. Baldwin (Eds.), *Research in literacy: Merging perspectives; Thirty-Sixth yearbook of the National Reading Conference*. Rochester, NY: National Reading Conference.
- Pedulla, J., Abrams, L., Madaus, G., Russell, M., Ramos, M., & Miao, J. (2003). *Perceived effects of state-mandated testing programs on teaching and learning: Findings from a national survey of teachers*. Chestnut Hill, MA: National Board on Educational Testing and Public Policy.
- Polio, C., Fleck, C., & Leder, N. (1998). "If I only had more time": ESL learners' changes in linguistic accuracy on essay revisions. *Journal of Second Language Writing*, 7(1), 43-68.
- Popham, W.J. (1987). The merits of measurement-driven instruction. *Phi Delta Kappan*, 66, 628-634.
- Purves, A. C. (1995). Apologia not accepted. *College Composition and Communication*, 46, 549-550.
- Quality counts. (2002). *Education Week*, 21(17). Retrieved from <http://www.edweek.com/sreports/qc02/>
- Roeber, E., Bond, L.A., & Braskamp, D. (1997). *Trends in statewide student assessment programs, 1997*. North Brook, IL: North Central Regional Educational Laboratory and the Council of Chief State School Officers.
- Shay, S. (1997). Portfolio assessment: A catalyst for staff and curricular reform. *Assessing Writing*, 4(1), 29-52.
- Sheingold, K., Heller, J., & Paulukonis, S. (1995). *Actively seeking evidence: Teacher change through assessment development*. Princeton, NJ: Educational Testing Service.
- Scherff, L., & Piazza, C. (2005). The more things change, the more they stay the same: A survey of high school students' writing experiences. *Research in the Teaching of English*, 39(3), 271-304.
- Shepard, L. (1995). Using assessment to improve learning. *Educational Leadership*, 54(5), 38-43.

- Simmons, J. (1992). Don't settle for less in large-scale writing assessment. In K. Goodman, L.B. Vird, & Y. M. Goodman (Eds.), *The whole language catalog: Supplement on authentic assessment* (pp. 160-161). Santa Rosa, CA: American School Publishers.
- Smith, M. L. (1991). Put to the test: The effects of external testing on teachers. *Educational Researcher*, 20(5), 8-11.
- Spaulding, E., & Cummins, G. (1998). It was the best of times. It was a waste of time: University of Kentucky students' view of writing under KERA. *Assessing Writing*, 5(2), 167-200.
- Stecher, B.M., Barron, S.I., Kaganoff, T., & Goodwin, J. (1998). *The effects of standards-based assessment on classroom practices: Results of the 1996-97 RAND survey of Kentucky teachers of mathematics and writing* (CSE Tech. Rep. no. 482). Los Angeles: Center for Research on Evaluation, Standards, and Student Testing.
- Storms, B. A., Sheingold, K., Nunez, A., & Heller, J. (1998). *The feasibility, comparability, and value of local scorings of performance assessments* (Tech. Rep.) Princeton, NJ: Educational Testing Service, Center for Performance Assessment.
- University of California Davis University Writing Program. (2004). *Expanded course description English 101, Advanced Composition*. Retrieved September 25, 2004 from <http://cai.ucdavis.edu/advancedclasses/101.doc>.
- Wallace, V.L. (2002). Administrative direction in schools of contrasting status: Two cases. In G. Hillocks Jr. (Ed.), *The testing trap: How state writing assessment control learning* (pp. 93-102). New York: Teachers College Press.
- Wiggins, G. P. (1993). *Assessing student performance: Exploring the purpose and limits of testing*. San Francisco: Jossey-Bass.