

**The Politics of State Testing:
Implementing New Student Assessments**

CSE Technical Report 424

Lorraine M. McDonnell
CRESST/University of California, Santa Barbara

February 1997

National Center for Research on Evaluation,
Standards, and Student Testing (CRESST)
Graduate School of Education & Information Studies
University of California, Los Angeles
Los Angeles, CA 90024-1522
(310) 206-1532

Copyright © 1997 The Regents of the University of California

The work reported herein was supported by the Educational Research and Development Center Program cooperative agreement R117G10027 and CFDA catalog number 84.117G as administered by the Office of Educational Research and Improvement, U.S. Department of Education.

The findings and opinions expressed in this report do not reflect the position or policies of the Office of Educational Research and Improvement or the U.S. Department of Education.

CONTENTS

ABSTRACT	v
I. INTRODUCTION	1
Study Purpose	2
Study Data and Analysis	3
II. THE POLITICS OF ENACTMENT	5
Assessment Policies: A Brief Overview	5
California	5
Kentucky	6
North Carolina	8
Ideas and Expectations	9
Opportunities for Policy Action	15
Policy Constraints	21
III. THE POLITICS OF IMPLEMENTATION	29
Opposition Groups and Their Concerns	30
Opposition Group Tactics	44
Explaining State Responses	48
The California Response	48
The Kentucky Response	58
The Technical Becomes Political	62
IV. CONCLUSIONS	65
REFERENCES	76

**THE POLITICS OF STATE TESTING:
IMPLEMENTING NEW STUDENT ASSESSMENTS**

**LORRAINE M. MCDONNELL
CRESST/UNIVERSITY OF CALIFORNIA, SANTA BARBARA**

ABSTRACT

Although advocates of alternatives to multiple-choice testing recognize that such assessments pose significant technical problems, the political dimensions of assessment policy are typically overlooked. Yet because of their link to state curriculum standards, these assessments often embody unresolved value conflicts about what content should be taught and tested, and who should define that content. Consequently, it is not surprising that the process was politicized in several of the states that were early implementors of new assessments.

This study examines the politics of state testing by analyzing the experience of California, Kentucky, and North Carolina as the three states implemented new assessments during the early 1990s. It is based on 51 elite interviews conducted in 1992 and 1994, six years of coverage of the state assessment by a major newspaper in each state, and relevant documentary data.

The three states are notable for both the similarities and differences among them. Despite North Carolina's incremental approach to altering its assessment, the state faced resource constraints and professional development demands similar to those of California and Kentucky. However, because the North Carolina assessment represented only a modest departure from the previous state test and was linked to a well-established state curriculum, it did not face the political opposition experienced by the other two states with their more sweeping changes in assessment policy. In California and Kentucky, interest groups mobilized in opposition to the new assessments. Although these groups represented only a small minority of parents and teachers, they tapped broader public concerns about recent curriculum reforms. Opposition to the state assessment resulted in its demise in California, while the Kentucky assessment continues to operate largely because state officials responded differently there than in California. Kentucky officials were motivated to act decisively to preserve the new assessment, while at the same time modifying it, because the broad scope of the state reforms had required a major investment of their political capital. California policymakers, in contrast, had never agreed on what the assessment was to accomplish and, in the face of political opposition, had little incentive to defend the new test on its own merits.

The three states' experience suggests that not all states have the political will and capacity to undertake fundamental changes in their curriculum and testing policies; that strong, flexible political leadership is necessary for such changes; and that the politics of state assessments requires that they be developed through an open, public process.

THE POLITICS OF STATE TESTING: IMPLEMENTING NEW STUDENT ASSESSMENTS¹

**LORRAINE M. MCDONNELL
CRESST/UNIVERSITY OF CALIFORNIA, SANTA BARBARA**

I. INTRODUCTION

Despite its centrality to policy, student assessment has traditionally been framed as a technical issue. Discussion has focused on whether specific types of tests can measure student learning reliably and validly, and on the appropriate uses of assessment results. Increasingly, however, a broader range of values have come to dominate debates over student assessment. These values represent beliefs and ideas about what purpose assessments should serve, what content should be tested, and who should define that purpose and content. They encompass professional norms about what constitutes effective teaching and how student achievement should be measured, as well as political norms about accountability and the relationship between schools and the larger community. These values may not always be compatible and, at times, may even clash. In fact, as the last few years have demonstrated, the politics of testing often revolves around conflicting values about the appropriate format and uses of testing.

In past work, I examined how the values that underlie testing policy can work as a policy instrument (McDonnell, 1994a). Considerable research has documented that school and classroom practices change in response to assessments (e.g., Corbett & Wilson, 1991; Herman & Golan, nd; Madaus, 1988). In some instances, that change is prompted by the rewards and sanctions that states attach to assessment results. However, even in states where no tangible consequences are linked to test scores, changes in teaching have also occurred. In those cases, policymakers have relied on a hortatory policy instrument that is based not on material rewards and sanctions, but on deeply-held values and the ability to persuade. These values include specific ideas about how schools can be improved, what should be taught and tested in public schools, and a more general

¹ I am grateful to Craig Choisser, Kirstin Falk, Lawrence Messerman, Sara Wakai, and Leah Watts for their assistance in collecting the data analyzed in this report.

belief that one way to effect greater educational accountability is to generate information about school and student performance. Policymakers believe that assessments will make teachers' and students' behavior more consistent with the curricular values underlying the assessments, and that citizens will take action to improve schools in response to the information offered through the publication of test scores (McDonnell, 1994b).

Study Purpose

This monograph builds on prior conceptual work by examining the politics of state assessments through the lens of the values influencing enactment and implementation. However, the values of various participants are not the only factors shaping the state politics of testing. Other variables such as available resources, the interests of participants, and opportunities to take action are also significant in explaining policy outcomes. These other factors are also examined as they facilitate or constrain policymakers and their constituents in promoting ideas about education reform, appropriate curricula, and educational accountability through assessment policy.

The current research is part of a larger study that has examined the design and implementation of new state assessments in California, Kentucky, and North Carolina over the past four years. These three states were selected for study because their policies represent a range of assessment formats and uses, and because they were among the earliest states to change their assessment systems. California's and Kentucky's state assessments represented a major departure from the traditional reliance on multiple-choice testing. Both states included constructed response items and performance tasks in their tests, and Kentucky's assessment also includes portfolios of students' work. Kentucky attached school-level rewards and sanctions to its assessment system, while California did not. North Carolina's assessment, on the other hand, represents a less significant departure from its previous testing program. Most of the test still consists of multiple-choice items, but it also includes some constructed responses and a writing sample in several grades. It does not include any performance tasks or the use of portfolios. In all three states, the new assessment was directly linked to state curricular standards.

A guiding assumption throughout this study has been that different policy designs will lead to variation in both the implementation process and in its

eventual outcomes. Consequently, it was important to include several different design dimensions in the sample—viz., whether or not consequences were attached to assessment results and the extent of departure from past testing practices. At the time this study began, only ten states had moved to alter their assessment systems, so these three states represent a significant proportion of the early implementors.²

The main question framing the research is: To what extent are state assessments producing results consistent with policymakers' expectations? Behind this seemingly straightforward question lie a set of prior questions that must be addressed. They include:

- What ideas and values define new state assessments?
- Are policymakers' expectations about these assessments consistent with those of parents, the public, and professional educators?
- What role do the technical issues that have typically dominated debates about standardized testing play in the current politics of student testing?

This monograph will address these questions by focusing on the enactment and implementation of new assessments at the state level. A subsequent report will examine local implementation, focusing particularly on the extent to which teaching is consistent with the goals of the state assessment.

Study Data and Analysis

The analysis is based on three data sources: elite interviews conducted in 1992 and 1994, articles published in a major newspaper in each of the three states, and a review of relevant documentary data. In the first round, 30 interviews were conducted across the three states with state-level respondents (legislators and their staffs, governors' education aides, state department of education leaders, and interest group representatives). A second round of 29 interviews was conducted in 1994: Eight of these were follow-ups with policy elites who had been interviewed in 1992 and the rest were with additional participants in the assessment policy process. All of the interviews were conducted face-to-face

² Arizona and Vermont, two other states that implemented new forms of assessment during this period, were also studied by researchers affiliated with CRESST (see Koretz, Stecher, Klein, & McCaffrey, 1994; Noble & Smith, 1994).

using a structured, open-ended protocol. The interviews averaged forty-five minutes in length.³

All news articles, columns, editorials, and letters to the editor about the state assessment and related education reforms that appeared in the *Los Angeles Times*, the *Lexington Herald-Leader*, and the *Charlotte Observer* between January 1990 and February 1996 were obtained through a computerized search. Articles appearing in other newspapers in the three states were obtained on a less systematic basis, most notably when respondents mentioned particular articles as relevant to debates about the state assessment.

A variety of documentary data were also collected and reviewed. These included copies of all the relevant legislation; materials produced by groups supporting and opposing the state assessments; state curriculum documents, including curriculum frameworks; technical reports on the state assessment produced by the state department of education and by outside monitoring bodies; test items that have been publicly released; and in the case of Kentucky, several public opinion polls on its larger education reform and state assessment.⁴

These data were analyzed by making structured comparisons among the three states on several key variables. The most important variable was policy design, particularly the type of policy instrument used. Given past evidence about the impact of testing on classroom practice, I assumed that local context would be less important in explaining implementation outcomes for assessment policy than is typically the case for other top-down education policies. Consequently, initial policy design was likely to be a more significant explanatory factor. I also assumed that assessment policies with tangible consequences attached to test results would be more powerful in changing local practice than assessments without such consequences. However, after the first round of state-level data collection, but

³ Respondents were assured that their responses would be confidential, so only their role positions are noted in the interview excerpts used throughout this monograph. Those involved in assessment policy in the three states are only identified by name when the information presented is from newspaper accounts or other public documents.

⁴ In addition to the data collected at the state level, 292 interviews were conducted with local school administrators, teachers, and parents between November 1993 and March 1995. Twenty-three teachers in Kentucky and the same number in North Carolina were interviewed twice, and a sample of their classroom assignments were collected and coded to determine how consistent they were with the goals of the state assessments. Because of the controversy over the state assessment in California, in-depth case study research was also conducted at seven schools where grass roots opposition to the test was particularly strong. The analysis of these data are presented in a separate monograph on local implementation.

prior to the local fieldwork, it became apparent that curricular values were critical in all three cases and that these values could be as important in shaping implementation outcomes as rewards and sanctions. Theoretical work on hortatory instruments that rely on values to change policy targets' behavior is limited (Schneider & Ingram, 1990; Stone, 1988). Nevertheless, it does suggest that political factors surrounding policy implementation might be as significant in explaining outcomes as other variables more directly linked to the implementation process, such as the technical soundness of a policy, its implementation time frame, and the level of available resources.

Consequently, the analysis presented in this monograph analyzes three categories of variables: the ideas shaping assessment policy and expectations for what the new tests should accomplish; the politics of assessment policy as evidenced in the enactment and implementation processes; and the technical and resource issues that affected implementation. The next section provides a brief overview of the three states' assessment systems and then examines the politics of enactment, concentrating on the ideas that framed each state's assessment, policymakers' expectations for the new assessments, and the opportunities and constraints that shaped policy formulation. A third section focuses on the state politics of implementation with particular attention to the political controversies and technical challenges influencing implementation outcomes. A final section discusses lessons drawn from the experience of California, Kentucky, and North Carolina that can inform deliberations in other states considering new forms of student assessment.

II. THE POLITICS OF ENACTMENT

Assessment Policies: A Brief Overview

California

Educators and testing experts looked to the California Learning Assessment System (CLAS) as an exemplar for new forms of more authentic student assessment. When fully operational, CLAS was intended to test all the state's students in Grades 4, 8, and 10 in reading, writing, mathematics, science, and social studies. Although the tests contained multiple-choice items, students were also required to answer open-ended questions that asked them about their

reaction to literary passages or how they arrived at a particular solution to a mathematics problem. The tests were tied to state curriculum frameworks that stressed the ability to understand underlying principles and to apply them in solving real-world problems, as well to master subject matter knowledge. Each assessment was scored on a scale of 1 to 6 that measured the extent to which students had mastered the state skill standards.

The immediate impetus for CLAS was Republican Governor Deukmejian's 1990 veto of appropriations for the state's prior test, the California Assessment Program (CAP).⁵ Fifteen months after the CAP veto, CLAS, or the "new CAP" as it was originally called, was enacted through a rare consensus among the three centers of education power in California—Deukmejian's successor, Pete Wilson; the Legislature in the person of former State Senator Gary Hart, then chair of the Senate Education Committee; and former State Superintendent of Public Instruction Bill Honig. Yet after only two administrations of the test in 1993 and 1994, a bill authorizing the continuation of CLAS was vetoed by Wilson in October 1994. Because of its innovative approach to assessment, CLAS encountered many technical problems, and its development and administration were further constrained by cost considerations. However, as the subsequent analysis will indicate, political factors reflecting the different values held by key policymakers and their constituents were more significant in explaining both the initial enactment of CLAS and its quick demise.

Kentucky

The student assessment system in Kentucky, the Kentucky Instructional Results Information System (KIRIS), is one component of what is probably the most comprehensive state education reform in this century. In 1989, in response to a lawsuit challenging the constitutionality of the state's school finance system, the Kentucky Supreme Court ruled not just that the finance system was unconstitutional, but that the entire state school system was unconstitutional. The court ordered the legislature to redesign Kentucky's education system in its entirety. The Kentucky Education Reform Act (KERA), signed into law in April 1990, pumped over \$700 million in new funds into the system in its first two years

⁵ Deukmejian vetoed the CAP appropriation in a dispute with the Democratically controlled Legislature over the funding source. He argued that the \$11 million to support CAP should come from funds that had specifically been reserved for education through a ballot initiative, rather than from the state's General Fund.

of implementation. KERA made a number of sweeping changes that ranged from how the duties of local school boards are defined to how teachers are licensed and what is taught in classrooms. It required that elementary schools teach younger children in “ungraded primaries” that combine students from kindergarten through third grade in the same classrooms; it mandated that each school establish a site council to govern its curricular, personnel, and budgetary decisions; and it created a network of family-service and youth-service centers located at or near schools with large concentrations of poor students.

KIRIS stands at the heart of the KERA reforms. As part of the comprehensive overhaul of its educational system, Kentucky developed six learning goals. From those goals, 75 “valued outcomes” were defined, setting the standards for what Kentucky schools would be expected to teach and students to master. All students in Grades 4, 8, and 12 are tested in five subjects using an assessment system that consists of multiple-choice items, short-answer questions, performance tasks, and portfolios.⁶ Student performance on these assessments is scored as “novice,” “apprentice,” “proficient,” and “distinguished,” based on a set of absolute standards for what students are expected to know. Once baseline scores (derived from a combination of student assessment scores and other data such as attendance and graduation rates) are calculated for each school, schools are required to improve by a specified percentage or threshold score every two years. Those that exceed their threshold score are eligible for monetary awards; schools that fail to improve or that decline by less than five percentage points are provided various forms of assistance; those that decline by more than five percentage points will be put on probation. Under the latter condition, parents have the option of transferring their children out of the school, and school staff face the possibility of dismissal.

KIRIS confronted many of the same technical and political problems as CLAS, but state policymakers responded differently than in California, and the assessment is still being administered. Nevertheless, changes were made in the

⁶ As in California, the format and content of the KIRIS assessment represented a marked departure from more traditional forms of testing. Students are required to write more; the initial stages of some tasks are performed in groups although students give individual answers; and students are given “real world” problems to solve, such as siting a new Wal-Mart in their community or designing an experiment to determine which of two spot removers works better (both examples are from the fourth-grade test). Student portfolios in mathematics and writing contain five to seven examples of students’ work and are selected to show the breadth of their understanding of core concepts and principles.

test and in its consequences. State officials eliminated two of the state's six learning goals—dealing with individual self-sufficiency and responsible group membership—that were targets of the most intense criticism; they delayed the imposition of sanctions on underperforming schools until 1996; and they shifted the high school assessment from the twelfth to the eleventh grade. In February 1995, the state awarded \$26 million to the 480 of the state's approximately 1400 schools that had reached their improvement goals. Depending on the degree of improvement, awards to individual schools ranged from \$1300 to \$2600 per teacher.

North Carolina

The state's new assessment, first administered in 1993, represents a less significant departure from its previous testing program than for either California or Kentucky. Students in Grades 3-8 are tested at the end of each year in reading, mathematics, and social studies using a combination of multiple-choice items and open-ended questions. In addition, students in Grades 4 and 6-8 are required to produce a writing sample.⁷ The North Carolina assessment does not include any group exercises or performance tasks (e.g., working with scientific apparatus), nor does it require student portfolios. Exams are scored at three levels of proficiency—basic, proficient, and exemplary or exceptional.

North Carolina has attached few tangible consequences to test results. The state does appropriate about \$250 per teacher, which local districts, and eventually individual schools, can either allocate as part of a differentiated pay plan or use for staff professional development. There has been some discussion that test results might be used as one criterion in making decisions about the local use of those funds. But most respondents argued that because of the small amount of funding available and the likely opposition of the teachers' union, test results would not be used in rewarding particular schools or their staffs. Similarly, North Carolina has a "take-over" plan that allows the state to intervene directly in local school districts performing at substandard levels. However, state take-over is not a realistic threat for most districts: No more than 3% or 4% of the

⁷In addition to the end-of-grade tests for elementary and middle school students, North Carolina also administers end-of-course tests to high school students in about 14 different subjects. However, these tests still largely consist of multiple-choice items and are only gradually being revised to include some open-ended items. The state does not intend to have most of these tests revised and operational until 1996-97.

state's districts are likely to be affected, and those most likely to be are chronically poor, low-performing districts that are already well known.

Although the end-of-grade test continues to be administered in North Carolina, its future is uncertain. A Standards and Accountability Commission, initially established by the governor but now working under the auspices of the State Board of Education (SBE), is exploring new forms of assessment in connection with new student standards and a possible exit examination for high school seniors. Although the Commission has yet to issue a final report, there is some indication that it will recommend less state-level testing for accountability purposes, coupled with a system that encourages classroom-centered performance testing on an ongoing basis. At the same time, the SBE is emphasizing core academic skills, so the state's end-of-grade test will now only include reading, mathematics, and writing. Students are no longer tested in social studies, and the planned addition of a science test has been canceled. These changes in the assessment system have come during a period of political uncertainty with a Democratic governor up for re-election, the Republican party in control of the lower house of the state legislature for the first time in this century, and a 38% reduction in the state Department of Public Instruction's (DPI) budget initiated by the legislature with the consent of the governor.

Ideas and Expectations

Most public policies embody a set of ideas, as well as advance the interests of those who will benefit from them. These ideas or values reflect goals in addition to self-interest and are often based on theories about how the world works or ought to work (Kingdon, 1993). For example, even policies that serve to benefit the interests of particular groups, such as organized teachers or textbook publishers, usually seek to achieve broader goals such as improved student learning. They also typically make assumptions about the causal process by which policy goals can be achieved—for instance, how students learn or the conditions under which teachers are likely to change their instructional strategies.

Policies that represent a marked departure from the past, as is the case for the new assessment systems, are even more likely to reflect particular ideas or values. Those ideas then define the expectations that policymakers have for what the policy should accomplish. They are reflected in statements such as:

KERA is based on three un-American assumptions: (1) all children can learn; (2) all children can learn at high levels; and (3) effort, not innate ability, can achieve success. (Kentucky Department of Education official)

The expectation is that we will have better educated kids. Policymakers want to close the gap between what students need to know and what they already have—viz., basic skills. They want students to have the ability to think productively, to have different ways to answer the same question, or to have multiple solutions to problems. And we have test items like that. We're big on judging in North Carolina and reporting performance. Over time, performance is expected to be better, as compared with peers in other states. (North Carolina Department of Public Instruction official)

Testing is one of the most powerful leverage points in improving student learning. (California Department of Education official)

Even though the three states differ in the origins of their assessment policies and in test format and use, three ideas are common to each. The first is that the assessment should be linked to a set of standards or to a specific curriculum. In both California and North Carolina, the state assessment was specifically aligned with the state curriculum frameworks. In the eyes of many educators, the lack of alignment between CAP and the state frameworks had been a major obstacle to full implementation of the frameworks: They assumed a constructivist approach to instruction, while CAP with its multiple-choice format was reinforcing more directive, traditional forms of teaching (Cohen & Peterson, 1990; Marcano, 1990). In North Carolina, a major impetus for the new assessment was closer alignment with the statewide curriculum (the *Standard Course of Study*) than had been possible when the state was relying on a generic, commercial test. Although the state frameworks are not as specific in Kentucky and were not completed until after the first administration of KIRIS, the state assessment was tied to a concrete set of performance standards.

Directly related to this link between curriculum and assessment is a second idea underlying the three state assessments. While also functioning as an accountability mechanism, assessment should influence classroom instruction. State policymakers assumed that the curricular values embodied in the assessment system would persuade teachers to teach differently.

As powerful as the rewards and sanctions are, the clarity of the standards to teachers are going to be more powerful. I think teachers definitely want their

students ready. If their peers decided what all 4th graders should know and be able to do, that kind of clear feedback is going to be very powerful. (Former Kentucky Commissioner of Education Thomas Boysen, as cited in Rothman, 1992, p. 20)

I see the assessment as accomplishing both accountability and curriculum change. Some people see the assessment as only being about accountability. But if you have a test of this importance, it will drive the curriculum. (Kentucky legislator)

The General Assembly thinks that the best shot at changing the curriculum is to have a test aligned with the curriculum. The DPI has now changed its rhetoric, and is telling teachers to teach to the test. (North Carolina DPI official)

The third idea embodied in these assessments is a particular view of how students learn most effectively. The assumption is that rote memorization of facts does not result in the kind of sustained, worthwhile knowledge that students need. Rather, students should learn in a way that will allow them to think critically, solve problems, and apply knowledge in unfamiliar situations.

What the legislature had in mind was designing a system that is performance-based. They want our kids to do things, not just know things. (Kentucky Department of Education official)

I was a math and science teacher before I took this position. The typical student is thinking, "Tell me, I may remember. Show me, I may remember. Involve me, I will understand it." I saw the importance of this as a basketball coach: students with glazed eyes wake up with sports. The performance events will motivate students. Working with other folks, youngsters will be able to see, to be involved, to do experiments. The results of all this will be that students will have a different experience. (Kentucky teacher union official)

According to this view, teaching for understanding is most likely to occur if students have opportunities to construct their own learning through hands-on experiments and projects, interactions with other students, and the application of knowledge and skills to real-world problems. In this model, there is less lecture and teacher-directed instruction, and a greater emphasis on the teacher facilitating students' own active learning. Although North Carolina's assessment does not go as far as the other two states in measuring this kind of learning, the curriculum frameworks in all three states reflect this philosophy.

These ideas were embodied in policies designed to meet particular state needs and circumstances, but the policies were also connected to a broader national

movement. Public intellectuals such as Marshall Smith (later to be Under Secretary of Education in the Clinton administration) were writing about an education reform strategy whose core is a coherent set of state curriculum standards linked to policies dealing with teacher training and licensure, curricular materials, and student assessment (Smith & O'Day, 1991). Others promoted the idea of changing instruction and learning through the testing process, and stressed the importance of a constructivist approach to student learning (Resnick, 1994). Some of the professional associations, most notably the National Council of Teachers of Mathematics (NCTM), were also developing content and performance standards that outlined what students should know and be able to do (National Council of Teachers of Mathematics, 1989).⁸

At the national level, the policy manifestation of these ideas was first seen in the six national education goals promulgated at the 1989 Charlottesville “education summit” between President George Bush and the nation’s governors. Subsequently, the National Council on Education Standards and Testing (NCEST), a bi-partisan group established by Congress and comprised of policymakers, researchers, and educators, issued a report endorsing the concept of voluntary national standards tied to assessments. NCEST envisioned these standards as including “substantive content together with complex problem-solving and higher order thinking skills” (1992, p. 3). In 1994, Congress passed Goals 2000, a Clinton administration initiative that codified and expanded to eight the national education goals, and sought to encourage the states to adopt voluntary content, performance, and opportunity-to-learn standards.

Since the NCEST report was issued, the emphasis has moved away from a single set of voluntary national standards to ones developed by the states. Still, the national reform movement and its accompanying policies continue to influence

⁸ The curriculum standards developed by NCTM combine critical thinking skills, such as problem solving and making connections among topics and their applications, with specific content knowledge. The standards are characterized by learning goals that emphasize understanding the conceptual basis of mathematics, reasoning mathematically and applying that reasoning in everyday situations, offering alternative solutions to problems, and communicating about mathematical concepts in meaningful and useful ways. Consistent with those goals, NCTM advocated changes in both mathematics content and instructional strategies. Particularly prominent in this reform vision of the mathematics curriculum is a changed view of the teacher’s role. Because students are expected to play an active part in constructing and applying mathematical ideas, teachers are to be facilitators of learning rather than imparters of information. In terms of instructional activities, this shift means that rather than lecturing and relying on a textbook, teachers are to select and structure mathematical tasks that allow students to learn through discussion, group activities, and other modes of discovery.

the general direction of state policies. For example, as a condition for receiving federal Title I funding, states must now demonstrate that they are using challenging content and performance standards, and assessing students with tests aligned with the standards and measuring higher order thinking skills (Smith & Scoll, 1995).⁹

In California and Kentucky, the links between the national movement and the design of state assessment policies were quite direct. The initial blueprint for CLAS was conceived by an Assessment Policy Committee (APC), appointed by the SDE and chaired by Thomas Payzant, then superintendent of the San Diego city schools.¹⁰ The APC's design was then incorporated into SB 662, the CLAS legislation. Its members included policy staff to Hart and Wilson, local school superintendents, teacher union representatives, and a representative of the state's business community. However, it also included several academics who were active in the national standards and assessment movement.¹¹

Although the state's valued outcomes originated in the work of an in-state Council on School Performance Standards, established prior to the state Supreme Court decision, Kentucky was also influenced by national reform trends. In fact, policymakers acknowledged that they had intentionally modeled KERA on the national reform movement. "Most of the ideas came from national discussions," noted one of the legislators instrumental in the reforms. The governor's education aide indicated that they had "talked to everyone prominent in alternative assessment."¹² Like a number of other states, Kentucky also designed its mathematics standards and assessment to be consistent with those recommended by NCTM.

⁹ In contrast to the Goals 2000 program with Fiscal Year 1995 appropriations of \$403 million, the Title I program provides over \$6 billion a year in federal funding to states and local school districts. Most observers believe that the federal government's influence over the standards and assessment process in individual states will be considerably greater through Title I than Goals 2000, even though the former is targeted on only a subset of students.

¹⁰ Payzant then served for nearly three years as the Assistant Secretary for Elementary and Secondary Education in the Clinton administration; he is now superintendent of the Boston City Schools.

¹¹ Eva Baker was a member of NCEST, and the research center she directs at UCLA has been in the forefront of research on alternative assessments. Similarly, Richard Shavelson, another APC member, had been conducting research on alternative forms of science assessment for some years with funding from the National Science Foundation.

¹² The legislature hired David Hornbeck, the former chief state school officer in Maryland and a member of NCEST, as a consultant to the subcommittee dealing with curricular issues. Grant Wiggins, a nationally visible advocate of performance testing, was one of the consultants hired by the SDE to assist in the initial design of KIRIS.

Although North Carolina's assessment was less influenced by national reform trends, it too revised its mathematics curriculum to be consistent with the principles espoused by NCTM.

Ideas about the link between curriculum and assessment and the role of assessment in leveraging changes in classroom instruction were joined in the three states with the more traditional notion that testing should serve an accountability function. As they described the curricular functions of assessment, respondents also discussed its role as an accountability mechanism. For example, a member of the Kentucky State Board of Education noted that KIRIS "will fix responsibility and accountability," and that "holding schools responsible will put peer pressure on teachers who are not doing well." He went on to describe the state assessment as a way "to sell patience to people whose taxes went up two years ago." Even teacher union leaders, whose support of the new state assessments has been tepid, acknowledge its accountability function:

There's no question that testing is a *quid pro quo* for the additional money that came to schools with the Basic Education Program [a program that increased state funding to local districts and guaranteed a base level of support] . . . The philosophy is: "We've poured all this money into the Basic Education Program, so now we want to make certain that it pays off." (North Carolina teacher union official)

A growing number of states resemble Kentucky in defining educational accountability as the rewarding and sanctioning of schools, based largely on assessment results.¹³ Other states, however, resemble California and North Carolina in conceiving of the accountability functions of their assessment systems as informational. The assumption is that, by sending individual student scores to parents and by reporting school, district, and state scores publicly, parents and concerned members of the public will act on that information by pressuring for improvement where it is needed.

Despite their different approaches, policies in the three states embodied a clear set of ideas, not only about the ultimate goal of assessment, but also about the theory or causal processes leading to the achievement of that goal. The assessment would be linked to well-defined standards and curricula; the underlying

¹³ A 1988 survey of all 50 states found that half of them had policies that triggered specific state actions on the basis of data generated by their assessment systems. These actions included assistance to districts deemed in need of improvement as well as a variety of rewards and sanctions based on performance (Office of Educational Research and Improvement State Accountability Study Group, 1988).

curricular values, combined with the public notification and consequences associated with an accountability system, would prompt changes in teaching; and as a result, students would not only learn more effectively, but would also acquire knowledge of greater worth. Subsequent events demonstrated that these ideas and the expectations they generated were overly ambitious. Nevertheless, the new assessment policies were notable for the powerful values undergirding them. They combined a long-standing assumption that educators would alter their classroom teaching in response to the demands of a public accountability system with a belief that assessment could and should shape the instructional process.

Translating these ideas into legislation and eventually into an actual student assessment depended on a variety of factors, some of which were common to the three states and some that were unique to each.

Opportunities for Policy Action

In each of the three states, the new assessment was a product of unique opportunities for policy action. The California and Kentucky cases closely followed Kingdon's (1995) model of agenda setting. He hypothesizes that policy emerges from the coupling of three independent process streams—problems, policy proposals, and politics:

They are largely independent of one another, and each develops according to its own dynamics and rules. But at some critical junctures the three streams are joined, and the greatest policy changes grow out of that coupling of problems, policy proposals, and politics. (p. 19)

A problem is recognized, a solution is developed and available in the policy community, a political change makes it the right time for policy change, and potential constraints are not severe. (p. 165)

In California and Kentucky, a policy window opened when a solution was joined to a set of problems and both were joined to favorable political forces. In all three states, the solution was a new assessment linked to state curricular standards and relying on a broader range of testing formats than just multiple-choice questions. But the problems and the politics differed in each state.

Three related factors dominated the problem stream in Kentucky: the state's chronically low educational achievement, persistent regional inequities, and the need for greater economic development. Kentucky has traditionally had one of the

highest adult illiteracy rates in the nation, and the 1990 census showed that only Mississippi had a lower percentage of adults graduating from high school. Related to this low academic achievement is the widespread poverty that has plagued the eastern part of the state for generations. It led to high levels of outmigration from the state, and the isolation of the Appalachian counties from the greater prosperity enjoyed by the central Bluegrass region and the western part of the state. Kentucky has moved aggressively over the past decade to attract new industries to the state as a way to lessen its traditional reliance on coal mining, tobacco farming, and heavy manufacturing. The symbol of both the state's success at economic development and its as yet unfulfilled aspirations is the Toyota plant in Georgetown, Kentucky. Policymakers noted the connection between low levels of educational attainment and the need for greater economic development:

People in this state make the connection between education and economic development. Toyota in Georgetown would only hire high school graduates, and they found that many of the applicants were not qualified. Toyota Georgetown is changing people. (Member, State Board of Education)

At heart, KERA is an economic reform to prevent the loss of people. The state needs a viable labor force. The Kentucky education reforms are more tied to the state's economic future than in other states. (State Department of Education official)

In 1988, Governor Wallace Wilkinson sought to address the state's persistently low educational achievement by proposing a series of reforms based on a set of student standards, an alternative assessment to measure progress on those standards, cash awards for teachers and schools that showed improvement on test scores and attendance, and greater instructional autonomy for individual schools. However, legislators balked at the \$75 million a year price tag and based on the state's recent, unsuccessful experience with standardized testing, they also questioned whether an appropriate test could be developed to measure student performance. Consequently, Wilkinson's proposals stalled in the legislature.

During the 1988 legislative session, his initiative was passed by the state Senate, but the legislation never got out of committee in the House. The governor then created the 12-member Council on School Performance Standards to determine what skills and knowledge Kentucky students were to master and how best to measure that learning. When the state Supreme Court decision was

announced in June 1989, the Council had completed about two-thirds of its work, and Wilkinson was hoping to call the legislature into a special session in August to try once again to deal with school reform.

The state Supreme Court decision provided an opportunity that changed the nature of the politics stream, and necessitated an even broader reform than Wilkinson had proposed. In *Rose v. Council for Better Education* (1989), the Kentucky Supreme Court directly addressed the major educational problems facing the state. It agreed with the trial court that each child in the state was entitled to an education that provided:

- (i) sufficient oral and written communication skills to enable students to function in a complex and rapidly changing civilization; (ii) sufficient knowledge of economic, social, and political systems to enable the student to make informed choices; (iii) sufficient understanding of governmental processes to enable the student to understand the issues that affect his or her community, state, and nation; (iv) sufficient self-knowledge and knowledge of his or her mental and physical wellness; (v) sufficient grounding in the arts to enable each student to appreciate his or her cultural and historical heritage; (vi) sufficient training or preparation for advanced training in either academic or vocational fields so as to enable each child to choose and pursue life work intelligently; and (vii) sufficient levels of academic or vocational skills to enable public school students to compete favorably with their counterparts in surrounding states, in academics or in the job market. (790 S.W.2D 186, *212)

It also ruled that the maintenance and funding of common schools in Kentucky is the sole responsibility of the General Assembly (state legislature), that common schools should be available and free to all Kentucky children, that they should be substantially uniform throughout the state, and that common schools should “provide equal educational opportunity to all Kentucky children, regardless of place of residence or economic circumstance” (790 S.W.2D 186, *212).

The governor and his staff had expected the Supreme Court decision “to be purely fiscal.” But its broader scope gave Wilkinson “an unexpected vehicle.” According to one of the governor’s aides:

. . . the court decision came down, and we had just come out with our six goals. The decision had seven—it was made for us! The Court said that the constitutional solution had to be performance-based. So the court order broke the deadlock with the General Assembly.

In the case of Kentucky, then, the assessment was the result of a confluence of clearly articulated problems, the expansion of an existing policy proposal, and a unique set of political circumstances that provided the opportunity for a marked departure from past policies.

The situation in California was considerably less dramatic than in Kentucky. Nevertheless, the same pattern was present with a new form of assessment identified as the solution to the problem of a mismatch between the state's new curriculum frameworks and its old, multiple-choice test. In this case, the precipitating political event providing an opportunity for action was Deukmejian's veto of the CAP appropriation, and the election of a new governor interested in providing parents with standardized information about their children's academic performance.

The enactment of a new state assessment in North Carolina followed the Kingdon model of agenda setting less closely. Still, while the politics stream lacked a single catalytic event like that of the other two states, North Carolina faced the same problem as California. The lack of alignment between the state assessment and the curriculum framework was a hindrance to full implementation of the content and pedagogy reflected in the state's *Standard Course of Study*. The original impetus for a new state testing program came from the Public School Forum, a foundation- and business-funded education support group whose board is comprised one-third from business, one-third elected officials, and one-third from education. The initiative was then picked up by the Department of Public Instruction (DPI), which convinced the legislature that the state needed an assessment more closely tied to the statewide curriculum and one that demanded more of students. The General Assembly authorized the development of such an assessment as part of legislation designed to grant local districts and schools greater autonomy. North Carolina has traditionally had a highly centralized school system, with state government circumscribing local operations through a finance system that specifies with considerable precision allowable expenditure categories and through a statewide curriculum. The legislature saw its Accountability Act as a way to grant local units greater autonomy by allowing them to apply for waivers from state regulation and to begin to move toward site-based management. The assessment was conceived as an exchange for this decentralization of authority.

What was perhaps most striking across the three states was the lack of any opposition to the new assessments when they were first enacted. The teacher

union in each state was a group likely to oppose the new assessment or at least, to express reservations about it. Because the state assessments were expected to alter curriculum content and instructional strategies and to ensure greater school accountability for student learning, teachers would have to bear much of the cost associated with responding to the new tests. Yet in each state, the teacher union had a strong reason to support the new assessment or at least, not oppose it. As noted previously, in North Carolina the assessment was understood to be a *quid pro quo* for increased state resources for the schools, much of which was spent on teacher salaries, including the hiring of 13,000 additional teachers statewide (interview, teacher union official). The same motivation helps explain the response of the Kentucky teachers' union:

Kentucky used to be among the worst in terms of level of state funding for education. The reforms meant more money for education, so as far as teachers were concerned, there was little bargaining over the reform package. (Kentucky teacher union official)

The California Teachers' Association (CTA) supported the new state assessment because the CTA viewed the legislation as an improvement over the old system of multiple-choice testing and because it saw the bill as opening the door for more alternative assessments in the future, including ones designed by teachers and having student portfolios as a component (interview with teacher union official).

Disagreements among other participants in the policy process were either minor or involved issues incidental to the state assessment. For example, in California, the Business Roundtable representative on the Assessment Policy Committee wanted students assessed in Grade 10 because, in his view, that was the "fork in the road" where students should be tested on the common curriculum they had experienced to that point and where they would then choose either a largely academic curriculum or a more vocationally-oriented one. The testing and curriculum experts on the APC argued for testing in Grade 11 because it is a more efficient time in terms of the amount of curricular exposure students have had and because they were concerned that the test not be used for tracking purposes. In order to ensure the Roundtable's support, it was decided that students would be assessed in Grade 10, but that they could retake the test in Grades 11 or 12 if they wished (interviews with two APC members).

In Kentucky, disputes arose in the General Assembly over some provisions of KERA. Debate centered on the size and method for funding the tax increase

needed to support the reforms, provisions to minimize nepotism in local school districts, and limits on the involvement of school employees in school politics. However, the state assessment was never an issue of dispute and the assumption that schools would be held accountable to a set of standards reflecting the areas outlined in the Supreme Court decision was never contested by any of the participants.¹⁴

Perhaps most notable, given subsequent developments, was the lack of vocal opposition from the groups that would later oppose the state assessments in California and Kentucky. The absence of their involvement in California is not particularly surprising because the assessment legislation was essentially developed within the APC by representatives of the state's major education policymakers and then enacted by the Legislature with little discussion or debate. Even legislators who later opposed CLAS acknowledged that they had not been paying attention either to the assessment or to the state curriculum frameworks that had preceded the assessment by some five years. The opposition groups' concerns seemed to be have stemmed initially from individual test questions that offended them. They then moved from those specific items to the general principles underlying CLAS. Consequently, the opposition did not emerge until after the test had been administered.

In Kentucky, several opposition group leaders reported attending public meetings in Frankfort when KERA was first being designed and becoming concerned that "only one educational philosophy was reflected" (interview with opposition group leader). Another opposition group leader ran unsuccessfully for state superintendent in 1991—even though the elective post would soon be abolished as a result of KERA—simply as a way to publicize the anti-KERA position. However, even though KERA was highly visible in Kentucky, its individual provisions—and particularly the assessment—were not well understood by the public, and opposition groups did not become well organized until after the legislation was enacted and the test actually administered.¹⁵

¹⁴ Opposition to specific provisions of KERA was greater in the House than in the Senate, as indicated by the differing margins on the final KERA vote in each chamber: 58:42 in the House and 30:8 in the Senate.

¹⁵ Part of the reason that opposition was limited may be due to the process by which KERA was enacted. It was essentially the product of negotiations among legislative leaders and the governor. According to a study of KERA's passage, most major decisions were not debated or discussed openly. Public discussion was minimized so that individual legislators would not water down or significantly change the bill by adding amendments. The legislation was then

Policy Constraints

For the most part, then, the political constraints that would eventually derail CLAS and threaten KIRIS did not exist when the original assessment legislation was enacted. In all three states, a new assessment was perceived as a partial solution to pressing educational problems. Facilitating political conditions then provided an opportunity for placing it on the state policy agenda. However, even in this positive environment for policy action, political and fiscal factors acted as constraints on the clarity and scope of assessment policy.

In California, the political circumstances that created CLAS led to constraints that would eventually hamper its implementation. Behind the consensus among the governor, the state superintendent, and the chair of the Senate Education Committee lay different expectations for what the new assessment should accomplish. Each of these men supported CLAS for different reasons, and they expected it to accomplish very different things.

Governor Wilson wanted CLAS for one major reason: As a condition for his support, he demanded that it include individual-level student scores that would provide information to parents and would, in his view, allow for the evaluation of individual teachers. Wilson wanted eventually to move to a system of merit pay where teachers with high-scoring students were rewarded. As one of the governor's aides noted:

We could care less about authentic assessment. It costs more money, and we don't know if it is any better. For us, having individual student scores is really powerful: it brings accountability into a system where it isn't there now. It will be possible to assess students, teachers, schools, and districts. Parents can then say they don't want their child in Ms. Smith's classroom because they will have the assessment information.

These expectations contrasted with Senator Hart's who also viewed the primary purpose of the new assessment system as accountability, but a very different kind of accountability. Hart saw a more valid form of assessment as the *quid pro*

enacted with a large number of roads and other public works projects attached that were funded out of the major tax increase attached to KERA (Bishop, 1994).

KERA opponents were later to charge that legislators who agreed to support the reform legislation were offered these "pork barrel" projects for their districts and that the Louisville delegation was threatened with losing an airport expansion project if its members did not vote for KERA. Opponents also argued that the full text of the legislation was not available until nine days before the final vote (interview, 10/26/94).

quo for allowing schools to move to site-based management. For him, the school site was the unit of accountability, not the individual teacher, and his model was that of an exchange—greater accountability for outcomes in exchange for greater autonomy over instruction and school operations. Honig’s motivation was not inconsistent with either of the other two policy actors, but it emphasized different expectations. He was interested in assessments that are more congruent with a particular kind of curriculum, that measure real world performance, and that influence teaching. Although he too was interested in accountability issues, Honig came to the enactment of CLAS from a different angle—a curricular philosophy, a set of performance standards that flow from that philosophy or framework, and an assessment appropriate for measuring performance on those standards.

The immediate result of these differing expectations was that the SDE agreed to produce individual-level test scores within a time frame that exceeded its technical capability. As early as 1992, one of the governor’s aides argued that “Honig probably agreed to things he can’t deliver within the Prop. 98 limits” (state constitutional revenue limits for K-14 education funding). After the CLAS continuation had been vetoed, one of the SDE officials responsible for the assessment noted, “It’s true that Bill [Honig] wanted the new test so badly that he said we’d do individual student scores—‘we’ll do whatever it takes.’” But the SDE felt that it had no choice but to overpromise because at the governor’s insistence, a sunset provision was included in the original legislation repealing the program after only three years (on January 1, 1995). Time became one of the major constraints imposed as a result of state policymakers’ fundamentally different notions about the purposes of assessment.

From the beginning, funding was another significant constraint in all three states. Assessments that include formats in addition to or in place of multiple-choice items cost more. Moving from multiple-choice tests that can be scored by machine to tests that include open-ended or constructed responses, requiring the judgment of a trained scorer, adds substantial costs. In North Carolina, because students in so many grades are tested, items requiring hand-scoring number five million. In California, the number was three million. Estimating the total cost of alternative assessments is difficult because the state, local districts, and individual schools all bear those costs. In addition, some costs such as the distribution and collection of tests are less visible than the obvious ones for development, teacher training, scoring, and reporting. Picus (1996) estimated that

the state share of KIRIS costs averaged about \$46 per student tested for each annual test administration between 1991 and 1994, while the more traditional format of North Carolina's end-of-grade assessment cost about \$7 per student tested for each year between 1992 and 1995.¹⁶ No cost estimates for CLAS are available, but based on state appropriations for the test, it cost at least \$30 per student tested. These estimates contrast with \$2 to \$20 per student for commercially developed, multiple-choice tests.

In each state, fiscal and time constraints meant that the original assessment plans had to be scaled back, with the emphasis placed on the state accountability portion of the assessment at the expense of continuous, classroom-based assessment. In Kentucky, the experts involved in the design of KERA, such as David Hornbeck and the governor's education aide, Jack Foster, assumed that it would include an assessment component that would be ongoing, embedded in the curriculum of each classroom, and administered by teachers. This element would be complementary to the assessment administered for accountability purposes, but it would be used more frequently and would be designed by teachers with assistance from the state. However, that assessment component has never been developed in Kentucky. The commissioner of education estimated that it would cost an extra \$50 million over five years, in addition to the \$30 million the state was spending on the accountability assessment. Respondents disagreed about how the continuing assessment was to function. A staff member in the Governor's Office argued that the SDE misunderstood the purpose of the continuing assessment, and assumed that it meant testing every student in every grade and having the test administered by the state. Instead, according to this respondent, the KERA law specified that the state should create the prototype, but then teachers would design the actual test. A legislator felt that the lack of support among legislators for the continuous assessment was partly the result of the state's unsuccessful experience with past assessments. In his view, legislators were skeptical about how well such an assessment would work and whether it would produce its intended effect. Consequently, his colleagues would not consider expanding to a continuous assessment until they saw both the scores on the accountability assessment and whether or not it was "driving the curriculum."

¹⁶ Although North Carolina's per student costs are lower than Kentucky's, the fact that there are more students in North Carolina and that it tests all students in grades 3-8 (and all high school students at the end of required courses) mean that its total testing costs are significantly higher than Kentucky's. Picus (1996) estimated North Carolina's annual costs to have averaged about \$ 9.2 million, while Kentucky's annual costs have averaged about \$6.7 million.

However, both these respondents and three others who talked about the concept agreed that the cost of continuous assessment was a major constraint, independent of how the assessment would be designed or the instructional benefits it might produce. Although \$1 million was appropriated for research on continuous assessment, it has never been made operational on a broad scale. As one legislative staffer noted, “pragmatically, we just couldn’t afford it.”

The California legislation assumed, like Kentucky’s, that in addition to the on-demand testing for accountability purposes, there would also be assessments that were embedded in the instructional process of individual classrooms and that these would include student portfolios in addition to other types of tasks. The constraint in California was less one of funding than of time. Although the state supported a research project aimed at developing a student portfolio system and also participated in the New Standards Project that was designing curriculum-embedded assessments, most of the state’s effort was devoted to implementing the on-demand testing system for accountability purposes. Talking several months after the CLAS continuation had been vetoed by the governor, a high-level SDE official described what in retrospect seemed like a tactical error:

In the public’s mind, CLAS became the on-demand test. We misallocated resources. We should have spent more resources and time at the local level internalizing the standards and developing the portfolios. When the controversy erupted, no one was paying attention to the other parts that were supposed to be in CLAS. The on-demand part of CLAS is important for accountability and pedagogical purposes. . . . [However,] the individual student scores ought to come primarily from the portfolios, not the state on-demand test. The on-demand component could be one component of Johnny’s score.

Nevertheless, the looming sunset of the legislation and the technical challenges associated with generating individual student scores meant that the SDE had little choice, but to concentrate on the state-level, on-demand assessment.

In North Carolina, funding was a greater constraint than time. The legislature appropriated about 30% less than the DPI requested for the state assessment. One consequence was that the DPI slowed development of an item bank that local school systems could use in aligning their own assessments with the state test and the *Standard Course of Study*, and that teachers could use for classroom assessments.

In all three states, the initial constraints were less overtly political than trade-offs among time, money, and the technical limitations of a new testing technology. The common response in the three states was to concentrate on the on-demand, accountability assessment at the expense of more decentralized options that could also be tied to the state standards and could rely on alternative testing formats.

Professional development for teachers was another area where the effect of time and fiscal constraints was most obvious. One of the explicit goals of these assessments was to change teaching, and several decades of implementation research indicated that such change could not occur unless teachers were given sufficient training and the time needed to adapt new approaches to their classroom routines (Fullan, 1991; McLaughlin, 1990). Yet the average teacher in the three states received very little professional development in preparation for the new assessments. Kentucky made the greatest commitment, and by the standards of what states typically spend on such capacity building, it was substantial. School districts were allowed to use up to nine days a year for professional development, and in addition, \$400 was allocated per teacher for professional development with 65% of that sum under the control of the local school site. The state and local districts also created a variety of professional development and technical assistance networks. Still, given the magnitude of changes expected under KERA, such as the transition to an ungraded primary, even this significant resource commitment was inadequate to the task.

In California, no additional resources were allocated for professional development associated with CLAS. State policymakers expected that state-funded, subject matter projects would provide the necessary assistance to teachers. These professional development networks focus on particular subjects such as mathematics and writing, and are run by the University of California. Although they are highly regarded by practitioners, they can serve only a small fraction of the state's teachers. North Carolina also did not allocate additional funding for professional development related to the new assessment. The state distributed a booklet to all teachers in the tested grades that explained the purpose of the assessment and that included examples of test items and their scoring rubrics. The state also provided some assistance to district officials through its regional technical assistance centers, and the DPI hoped that because a large number of teachers were involved in scoring the state assessment, they

would convey their experience with the scoring rubrics to their colleagues. Despite these attempts to offset the shortfall in professional development, fiscal constraints have meant that the goal of changing assessment instruction through the state test has been considerably more difficult to achieve.¹⁷ As one teacher union official noted:

The tests are supposed to deal with thinking skills and call for judgment. But who's prepared the teachers when they have been bombarded for ten years with fixed content and six-step lesson plans?

In addition to these common responses to the constraints they faced, each state also made some unique accommodations. For example, the major reason that North Carolina did not include performance items on its state assessment is that testing experts in the DPI judged them to be too unstandardized for accountability purposes and too expensive to score. They reasoned as follows:

Multiple choice gives reliability, is less costly, and means less time is spent on testing. Now it's true that some things are only measured indirectly by multiple-choice items. Therefore, we also included open-ended, constructed responses. The only thing left was whether students could do a task. But performance tests are so expensive, and they're unlikely to be standardized. We decided that they should be kept as classroom exercises.

Kentucky officials, on the other hand, were firmly committed to performance tasks and saw multiple-choice items as sending the wrong instructional message to teachers and students, even though they may be more reliable and cheaper to score. However, Kentucky also had to accommodate competing demands for available time and resources. One of its major compromises was to design and administer the assessment before the state curriculum standards were fully developed.

The Kentucky legislation mandated that the SDE develop standards and then assess students on those standards, not promulgate a statewide curriculum. Therefore, the SDE was initially wary of being too prescriptive about curricular content. It also had only a short time before the first administration of the assessment—two years during which a number of other major reform programs also had to be implemented. In 1993, after the first administration of KIRIS, it

¹⁷ The relationship between resources for professional development and local responses to the state assessment is discussed in greater detail in a separate monograph on the local implementation of new state assessments.

published what SDE officials considered to be a resource document called *Transformations*. The document outlined the state learning goals and outcomes, and it provided numerous examples of instructional activities that might be used in teaching concepts related to those goals and outcomes. However, the emphasis was on pedagogy, rather than on the curricular content to be covered. Yet *Transformations* was the only official guide that teachers had for discerning how they needed to change their teaching in response to KIRIS. Without more guidance about content, they reported having to infer what that they needed to teach from released test items. Still, there was an expectation on the part of state officials that the test should influence what teachers taught as well as how they taught.¹⁸ Eventually, after growing demands from local educators, the SDE issued a second edition of *Transformations* that included content guidelines and suggested course outlines. But this more specific guidance came only after the assessment had been administered for several years. In looking back over its implementation, one respondent involved in developing KIRIS gave this advice to other states considering the design of new assessments:

You need to make certain that the test doesn't lag the curriculum standards and frameworks too much because the test is a measure of how well teachers understand what's expected of them. Giving the test motivates them to change, but they can only do so as they understand the frameworks. If their understanding is "wrong," the scores won't improve.

By some measure, the politics of enacting new assessments in the three states was unremarkable. Because of KIRIS' tie to the historic KERA reforms, it garnered greater attention at the time of enactment than either California's or North Carolina's assessment policies. However, none of the three assessments were highly visible or contentious issues. They were viewed by policymakers as meeting identified problems, and their champions were able to leverage favorable political conditions to secure legislative passage. One goal of these policies was entirely consistent with that of past state assessments. Policymakers assumed that by testing students and then publicly reporting the scores on a school-by-school basis, parents and the general public would use that information to hold the

¹⁸ One high official in the SDE offered the following as an example of how KIRIS was meant to influence curricular content:

Physical science content is not covered in the eighth-grade course of study; most schools emphasize life science at that level. That's why so few students scored *distinguished* in science. But about one-third of the eighth-grade assessment in science deals with physical science. That message will get out and schools will emphasize physical science more.

schools accountable. The Kentucky legislature took the notion of accountability one step further, using the test results as a basis for rewarding or sanctioning individual schools. The state assessments were subject to the same kinds of time and fiscal constraints as most policies. Policymakers expected the assessments to be developed and ready for administration within a short time frame, and funding was insufficient to implement fully all the assessment components.

However, two aspects of these assessment policies make their enactment distinctive. First, they were premised on a clear set of ideas about what should be taught in each state's public schools and an expectation that a particular kind of assessment would lead to teaching reflective of those curricular values. Although there is evidence that some policymakers in California and Kentucky did not understand the full import of either the assessment or its underlying values and that some supported the new policy for reasons unrelated to its curricular implications, the legislative mandate was clear. The assessment was to measure student learning defined as the ability to think critically, solve problems, and apply knowledge in new situations.

A second factor that made these assessment policies special—though by no means unique among public policies—was that they were based on a testing technology that was as yet unproven on a widespread basis. States like California had experience with writing assessments that demonstrated that essays and other types of open-ended responses could be scored reliably and that such assessments did exert a positive influence on teaching and learning. Nevertheless, with the exception of specialized tests such as the Advanced Placement exams, no large-scale system for testing elementary and secondary students had moved as far toward alternatives to multiple-choice tests as states such as California and Kentucky were now planning. Knowledgeable policymakers understood that they were embarking on uncharted territory and they accepted the cautions of testing experts, but they also assumed that current limitations and problems could be addressed simultaneously with the implementation of new assessment strategies (McDonnell, 1994b). In sum, while the policy formulation process might not have held any apparent clues about how implementation of the assessments would proceed, their underlying values and technical novelty did suggest that implementation might not be as smooth as the enactment process had been.

III. THE POLITICS OF IMPLEMENTATION

Policy implementation is typically viewed as a set of administrative activities, rather than as a political process. The factors traditionally considered to be most important in explaining implementation outcomes are those associated with either the implementing agency (e.g., its mission and structure, how it communicates policy intent to local practitioners, the level of resources provided), or with the local context (e.g., the capacity and will of local implementors). However, analysts who focus on policy design variables and those who examine the role of “street level bureaucrats” such as teachers emphasize that implementation also has a political dimension (e.g., Ingram & Schneider, 1993; Palumbo & Calista, 1990). Other analysts have noted that conflicts or ambiguities about policy intent not resolved during the enactment process will only persist during implementation, and will constrain that process as well (Bardach, 1977).

Consequently, it should not be too surprising that in two of the three states studied, implementation of the new assessments was politicized. What is notable, however, is that it occurred after a low-key, uncontroversial enactment process. In both California and Kentucky, interest groups were mobilized, the state assessment became a more visible focus of media attention during implementation than it had been during enactment, policymakers were pressured to reconsider their initial decisions, and in California, lawsuits were filed against the SDE and several local districts. By all objective measures such as numbers of people attending meetings and rallies or the proportion of parents who forbade their children to take the state test, the politics of implementation involved only a small minority of each state’s students and parents. Nevertheless, this political activity was largely responsible for the demise of CLAS in California and it exposed KIRIS to intense public scrutiny in Kentucky. In both states, the assessment also became an issue in the gubernatorial election campaign.

This section analyzes the political dimensions of implementing the new assessments in California and Kentucky.¹⁹ It focuses on the debate over the

¹⁹ It is not entirely clear why the testing program in North Carolina did not face the kind of opposition encountered in the other two states. It may be because the test there represents a less radical departure from more traditional approaches, that it is tied to a curriculum that has been in place for some years, or that the state was careful to limit to a few pilot projects its foray into “outcomes-based education,” an approach that became a major target of opposition groups in California, Kentucky, and a number of other states.

assessments' underlying values by examining the groups that opposed CLAS and KIRIS and the differing responses of officials in the two states to that opposition. Because technical concerns about the test eventually became a factor in the political dispute, those issues are also examined. The purely administrative dimensions of state-level implementation such as technical assistance to local districts are not considered in this analysis; those factors are examined as part of a separate monograph on local implementation of the state assessments.

Opposition Groups and Their Concerns

Opposition to the California and Kentucky assessments was led by a loose coalition of conservative groups that are part of what has become known as the “religious right.” Several of these state-level groups such as the Eagle Forum and the Rutherford Institute are chapters of national organizations, while others such as the Family Foundation in Kentucky and the Capitol Resource Institute in California have

informal ties with national movements. The groups, particularly those in California, had an informal division of labor during the CLAS controversy with a few concentrating on lobbying state officials, while others filed lawsuits against the SDE and local school districts and still other groups were primarily concerned with mobilizing grass-roots opposition. In both states, local groups with ties to the state organizations also worked to mobilize parents against CLAS and KERA.

Opposition groups in both states represented a small minority of parents and teachers. The largest group in California has about 2000 subscribers to its newsletters and the largest in Kentucky, about 900. The rallies at the state capitol that the groups organized in each state drew no more than several hundred people.²⁰ Another indicator of the extent of opposition in Kentucky was that a petition to the legislature calling for KERA's repeal was signed by about 11,000 people (May, 1994a). In California, the best indicator was probably the number of students in grades in which the CLAS test was administered who were “opted-out” of taking it with their parents' permission—a protest tactic recommended to

²⁰ In both states, active opposition at the local level seemed to be concentrated in a few geographic regions. In Kentucky, opposition was centered in the western portion of the state; in California, it was concentrated in the south in Orange, southern San Diego, Riverside, and Santa Barbara counties. Although groups formed around Lexington and Louisville, Kentucky, the opposition in California was primarily concentrated in small cities and suburban areas. There was no significant opposition in large cities such as Los Angeles, San Diego, and San Francisco.

parents by the opposition groups. According to SDE records, that number equaled about 61,000 students or about 5% of the more than 1.2 million who were to be tested.

Despite their minority status, however, there is evidence from national public opinion polls that the opposition groups were tapping into concerns shared by a wider segment of parents and the public. An overwhelming majority of the American public supports having local schools conform to a set of national achievement goals and standards and to requiring that standardized tests be used to measure student achievement on those standards (Elam, Rose, & Gallup, 1991, 1992). A majority also sees raised standards as a way to encourage students, even ones from low-income backgrounds, to do better in school (Elam & Rose, 1995, p. 47). The consensus breaks down, however, in moving beyond this belief in the need for standards and assessment to questions about what those standards should be and how students should be taught and tested. Public opinion data indicate that some of the questions opposition groups raised reflect broader public concerns. For example, recent surveys about the teaching of mathematics and writing point to fundamental differences between the curricular values of education reformers and large segments of the public (Johnson & Immerwahr, 1994).²¹ In a Public Agenda survey conducted in August 1994, 86% of the respondents in that national sample said that students should learn to do arithmetic “by hand”—including memorizing multiplication tables—before starting to use calculators (Johnson & Immerwahr, 1994). This opinion contrasts with the 82% of mathematics educators responding to an earlier survey who said that “early use of calculators will improve children’s problem-solving skills and not prevent the learning of arithmetic” (p. 17). This survey also found similar contrasts in the opinions of educators and the general public on how to teach writing and in the value of teaching students in heterogenous ability groups.²² It

²¹ National public opinion data are summarized here, because no polls were conducted specifically on a California sample. Surveys were conducted, however, in Kentucky. Probably the most important finding to emerge from those polls was the public’s low level of knowledge about KERA. Despite the fact that KERA had dominated the media for close to five years and the state and private groups had sponsored numerous public information campaigns, a statewide survey conducted in 1994, found that over half of the general public and more than 40% of parents with children in public schools reported that they still had either a low level of understanding or knew nothing at all about KERA. The same survey found that less than half of the general public judged KIRIS to be working well (Kentucky Institute for Education Reform, 1994).

²² Sixty percent of those in the Public Agenda sample rejected the educational strategy that encourages students to write creatively without a prior concentration on spelling and grammar.

was that concern and seeming inconsistency between reform goals and parental expectations that opposition groups in the two states attempted to mobilize.

In California, six statewide groups led the opposition to CLAS:

- The Traditional Values Coalition (TVC) is a lobby group that claims 31,000 churches nationwide as part of its network and a \$2 million annual budget. It is based in Anaheim and has lobbied in favor of prayer in the schools and against abortion and gay rights (Zoroya, 1995).
- Capitol Resource Institute (CRI) is “a conservative public policy center focusing on state and local issues impacting families in their communities” (Capitol Resource Institute, cover letter accompanying CLAS information packet). CRI, located in Sacramento, has about 2000 paid subscribers who receive its materials; it is affiliated with Focus on the Family, a conservative Christian coalition based in Colorado (Colvin, 1995d).²³
- The Eagle Forum is a national group, founded in 1972 by Phyllis Schlafly. Like most of the other groups, its agenda extends beyond education to include minimizing governmental intervention into what it considers to be the purview of families, maintaining a strong national defense, and encouraging private enterprise unfettered by governmental regulations.
- Parents Involved in Education (PIE) is a statewide umbrella organization whose major aim is to “reinstate” parental rights. It provides information and other types of assistance to local organizations with similar goals. PIE’s leadership overlaps with that of the state Eagle Forum, and it has established networks with similar organizations across the country.
- The Rutherford Institute is a conservative legal foundation, based in Virginia, with volunteer lawyers across the country willing to litigate cases where they believe constitutional liberties have been violated.²⁴ It

Instead most respondents endorsed the idea that “unless they are taught rules from the beginning, they will never be good writers” (p. 18). Similarly, the Public Agenda poll found that “only 34 percent of Americans think that mixing students of different achievement levels together in classes—’heterogenous grouping’—will help increase student learning. People remain skeptical about this strategy even when presented with arguments in favor of it” (p. 18).

²³ Focus on the Family has been described as a “\$100-million-a-year Christian broadcasting and publishing empire” (Stammer, 1995, p. E1). Its leader, child psychologist James Dobson, reaches three to five million people weekly through his radio broadcasts and is an influential figure in conservative politics nationally. Dobson sees his mission as altering modern cultural values as they affect families, and stopping “the nation’s moral free fall” (Stammer, 1995, p. E1).

²⁴ In a *Sacramento Bee* article published several months after the height of the CLAS controversy, Delsohn (1994) reported that more than 200 California lawyers have volunteered to assist the Rutherford Institute in its work for Christian and other related causes. He also reported that the national office raises more than \$6.5 million a year in direct mail solicitations and that it produces radio messages that are broadcast on more than 1000 local stations.

has a particular focus on parental rights and the exercise of religious freedom.

- The United States Justice Foundation (USJF) is also a conservative legal foundation. It was founded in 1979 and is based in Escondido, California. Although it has dealt with other education issues such as school desegregation, its past cases have been wide ranging and include ones dealing with tax issues, term limits, union decertification, and health care. Its basic aim in all these cases is to limit the scope of government.

In Kentucky, there were three major statewide groups opposing KIRIS:

- Parents and Professionals Involved in Education (PPIE) is a grassroots organization founded in 1990 specifically around issues related to KERA. The group holds workshops, distributes information, and is involved in various types of lobbying activities. With its subscriber base numbering about 900 statewide, PPIE is “probably the oldest and largest of citizen groups critical of the reform act” (Associated Press, 1994).
- The Eagle Forum in Kentucky is similar in its goals and strategies to the state chapter in California.
- The Family Foundation is a conservative organization, founded in 1989 to deal with issues related to the family. Although it has been a constant critic of KERA, the foundation has also focused on welfare, pornography, health care, and crime issues. It has an annual budget of about \$170,000 that comes largely from donors within Kentucky (Schreiner, 1996).²⁵

Although PPIE is the largest of the grassroots groups focused on KERA, there are several other smaller ones such as America Awaken, a Paducah-based group that circulated the petition calling for the repeal of KERA. A similar group in Owensboro, Families United for Morals in Education, regularly opposes KERA in newspaper advertisements (Associated Press, 1994). In October 1994, the three major opposition groups formed an umbrella lobbying group called Campaign for Academic Renewal in Education (CARE) and proposed an alternative to KERA centered around standards emphasizing traditional academic content, more reliable state tests, greater local control, and a voluntary nongraded primary program. It launched its campaign to get KERA modified during the legislature’s 1996 biennial session with a pamphlet asking “Why does this man [Shakespeare]

Although most of its cases involve Christians, the Institute has also defended Hare Krishnas, Orthodox Jews, and other non-Christians.

²⁵ The Family Foundation is the only one of the Kentucky opposition groups that has paid staff (two full-time and three part-time) in addition to volunteer assistance. The other groups are run entirely by volunteers. Because the California groups have more resources, most of them are able to employ some staff.

feel left out? Because he was excluded from state education goals. Give the Bard a break.”

Despite some differences in their size and numbers, opposition groups in California and Kentucky had very similar concerns about the state assessments. These concerns fell into two broad categories dealing with the values underlying the assessments and issues related to their technical quality. Opposition groups argued that the state assessments embodied inappropriate values because the outcome standards were not sufficiently academic, that they represented a social and cultural agenda that was offensive to some parents, and that they intruded unnecessarily into family life by asking students personal questions and thus violating parental rights.

Opposition to the state assessment was somewhat stronger in California, and focused primarily on the language arts portion of the test. Most of the reading passages that evoked opposition were from works by authors of color such as Richard Wright, Dick Gregory, and Maxine Hong Kingston. Critics charged that many of the selections on which students were tested promoted inappropriate values such as violence and the questioning of authority.²⁶ The selection that caused the greatest controversy was “Roselily,” a short story by Alice Walker that was included on the tenth-grade CLAS test in 1993. CLAS opponents argued that the story about a Black woman’s thoughts during her marriage ceremony to a Black Muslim was disrespectful of religious people. The kinds of questions that most upset critics were language arts prompts asking students to discuss family matters (e.g., “Why do you think some teenagers and their parents have problems communicating?”—a question on the eighth-grade exam); to question authority (e.g., “Think about a rule at your school—on the playground, in the cafeteria, or in the classroom—that you think needs to be changed. Write a letter to the principal about a school rule you want changed”—a fourth-grade question); or to delve into people’s psyches (e.g., an “open mind” question showing a picture of a head with nothing in it and students asked at the end of a literature passage to fill the head

²⁶ In some cases, concerns about the test became rather exaggerated. Perhaps one of the strangest instances was when two members of the State Board of Education proposed deleting an excerpt from Annie Dillard’s “An American Childhood” because it was about children throwing snowballs and therefore, might be viewed as violent. The same two board members argued against including Alice Walker’s “Am I Blue?” because it could be interpreted as advocating a vegetarian lifestyle. They maintained that the “anti-meat-eating” excerpt should be excised from the test to protect the sensitivities of rural students whose parents might raise poultry or livestock (Ingram, 1994). The State Board subsequently reinstated the passages after being accused of censorship by civil libertarian groups.

in with symbols, images, drawings and/or words showing what the character was thinking).²⁷

The points of controversy in Kentucky were less on specific passages and more on the questions asked of students and on what critics perceived as biased scoring rubrics. One example that several opponents cited as an inappropriate question came at the end of what they considered to be “charming” reading passage for fourth graders entitled “Your Dad’s a Wimp?” The question asked, “Would you like to be part of Jesse’s family? Why or why not?” One of the opposition group leaders described how her ten-year-old daughter reacted to the question:

. . . the open-ended question made my ten-year-old go nuts. She assumed that she wouldn’t be in her own family any more when the test asked if she would want to be a part of Jesse’s family. I had to ask the school to destroy my daughter’s test. When you’re testing ten-year-olds, you shouldn’t be investigating their attitudes.

Another opposition group leader made the same point about the “Jesse” question, but expanded it to argue that the scoring of this and other items reflected a bias:

. . . the state doesn’t make the scoring guide until they get all the answers in. That doesn’t give parents any confidence. The state says that whether students answer the question yes or no doesn’t make a difference. But then they include on the scoring guide an answer that goes against traditional family arrangements. If a student answers that she wants to be part of Jesse’s family, then she is saying that she wants to leave her own family. Also, a student doesn’t really know what it would be like to be part of Jesse’s family.

There’s a larger concern that there’s a bias in the scoring. That’s usually true in reading and social studies. Although in science, there’s so much about environmental subjects. Fourth graders were asked what to do about the problems of the homeless. Students have to give an opinion off the top of their heads; they really wouldn’t know unless they did some research.

²⁷ One of those responsible for developing the language arts portion of the CLAS assessment explained why the expert advisory committee that designed the test chose the kinds of passages and questions that it did:

First, it was because the student response was wonderful. These were pieces that are regularly taught in classrooms. We wanted passages by quality writers. A large proportion of the authors on the CLAS test were award winners. They also caught students’ attention and they had the complexity we were looking for as a good way to test reading achievement. Second, there was a great effort to make the test diverse. What we learned is that it’s hard to bring together a diverse state.

Similarly, the president of the Kentucky chapter of the Eagle Forum charged in an op-ed article that not only do KIRIS writing prompts require students to choose sides on controversial subjects such as euthanasia and citizens owning firearms, but the scoring rubrics advantage students whose answers reflect pro-multiculturalism, pro-women's rights, and pro-environmental positions (Daniel, 1995). In an *Education Week* article, she was quoted as describing opposition to KERA as a "classic case of cultural warfare" (Harp, 1994).

The following excerpts from interviews with several of the opposition leaders in the two states amplify these broad concerns:

We were opposed to it [CLAS] for two reasons. The first concern was the one that had a tendency to catch the attention of the media. It was that the test seemed to be delving into areas of students' psyches. The types of questions and the types of stories were emotional ones. It appeared that the people making the test had a philosophical agenda. The test was an invasion of student and family privacy.

But there was a much deeper concern that we didn't see until we delved into the test much more. CLAS was not challenging enough for kids academically. The standards being tested were unclear and not rigorous. I would define rigorous as testing kids for a knowledge of facts, but more than just a regurgitation of facts. A rigorous test would also challenge students' intellectual abilities. . . . We've been told by a math expert that in the CLAS test, basic skills were skipped over, but students were expected to do problems that college physics students can't complete . . .

There's a tendency for a lot of people involved in education to believe in helping children develop. They want to get kids excited, and one way to do that is to get them writing about their personal feelings. These are people who are pushing the edge of the envelope about where education is going, and about educators playing a parental role. In the right circumstances, questions like "what makes you happy?" are all right. But teachers are not trained to be therapists and classrooms aren't meant to be therapy sessions. You need someone who is trained to bring closure to such discussions, and parents have to be involved. I'm not certain that CLAS was intended to be the "great therapy session," but that's how it turned out. (California opposition group leader)

☺ ☺ ☺

They [the state tests] need to give parents something to evaluate. There was nothing published that would evaluate students' ability in language skills. It didn't test their ability to interpret the author's intent, write grammatically, spell correctly, or to punctuate. It didn't show a concrete level of mastery of reading and writing. This was a test that was supposed to be an individual student assessment that would test their basic skill in language arts.

But what the language arts section was really doing was using a situation prompt in order to enlist a student's ability to draw from a personal experience a similar situation. We realized all of the stories that were being used to draw from an experience were from a very negative content. There is nothing concrete being academically measured . . .

When asking students to relate to their own personal experience that could possibly be another way of gathering data on the life of the family. The eyes of our children will become the window of our household . . .

Our motto is "We want ABC (Academic Based Curriculum), not OBE [outcomes-based education]." We want mastery of academics, not student attitudes. (California opposition group leader)

ღ ღ ღ

We would like to see more emphasis on basic skills in teaching and less emphasis on creating new cultures. And we would like to have the teachers not usurp the rights of parents by deliberately trying to educate their students to another point of view.

The stories that I read all show the father in a very poor light. It would make me wonder if there was some sort of feminist thing here trying to show that men are bad. The fathers all did ridiculous things or they abandoned their children. And I guess the question is if we want to raise a generation of children that has the values that you work hard and value intelligence, how do you do that? Do we do it by showing the lowest actions of people where the father is mean to the children and tells them to kill the cat and they do? Is that how we want to raise them? Or do we teach by using more of the stories from "Fathers Knows Best?" Do we do this by promoting the role of responsibility in young men or by negating men from our culture? I think that these stories have already brought about some destructive things in our culture. (California opposition group leader)

ღ ღ ღ

Our number one concern with KERA is that it narrows the curriculum. It dumbs-down the curriculum; the less is more philosophy. . . . The outcomes are intentionally vague; they haven't communicated anything to teachers, students, and parents. There's all kinds of leeway for mischief. For example, *Transformations* has as one of the suggested activities for middle school students that students join or contribute to Act Up, NOW, or UNICEF. The department of education heard a lot of noise about that and Tom Boysen [the state commissioner of education] labeled it "an unfortunate mistake." Also the vagueness of the outcomes lends itself to mediocrity. (Kentucky opposition group leader)

Most of the groups framed their state-specific concerns in the larger context of opposition to outcomes-based education (OBE) and the federal Goals 2000 program. OBE is an education strategy designed to focus instruction on helping students master a specified set of skills or outcomes. It took on political overtones and became a target of conservative groups because the outcomes specified by SDEs and local school districts included not just academic goals, but also ones dealing with students' attitudes and their psychosocial development. The OBE controversy escalated in 1991 when conservative groups protested Pennsylvania's draft "learning outcomes." Critics argued that it was inappropriate for government to require that students demonstrate their ability to bring about and adapt to change in their own lives, or that they develop personal criteria for making informed moral judgments. The critics maintained that while they strongly supported students developing such skills, they feared that such outcomes would lead to the state specifying what behaviors were ethical and that students might be barred from graduation if they did not demonstrate the right attitudes. The critics distorted the state's intentions, but the vagueness of the outcomes and the implicit political content of some (e.g., dealing with the environment) provided a platform for the opposition (Ravitch, 1995).

A videotape featuring the leader of the Pennsylvania opposition, Peg Luksik, was used by opposition groups in both California and Kentucky to argue that what they viewed as the weaknesses of CLAS and KIRIS were specific instances of a larger national problem. In addition, materials such as a publication produced by Focus on the Family, entitled *Not With My Child You Don't: A Citizens' Guide to Eradicating OBE and Restoring Education*, were widely available to opponents and their supporters. In Kentucky, the anti-OBE crusade was aimed specifically at two of the state's "valued outcomes" that dealt with self-sufficiency and responsible group membership. A number of the outcomes specified under these two goals resembled the ones that had become controversial in Pennsylvania, and critics were particularly concerned about outcomes requiring students to demonstrate "an understanding of, appreciation for, and sensitivity to a multicultural and world view" and to demonstrate "an open mind to alternative perspectives." In talking about her group's opposition to OBE and its link to CLAS, Carolyn Steinke, the president of PIE, noted that OBE "is something that is difficult to explain, yet so dangerous" (Walsh, 1994).

Goals 2000 is a target of opposition groups because of its assumed link to OBE and because critics perceive standards-based reform as eroding local control and strengthening the hand of government in the raising of children (Pitsch, 1994). In the literature it distributed, the Capitol Resource Institute argued that “Goals 2000 is using CLAS to link California’s students with universities and jobs.” It then went on to imply through a rather circuitous set of assertions that information on students would be available across states in a computerized data base, and that “fulfilling Goals 2000, this computer identifier will link students to employers through WORKLINK.” The CRI document then poses these questions:

The key question is this: what will happen to students who are not politically correct? Will they be allowed to go on to college and compete for top jobs, or will they be discriminated against for not accepting the mandatory multicultural mindset of the schools? Will students who fail to measure up to the CLAS standard be graded lower and be refused entrance to college? These are not strange questions when we’re dealing with a completely new set of academic values such as that found in CLAS. (Capitol Resource Institute, 1994)

Although most of opposition groups’ concerns centered on the assessments’ underlying values, they also raised questions about the technical adequacy of the tests. They argued that the performance levels were vague and not understandable to lay people. There was also a widespread perception among opponents that the scoring of the state tests was subjective and unreliable:

CLAS couldn’t give individual-level scores because the tests take too long to score. We didn’t believe the SDE when they said they could produce individual level scores, because those scores are just too subjective. (Opposition group leader)

We were told in the hearings before the Assembly Education Committee that the state has 70% accuracy in scoring the CLAS test. Well, 70% is a C in my book. It means that 30% of the students have been misgraded. (Carolyn Steinke speaking at a community forum in Los Olivos, California, 7/13/94)

The test is all subjective; the department of education and the testing contractor can do whatever they want. The scores were all higher just before the [1994] election, but I don’t think that will take the edge off. (Kentucky opposition leader)

Some of the opponents’ concerns about the technical adequacy of the state tests evidenced a misunderstanding of how large-scale assessments are scored, but as

the technical reviews of CLAS and KIRIS would show, some of the weaknesses identified by assessment critics were confirmed by testing experts.

One of the issues that was raised in both California and Kentucky by those in the policy community who had to respond to the opposition groups was the extent to which the groups' concerns were sincere or were part of a strategic move to advance a broader political agenda. A number of policymakers argued that criticisms of CLAS and KERA were merely a component of a larger strategy to promote school vouchers. In California, Proposition 174, an initiative to fund vouchers that was supported by many of the same groups opposed to CLAS, had been rejected by the voters the previous year. Similarly, in Kentucky, groups such as the Eagle Forum had discussed voucher legislation with members of the General Assembly. When confronted with a public charge by former Governor Wilkinson that they were criticizing KERA as a way to get tax money for private school tuition, Martin Cothran of the Family Foundation and Donna Shedd of the Eagle Forum argued that the two issues were separate. Shedd maintained that although she likes the idea of school choice, that did not mean that she did not care about public education. Cothran noted that while the Family Foundation would probably be in favor of vouchers, it had not taken an official position on the issue (May, 1995a). However, both Shedd and another leader of the KERA opposition are active in the home school movement in Kentucky. The situation was similar in California with opposition group leaders acknowledging that they were active in the voucher movement, but seeing their opposition to CLAS as a separate issue. The response of one group leader exemplifies this perspective:

I was personally for [Proposition] 174 and spoke publicly in favor of it. [But] I don't think it would be fair to say that CLAS opposition was a transfer of causes, because the CLAS group was much broader than the Prop. 174 group. But it would be naive to say that some groups weren't involved in both because I knew they were.

The question of sincere versus strategic motives raises a related issue that also cannot be resolved unequivocally—viz., whether the criticisms made against CLAS and KIRIS were valid. Answering that question is difficult because so much depends on the religious, political, and cultural values one holds. Clearly, political and cultural conservatives, whether or not they are religious, will arrive at a different judgment than those who are more liberal in their beliefs. Nevertheless, it is possible to sort out some of the criticisms, separating those that are clearly false or exaggerated from those that have merit. For example, some claims, such

as the one made by a KERA opponent charging that the education reform act promotes the teaching of witchcraft, border on the preposterous (Ward, 1993).²⁸

Other claims were exaggerated or distorted, although it is not clear whether the misrepresentation was intentional or not. One example was the portrayal of offensive test items by CLAS opponents as being on the CLAS assessment when they were not.²⁹ One particularly egregious example was used in literature distributed by Capitol Resource Institute and in a special radio broadcast that Natalie Williams, CRI's lobbyist, did with James Dobson. This item was never on the CLAS test, but was presented as such in a CRI publication:

One math question goes like this: There are seventeen apples and four students. How would you divide the apples and explain why. The state-desired response that would score the highest (6 on a scale of 1 to 6) was this: Four apples go to three students and five apples go to the fourth, because the fourth student was the most needy. This clearly reflects a value judgment that has no basis in mathematical problem solving, unless that is what they mean by "the new math." (Capitol Resource Institute, 4/25/94)

Opponents admitted that some items that got into the "pot" of public debate were not official CLAS items, but they argued that even if an item was only a sample from a field test or a locally-developed one to prepare students for the test, it should be included as part of the CLAS system.³⁰

²⁸ This charge was made by the Rev. Joe Adams, a KERA opponent and chair of the Christian Home Educators of Kentucky. He argued that KERA mandated ethical values and beliefs that were not those of most Kentucky parents. He also said that ignoring Christian views of Christmas in the schools while celebrating Halloween amounted to an endorsement of witchcraft (Ward, 1993).

²⁹ The question of which items publicized by the opposition groups were really from the CLAS test was complicated by the fact that the SDE released very few items because it hoped to save money by using them in successive administrations of CLAS over several years. In addition, those who saw the test, such as teachers administering it, were not supposed to discuss its contents or distribute copies to others. However, several opposition group leaders admitted that they received copies of the test from teachers upset about CLAS or that copies were left anonymously in their offices or sent to them through the mail. One group was also given a complete set of the tests by a policymaker sympathetic to their cause who had obtained them from the SDE. In addition, once the opposition groups filed lawsuits, they collected sworn statements from students about the test content as they remembered it from the previous year's administration.

³⁰ One example of an item on a locally-developed tenth-grade sample test, but not on the official CLAS assessment, was a story by Hernando Tellez entitled "Just Lather, That's All." It includes an incident where a barber contemplates killing a murderer who has come in for a shave. The barber does not kill the murderer, and the story ends with the murderer saying, "They told me you would kill me. I came to find out. But killing isn't easy. You can take my word for it." Critics portrayed this story to the media as an example of a CLAS item that abandoned academics to focus on subjective, personal, and invasive material (Hageman, 1993). The

Another particularly distorted criticism of the test was the notion that students' responses would be computerized and made available to other government agencies, businesses, and individuals. In California, assurances from the SDE that information about students' ethnicity, family income, and their responses to test questions would be kept completely confidential were not believed by opposition leaders. Rumors started that students who were angry with their parents and who had written negatively about them had prompted calls to the family home by child protective services workers. Although it is difficult to identify the source of such rumors, it appears that in a few cases the teachers who were scoring the exams suspected possible child abuse, based on students' answers. Since teachers are legally required to report such abuse, they told their SDE supervisors and SDE staff then reported it to the principals at the students' schools. The assumption was that the local site administrator would investigate and decide whether legal authorities should be contacted. To the best of the SDE's knowledge, very few cases were ever reported to principals, and Child Protective Services never intervened as a result of the CLAS test. But as an SDE official noted, even if principals were contacted about responses on only one-hundredth of one percent of the CLAS exams, that represented more than 100 students. As we will see in subsequent sections, this rumor unique to California might have been avoided if the test development and administration had been more open to public participation and understanding.

Despite these examples of opposition concerns that either exaggerated or distorted the actual situation, the tests did appear to highlight a set of social values in addition to measuring academic skills. In a few cases, scoring judgments were also made that seemed to contradict parents' sense of what constituted an appropriate answer.³¹

Riverside Press-Enterprise, which initially ran a story about the item, subsequently printed a correction indicating that the SDE had never considered including this story on CLAS. The paper also noted that parents throughout Riverside county were outraged that the story might be used on the test.

³¹ One example that seemed to conflict with the public's notion of what constitutes acceptable academic performance became a topic at a legislative hearing. In solving a problem on the KIRIS test, a student multiplied four times four and got 17 as the solution. The student was able to explain the rationale behind his solution and the response was scored as "proficient." However, one Democratic legislator questioned whether KIRIS was measuring a student's ability "to shoot the bull" more than the academic content the student had mastered. Thomas Boysen, the state commissioner of education, promised to change the scoring rubrics to reflect this concern, saying that precision was important (May, 1994b).

A review of the first edition of *Transformations* (1993) indicates that joining Act Up was not a suggested activity for middle school students, contrary to what one opposition leader charged. Most of the community activities proposed were ones that the majority of the public would view as commendable (e.g., participating in community service projects such as Habitat for Humanity or the Red Cross blood drive; producing a video to explain school rules). However, the political groups listed as “learning links” or community resources are all liberal groups such as the ACLU, the NAACP, labor unions, the Sierra Club, and Amnesty International. None of the political organizations with which the opposition groups are affiliated were included in the original edition of *Transformations*. Similarly, a number of activities were listed where the underlying principle would likely be acceptable to people of different social and political beliefs, but the examples given reflected only a liberal viewpoint. For example, a high school activity involving a debate over rights gave as an example animal rights and the right to die. Such an activity might be more acceptable to conservatives if the list of examples also included a right valued by them such as property rights. Likewise, a middle school activity that had students interviewing people whose lives were affected by a social movement listed as examples the ERA and civil rights movements. A more inclusive list might also include religious fundamentalism as an example of a social and political movement that has transformed the lives of its adherents.

Achieving a balance among diverse perspectives was also an issue for CLAS. As a number of commentators have noted, the literature selections and test items that opponents found offensive represented only a small proportion of all the content included in CLAS (Wilgoren, 1994). Nevertheless, even CLAS’ strongest supporters understood in retrospect that the language arts portion appeared to reflect a narrow range of lifestyles and human emotions:

I had never read through the entire body of [language arts] tests, until I had to meet with the Legislature, so I sat down and read all of the tests. . . . My response when I sat down and looked through all of them was, “Oh my God, they are all so much about serious problems.” It was all good literature, and I believed in all the readings that we had. But the whole collection seemed so dark. I thought it was a shame. I don’t know if anyone looked at all of them. (Language arts assessment consultant)

Perhaps more significantly, CLAS also represented pedagogical approaches that a much broader segment of the public than just religious conservatives began

to question. That questioning was most evident in the debate over the use of “whole language” pedagogy which deemphasizes the teaching of phonics and emphasizes the use of literature in reading instruction. Although conservatives argued against the deemphasis on phonics in the state curriculum frameworks, it was Marion Joseph—a top aide to the liberal, former state superintendent of public instruction, Wilson Riles—who was the most influential voice in arguing that the state’s move to whole language instruction had gone too far and was not working (Colvin, 1995e). Joseph’s efforts, coupled with low reading scores on CLAS and the National Achievement of Educational Progress (NAEP) and arguments from classroom teachers that a balance had to be struck between traditional methods and whole language, led to a change in the state’s recommended approach to reading instruction (California Reading Task Force, 1995).³²

Like the findings from the Public Agenda survey, the debate over whole language in California suggests that the opposition groups were tapping into concerns shared by a broader segment of parents and the general public. There is no question that the debate over CLAS and KIRIS was infused with misinformation, but when all the exaggerations and distortions are stripped-away, there still remains a core of concerns that were viewed as legitimate by people outside the small group of vocal opponents.

Opposition Group Tactics

The substance of their concerns helped advance the opponents’ cause, but their tactics were also instrumental in ensuring that their influence exceeded what might have been predicted, given their numbers. In both Kentucky and California, the opposition groups relied on lobbying and grassroots mobilization to promote their agenda. The work of the Rutherford Institute and the United States Justice Foundation in California, added a third tactic there, the filing of lawsuits.

Opposition groups in both states testified before legislative committees and the state board of education, organized rallies at the state capitol, and lobbied individual legislators. In California, CRI and the TVC were the two opposition

³² Public concern about the reading curriculum grew more intense after fourth-grade NAEP scores in reading showed that in 1993, California ranked ahead of only Mississippi, the District of Columbia, and the island of Guam among the fifty states and the territories. NAEP reported that more than half of the students tested could not understand simple texts, identify themes, or summarize what they had read. CLAS scores identified similar problems in students’ reading achievement (Colvin, 1995a).

groups most focused on lobbying; while in Kentucky, all the opposition groups were involved in lobbying activities. The opposition groups primarily worked with Republican legislators in both states, and some endorsed electoral candidates. For example, in the 1994 election, PPIE in Kentucky endorsed four congressional candidates and approximately 40-50 candidates in state legislative races. Although a few Democrats were acceptable to PPIE, the group decided not to endorse any Democratic candidates publicly because they feared that its endorsement might get the candidates in trouble with the Democratic leadership. PPIE also checked with all the other candidates it planned to endorse to make certain that such an endorsement would not hurt the candidate in his or her own district.

In California, the opposition groups worked closely with each other, but did not coalesce with any traditional education interest groups. In Kentucky, however, the Family Foundation joined with the Kentucky School Boards Association, the Kentucky Congress of Teachers and Parents, the Kentucky Association of School Administrators, and the Kentucky Association of School Superintendents in sending a letter to the State Board of Education outlining their concerns about KIRIS. The letter listed three major problems with KERA: a lack of public confidence and ownership in the testing system; the need to ensure that it is fair and that the state is providing incentives to improve teaching and learning; and a resolution to the technical problems related to KIRIS' design, administration, and scoring. In explaining why the four groups representing key segments of the education establishment had included the Family Foundation in their letter, they noted that their recommendation to have an open process for making decisions about the test would not have been credible if they had excluded a group that agreed with their position on this particular issue (May, 1995b).

Grassroots mobilization took a variety of forms. Representatives of opposition groups spoke in local churches, community forums, and at "coffees" arranged by local groups in individual homes. One leader of a Kentucky opposition group reported that she had logged 100,000 miles on her van traveling around the state. One of the most effective tactics that groups used to publicize their position to audiences beyond their traditional constituents was use of the media. The Kentucky groups were particularly effective in gaining visibility through the general news media. Martin Cothran of the Family Foundation wrote op-ed

articles that were printed in several of the state's major newspapers.³³ Donna Shedd, vice-president of the state Eagle Forum, wrote four opinion pieces, and sent them to all the state's daily and weekly newspapers. A number of those newspapers printed one or more of the pieces. Cothran and Shedd also became the opposition spokespersons that journalists contacted when they needed a critical response to various KERA developments. Consequently, they and the positions they represented became quite visible across the state. In addition, other opponents wrote letters to the editor, reiterating the groups' position in terms of specific incidents or the comments of KERA supporters.

Probably because CLAS was a considerably less visible policy than KERA and because CLAS was the subject of intense controversy for only a year, there was significantly less media coverage in California than in Kentucky. Nevertheless, opponents used some of the same tactics as in Kentucky, particularly the use of local letters to the editor. A few opposition leaders such as Gary Krep of USJF, Carolyn Steinke of PIE, and the Rev. Lou Shelton and his wife Beverly of TVC became the spokespersons whom the media quoted in CLAS stories.

Although the California groups may have been less successful in using the secular media than their counterparts in Kentucky, the CLAS opponents had an additional tactic in the form of legal challenges.³⁴ The US Justice Foundation filed lawsuits against nineteen school districts seeking to block administration of the 1994 test, and the Rutherford Institute filed a similar number. These suits argued

³³ Cothran was credited by other opposition leaders with effectively focusing their arguments on issues of broad interest to the general public, such as getting schools to emphasize a traditional academic curriculum. Cothran's communication skills were evident in the CARE platform that helped transform the image of the opposition groups from that of extremists accusing the state of encouraging witchcraft to one of responsible citizens who were espousing greater academic rigor in the schools.

³⁴ One of the opposition leaders in Kentucky explained that the groups there would like to have a legal fund, but do not have the resources. They had been talking with a lawyer representing the Rutherford Institute in Kentucky, and were hoping to be able to file a suit if any districts began to use the twelfth-grade KIRIS test as a condition for high school graduation.

Only one lawsuit was filed in Kentucky by a family who wanted their children exempted from taking KIRIS because they argued that the test "promoted environmentalism, animal rights, feminism, multiculturalism, socialism, humanism, evolution, liberalism, and gun control," as well as overstressing "imagination," "negativism," and "depressive content." The case was dismissed in circuit court, with the judge considering each of their objections separately and finding no basis for them. In response to their charge that the test lacked traditional content, he wrote: "From Appalachian craft to the inner cities, from stories of Lincoln to the strengths of American industry, there evolves a glowing account of a country teeming with opportunity and liberty for all persons regardless of their race, color, creed, or religion" (Associated Press, 1996).

that at least some of the items on the CLAS test were covered under section 60650 of the California Education Code which states:

No test, questionnaire, survey, or examination containing any questions about the pupil's personal beliefs or practices in sex, family life, morality and religion, or any questions about his parents' or guardians' beliefs or practices in sex, family life, morality and religion, shall be administered to any kindergarten or grade 1 through 12 inclusive, unless the parent or guardian of the pupil is notified in writing that such test, questionnaire, survey, or examination is to be administered and the parent or guardian of the pupil gives written permission for the pupil to take such test, questionnaire, survey, or examination.

Consequently, according to the lawsuits, if school districts did not obtain permission from all parents prior to administering CLAS, they were violating the law. Parents who could serve as named plaintiffs were recruited through local opposition groups with the promise that either USJF or Rutherford would provide pro bono legal assistance. School boards that wished to oppose the state were also offered such assistance. In addition, parents who contacted Capitol Resource Institute or PIE and wanted to take legal action were referred to the two legal aid foundations.

The CLAS opponents lost their cases in court, with the first ruling coming in a suit brought by two parents with children attending school in a suburban area of the Los Angeles Unified School District. The judge in that case ruled that the CLAS materials “are obviously designed to elicit analytical comprehension and writing abilities . . . The questions are not designed to elicit the prohibited information. More importantly, they, in fact, do not call for a revelation of the pupil's parents' or guardians' beliefs and practices on the prohibited subjects” (Merl, 1994). That ruling set a precedent that allowed districts to avoid having to go to every parent to obtain permission before the CLAS test was administered.³⁵ However, several weeks before this ruling, a San Bernardino Superior Court judge had ruled that individual parents could have their children excused from or “opted out” of taking the CLAS test. The SDE subsequently agreed to such an option allowing districts to excuse students from taking the test if their parents objected to its content. As indicated in the previous section, that option was exercised by

³⁵ In addition to the legal action initiated by the opposition groups, the school boards in five of the state's 1000+ districts decided not to administer CLAS, arguing that while they were required by law to administer state tests, the test did not have to be CLAS. The SDE subsequently had to take legal action against recalcitrant districts.

about 5% of parents of students in tested grades. Various opposition groups provided many of those parents with the forms to opt their children out of the exam.

Explaining State Responses

For a variety of reasons, primarily related to differing political incentives, officials in California and Kentucky responded to the opposition in different ways.

The California Response

Because of the sunset provisions in the original CLAS legislation, the assessment had to be reauthorized during the height of the controversy. Gary Hart, the chair of the Senate Education Committee, was responsible for guiding the reauthorization through the Legislature. He made a number of changes in the reauthorizing legislation (SB 1273) in response to the demands of the opposition groups and their legislative allies. For example, SB 1273 contained a prohibition against asking students about their sexual, moral, and religious beliefs or those of their families. It also added parents and other members of the public to the teams that would develop the assessments, and it established a citizens committee, appointed by the Legislature, to review the final test. SB 1273 also required that the CLAS test contain a mix of performance assessments, multiple-choice items, and short-answer items that measure “building block” skills as well as higher order skills. In addition, the bill required that the SDE distribute sample exams as broadly as possible so the public would know what types of questions had been asked.

After the bill had passed the Senate and was awaiting passage in the Assembly, Governor Wilson used his power of the line-item veto to veto CLAS’ 1995 appropriations, arguing that the testing program had to be reformed and that SB 1273 did not meet his criteria. In September 1994, after SB 1273 had passed both houses despite opposition from conservative Republican legislators, Wilson vetoed the measure. In his veto message, Wilson argued that SB 1273 departed from the vision of CLAS embodied in the original legislation. Rather than making the generation of individual student scores the first priority of CLAS,

SB 1273 takes a different approach. Instead of mandating individual student scores first, with performance-based assessment incorporated into such scores as this method is proven valid and reliable, it mandates performance-based assessment

now and treats the production of individual student scores as if it were the experimental technology—which it clearly is not. In short, SB 1273 stands the priority for individual student scores on its head . . .

Performance-based assessment—if well-designed—can play an important role in a system in which multiple measures are used to accurately assess *each student's* academic achievement. But under SB 1273 it remains only a too-long deferred good intention—only a “goal,” and one not to be achieved until “the spring of 1999.” That’s not good enough. It’s too long to wait. (9/27/94)

Wilson then reaffirmed his commitment to a testing program that produced “valid, reliable individual-level student scores; objective world-class performance standards; and an appropriate mix of questions designed to assess students’ mastery of basic skills, as well as more sophisticated abilities to apply knowledge.” He urged the Legislature to give its highest priority to enacting such a program.

Several months after the governor’s veto, the 1994 CLAS results were released, showing extremely poor achievement among those students who were tested. For example, more than 40% of the tenth graders tested in mathematics scored at the lowest of the six performance levels, meaning that they had “little or no mathematical thinking and understanding of mathematical ideas.” In fact, the overwhelming majority of tested students scored at one of the three lowest levels in mathematics; the reading scores were only slightly better with 60% scoring in the three lowest levels. The newly-elected state superintendent of public instruction, Delaine Eastin, immediately appointed two task forces to examine the mathematics and reading curricula and to recommend how they might be improved. Debate continued not only over the most productive way to teach mathematics and reading, but also over whether the curriculum embodied in the state frameworks and CLAS had ever really been implemented in classrooms and thus given a fair test of its effectiveness (Colvin, 1995b).

Both task forces determined that a more balanced approach to instruction was necessary. In announcing the task force recommendations, Eastin argued that:

Both task forces concluded that many language-arts and math programs have shifted too far away from direct-skills instruction. . . . Both send a clear message that students need basic skills as well as more complex analytical and problem-solving skills. . . . Certainly we need to teach phonics, spelling, grammar, and computation skills, including adding, subtracting, multiplying, and dividing.

However, there was nearly universal agreement that no one is advocating simply returning to the repetitive skill and drill approach. (as quoted in Diegmüller, 1995)

During this same period, two liberal Democratic and one Republican member of the State Assembly introduced legislation requiring that spelling, phonics, and basic computation be taught in the state's schools. As one news report noted at the time, "the fact that legislators believe it necessary to make such pronouncements is probably more significant than the pronouncements themselves" (Colvin, 1995c). In addition to this symbolic action, the state Legislature also enacted a new assessment program, and over the opposition of conservative groups, the governor signed the legislation. The legislation was nearly defeated in the Legislature and Wilson, then running for president, waited until the last possible day to sign it. However, the new testing program was supported by the state Business Roundtable, representing the largest businesses in the state, and by the school boards association, the PTA, and the teachers' unions. The new assessment system has two tiers, with the first one allowing local school districts to use a commercial, "off-the-shelf" basic skills test for students in Grades 2 through 10. This testing is voluntary: Local districts receive \$5 per student if they choose to participate, and they can select from among several tests approved by the SDE and designed to produce comparable scores across districts. The second tier consists of a mandatory statewide test of applied academic skills, produced by an outside contractor, and administered to students in Grades 4, 5, 8, and 10. This system has not yet been implemented and for the past two years, California has had no statewide assessment.

The new test, to be called the California Assessment of Academic Achievement, will also contain all the safeguards that had been included in the earlier SB 1273. School districts will be required to exempt students from the assessment upon request of their parents; parents and the general public are to be included in all phases of design and implementation, including development of the assessment instruments; and a six-member Statewide Pupil Assessment Panel will be charged with reviewing the test and reporting its findings to the State Board of Education. However, the enactment of a statewide assessment that more closely fits the demands of both the governor and the CLAS opposition groups has not quelled the debate over curricular values in California. The controversy continues, revolving around the adoption of new mathematics textbooks and whether the curriculum stresses higher order skills at the expense

of basic, computational skills. Now, however, the opposition is spearheaded by well-educated parents questioning a set of academic, professional priorities, rather than the cultural and political values that dominated the earlier CLAS debate (Sommerfeld, 1996). Nevertheless, some of the same themes continue to characterize public discourse: whether there is sufficient evidence in support of new curricular approaches; what is the appropriate balance between a traditional and a more progressive curriculum; and how much weight should be accorded to professional judgment as compared with parental and community preferences.

The obvious question that emerges after a review of the California case is how could an assessment, heralded by education reformers and curriculum experts as the most innovative in the country, be decimated so quickly by groups representing a minority of the state's parents and voters? Five factors shaped the state's response to the CLAS opposition and help explain the policy's eventual fate. What is perhaps most surprising is that these factors deal primarily with the process by which the assessment was developed and monitored by political authorities. It is always easy to argue in hindsight that "things might have been different if only . . ." However, in this case, a different process might very well have diffused the worst controversies associated with CLAS and allowed it to continue with its basic features intact.

The first factor relates to the process by which the CLAS test was developed and to the types of participants in that process. The strong desire on the part of SDE officials to use the test as a lever to shape the curriculum led to a development process dominated by curriculum experts, psychometricians, and teachers who subscribe to a constructivist mode of pedagogy. It was an insular process with no public involvement. One opponent called the development group "curriculum Nazis." Although that was an unfair characterization, those involved in the development process acknowledged:

There was a strong bias [on the language arts test development committee]. First of all, we didn't select anybody who didn't have a good understanding of what current thinking is in the field, who wasn't an exemplary teacher. Remember the point of the test was to drive instruction in the direction of the framework. The framework represents a particular take on what it means to read and write, for work in the Language Arts. We only wanted on the development team teachers who understood that framework and as practitioners exemplified what the frameworks stood for. Insofar as it represents a particular bias, the test is mandated to support a

particular bias. Bias as to be the best current state of knowledge in the field. In that sense it is bias. (Language arts assessment consultant)

Talking several months after the CLAS reauthorization had been vetoed, one SDE official suggested that perhaps the test developers had overestimated the ability of the test to drive the curriculum. He characterized the curriculum experts involved in the test development as “the true believers who wrote the curriculum frameworks” and noted the “significant gap in expectations between the curriculum and the measurement people”:

As to the development process . . . After all, we’d all seen the light on the frameworks. At the time, we just didn’t realize the gap between what we were doing curriculum-wise and what we could do measurement-wise. [A psychometrician] was one of the members of the technical advisory committee least sympathetic to the curriculum people. He’d say, “Why should we measure it, when you’re not even teaching it?” [The curriculum experts on the committee] would say, “As soon as you test it, it will get taught.” I now realize that it was unfair to put the burden on the testing system to change teaching.³⁶

However, this debate over CLAS’ potential occurred at a purely technical level among experts. Until the controversy erupted in 1994, there had never been any consideration of broadening the group responsible for developing and reviewing CLAS:

Almost nowhere in discussions about the frameworks was there any consideration of where public involvement should enter into it. There had been such a battle among the experts and the academic community about the social studies frameworks, that public involvement would probably have caused even greater disagreements . . .

I watched other states with citizens’ commissions writing educational goals and the like. But the members of those commissions are typically leaders from business and other sectors; they’re not really representative of the general public. If we had involved the public, we would probably involved the wrong people—they probably would have been the “gung-ho” reform activists. (SDE official)

³⁶ The continuing debate between the testing and the curriculum experts about how CLAS should be configured and what it should be expected to accomplish manifested itself in a number of other ways. For example, the curriculum experts advocated moving to portfolios immediately and eliminating all multiple-choice items. Because they were concerned about validity issues, the testing experts argued against such an approach until the assessment technology was sufficiently improved.

SDE officials also reported they were certain that some curriculum staff in the Department had alerted the ACLU and People for the American Way to the fact that SDE officials and several State Board members were considering dropping “Roselily” from the assessment. These internal disputes further constrained the SDE’s implementation of CLAS.

A second factor, time and budgetary constraints, has already been discussed. One manifestation was the SDE's refusal to release test items that had become the subject of dispute. The Department started from the long-held premise that the release of test items might advantage some test-takers at the expense of others, and that unless some items could be used for more than one year, costs would increase and over-time trends would be more difficult to discern. However, the SDE persisted in requiring test secrecy even in the face of outrageous rumors about what was on the test. Only belatedly, did the SDE allow legislators to view the test and the general public was allowed to see [but not copy or take away] the literature portion of the test at 41 sites statewide for one month after the test had been administered to students.

Although there were clear bureaucratic reasons for the secrecy surrounding the test, the effect was to undermine the SDE's credibility. Opposition leaders could play upon the refusal to release test items, saying, "It makes you wonder what people are doing to your children if you have no right to see the test." After the controversy was over, an SDE official acknowledged the negative effect of the Department's secrecy: "Where we really lost it is that the more we said, 'you can't see the test,' the more we fed mainstream fears."³⁷

The limited information from the SDE about CLAS also meant that technical problems with the test were exaggerated, and that the SDE could do little to counter erroneous accounts. One prominent example related to the scoring of the tests. Because of the high cost of scoring, the SDE decided to score only a sample of the 1993 tests (about 42%). However, according to the expert review of CLAS, that decision was never adequately communicated to parents or even to local superintendents (Cronbach, Bradburn, & Horvitz, 1994). The SDE's failure to communicate how it was dealing with fiscal constraints permitted opponents to

³⁷ Looking back over the controversy and the SDE's response, another SDE official admitted that tactical mistakes were made, but blamed them on the constraints placed on the testing staff by the Department's lawyers, once the lawsuits were filed:

Secrecy for security's sake was a good idea until CLAS blew up. It much too quickly became a legal issue. I put the blame on the legal people who said that we didn't have enough passages to release them and still use them on the test. We didn't give as much attention as we should have to the middle ground. For example, we could have put copies of the test in county offices of education and had people who read them swear not to disclose their contents. But the Department lawyers said absolutely not. By that time, we were in court. The lawyers refused to let us tell a press person whether or not a particular passage was even on the test. As a result, we had to set back while the press reported that all kinds of bizarre passages were on the test like the one about the cat being dragged across nails. It was so ludicrous not to let us respond to the press because we were being crucified.

overstate those constraints by making incorrect assertions, such as “less than one percent were even scored of the students who took the test.”

A third factor explaining the California outcome was the lack of scrutiny by elected officials. Just as there was no public involvement in the development of CLAS, so there was little political oversight until after the controversy erupted. Two mechanisms existed for political oversight of CLAS: the traditional one of legislative hearings and another specific to CLAS, the Assessment Policy Committee that served as an advisory committee to the SDE and included representatives of all the relevant elected officials. However, legislative hearings on CLAS were few, and the APC never saw a copy of the CLAS test. One legislative staffer who was a member of the APC explained the situation in this way:

There wasn't a lot of legislative oversight once SB 662 was passed. There was really only one checkpoint: the SDE was required to produce a five-year plan and they were late in doing that. The lack of oversight on SB 662 was no different than for most legislation. There's really no mechanism for it to happen. The Legislature operates on a contingency basis; if there's some kind of crisis, then there's oversight. The checkpoints are really in the budgetary process—how much money is included for an activity in the governor's budget and then the legislative hearing on the budget. But there was never a hearing that focused on issues such as what the result of the CLAS pilots showed or what the SDE was communicating to parents.

The oversight for SB 662 was supposed to be through the policy committee, but that was like “singing to the choir.” The SDE basically did a “show and tell” for us, and we thought everything was fine because we believed in what the SDE was doing. We assumed (incorrectly perhaps) that the SDE was checking with different constituencies. We only started to hear later that there were concerns about the frameworks; Honig was such a cheerleader for the frameworks.

Another member representing a more conservative elected official confirmed this limited role of the APC:

We were repeatedly reassured by [SDE officials] that the whole issue was being blown out of proportion by the conservative right wing. I don't know why the policy committee didn't become concerned, but the SDE was saying everything was all right . . .

I guess it's pretty surprising, but the policy committee never got involved in test content. We relied on the expertise of the SDE. The test was “by teachers, for teachers.” The committee was going by the agenda of the SDE; we were an advisory group.

This limited political oversight was coupled with a fourth factor, a lack of political leadership. Bill Honig, the charismatic state superintendent of public instruction, had been the state official most committed to using CLAS to change the content taught in the state's schools and the way that content was taught. He was an articulate and passionate advocate for the state frameworks, and he understood that although the state could not mandate such a curriculum in local schools, the CLAS test would move classroom teaching closer to that goal. However, by the time of the CLAS controversy, Honig had resigned his position because he had been indicted on conflict of interest charges involving consulting fees paid to his wife. His replacement as acting state superintendent, William Dawson, was a career civil servant who by all accounts did not have Honig's political skills.

Several policymakers involved with CLAS argued that if Honig had still been in office, the CLAS debacle could have been avoided. That sentiment was expressed by both legislative and gubernatorial staff:

Dave Dawson did a remarkable job under the chaos and the pressure. The poor soul was like the character in the Peanuts cartoon carrying around the cloth. He was in there as a pinch-hitter for Honig who was a wheeler and a dealer. The SDE just wasn't geared up to respond. A lot of people said that the whole thing wouldn't have happened if Honig had still been in office. I would vote with that sentiment. The battle would have been different because Honig would have been in the forefront of the battle. (Legislative staffer)

It was Dawson's naiveté that fanned it more than anything else. If the first ripple had been handled right, it would have gone away. . . .

I think they [the CLAS tests] stink; they're instructionally horrible. They got that way because the curriculum Nazis—the content folks in the SDE and the people they used on committees—had an agenda. These people are a very insular, in-bred group.

That wasn't a problem when Bill [Honig] was there. I'm convinced that *Black Boy* and the open mind question would not have been on the test if he had been there. I know I sound like I'm saying "if only Bill, if only Bill . . ." But it's true, it would have been different if he had been there. (Member of the governor's staff)

An SDE official disagreed with the notion that Honig's presence would have changed the CLAS outcome:

I'm almost certain that it wouldn't have made a difference if Honig had been here and had read the test. He wouldn't have seen the problems, and when the

opposition emerged, he would have seen them as the enemies of education reform and there would have been a battle. We were psychologically incapable of putting ourselves in the critics' shoes. Most people aren't empathetic enough to do that.

Clearly, participants had different perceptions about the role that Honig might have played if he had remained in office, with some seeing him as a moderating influence and others assuming that he would have escalated the controversy by fighting in a spirited, visible way. It is also problematic to assume that one person, no matter how committed and skilled, could have reversed a political movement with such momentum. Nevertheless, the absence of Bill Honig meant that once the CLAS controversy began, there was no visible political leader strongly defending the testing concept. Among the three officials who had initially supported CLAS, only Honig had stressed its underlying curricular values as the dominant rationale for the test. Gary Hart tried to save CLAS because he saw its worth as a curricular and school accountability tool, but he lacked Honig's visibility and had other items on his policy agenda, notably charter schools and teacher training. Pete Wilson had never been a strong supporter of CLAS' specific content and format; his only interest was in individual-level student scores that could be obtained in a number of different ways, including traditional multiple-choice tests. Consequently, it is reasonable to assume that had Honig still been in office, CLAS would have received a more vigorous and skilled defense when it ran into trouble.

The final factor explaining the CLAS outcome was timing: The controversy occurred in an election year. Democratic respondents argued that Wilson's veto of the CLAS reauthorization could be explained by the fact that he was running for re-election. Further complicating the electoral picture was the campaign for state superintendent of public instruction. Delaine Eastin, the Democratic chair of the Assembly Education Committee, was running against Maureen DiMarco, who is also a Democrat and Wilson's appointed secretary of child development and education. DiMarco was in a particularly difficult position because she had opposed Proposition 174, the state voucher initiative that was supported by many of the CLAS opposition groups, and she was viewed as part of the education establishment by those groups and their supporters. At the same time, her affiliation with Wilson made traditional Democrats less likely to support her, so opposing CLAS would distinguish her from Eastin and increase her chances of

gaining support from those planning to vote for Wilson. The governor's staff argued against the notion that the election year had anything to do with his veto:

The religious right was long gone by the time of the governor's veto of SB 1273. That group was clearly going to vote for him; they had no place to go, they weren't going to vote for Kathleen Brown [his opponent]. It was like taking "coals to Newcastle." One way you can tell that the decision about the veto had little to do with the campaign is that during the election, the governor's staff looked at everything that they thought would impact on the campaign. They didn't even ask to see [his veto message].

It is difficult to know exactly how great an impact the election campaign had on the CLAS outcome. Still, most respondents assumed that there was some relationship. Even an opposition group leader expressed the opinion that given Wilson's past support for CLAS, "if it was not an election year, I don't know if he would have vetoed that re-authorization."

Suggesting that the CLAS controversy would not have occurred if there had been a more open, inclusive development process is only speculation. One might argue that such a controversy was inevitable in a state as culturally and politically diverse as California. Nevertheless, there is some evidence that a different process might have led to a different outcome. For example, an SDE official reported that after the CLAS controversy began, the Department convened a panel of parents and members of the public expressly chosen to represent disparate viewpoints. The purpose was to see if the panel could agree on an appropriate set of reading passages and test items to include on CLAS. Although the group reached consensus on only a small sub-set of passages and items, the exercise did demonstrate that people with diverse cultural, religious, and political beliefs could agree on curricular content that they all thought was appropriate and worthwhile. Unfortunately, by the time the panel was convened, CLAS was already facing extinction.

A second piece of evidence comes from Kentucky's experience. Like California's, Kentucky's development process was dominated by testing experts, curriculum specialists, and classroom teachers, and also like California, it encountered serious opposition. However, the state response in Kentucky differed significantly from California's, with the result being that KIRIS continues to be implemented, albeit in a somewhat modified form. The Kentucky case suggests that while controversy might be difficult to avoid, political leadership and oversight can be effective in diffusing such conflicts.

The Kentucky Response

During the 1994 legislative session, the political and business elites who support KERA mobilized to blunt the opposition and ensure the reform act's continued implementation. One strategy involved the legislature and the state commissioner of education making some "midcourse corrections" to protect KERA from further assault and to ensure that it would remain basically intact. As noted in a previous section, they eliminated two of the state's six learning goals—dealing with individual self-sufficiency and responsible group membership—that were targets of the most intense criticism; they delayed the imposition of sanctions on underperforming schools until 1996; and they shifted the high school assessment from the twelfth to the eleventh grade. The state's 75 "valued outcomes" were reduced to 57 "academic expectations" with an emphasis on goals that are academic and can be measured by the assessment. All references to "emotions," "feelings," "evolution," and "environment" were eliminated from the standards. In addition, the commissioner of education announced that any adult who was willing to sign a nondisclosure agreement would be allowed to see all the KIRIS test items.

KERA, and particularly KIRIS, became an issue in the 1995 gubernatorial campaign. The Republican candidate, Larry Forgy, argued that the state should be using a national, standardized test that would allow Kentucky's scores to be compared with those in other states; and that less emphasis should be placed on the writing portfolios because they consume too much time and the grading is too subjective. Forgy's opposition to KERA became a highlight of his campaign commercials, with him saying that "KERA is failing us in the classroom" and that it "is putting our children at risk." In focusing on this issue, he was capitalizing on voter sentiment. A state poll conducted during the campaign showed that about 42% of likely voters who had read or heard about KERA approved of the changes it was bringing about, while 45% disapproved (Cross, 1995a). The Democratic candidate, Paul Patton, largely defended KERA during the campaign. However, in response to Forgy's criticisms, Patton agreed that the state assessment had problems and suggested that the number of performance levels be increased and that more fact-based questions that everyone could agree upon be included on KIRIS. In his campaign commercials, Patton promised to undertake "a comprehensive review of KERA—keep the programs that work, get rid of those that don't" (Cross, 1995b). Patton won the governorship with 51% of the vote.

Most political commentators agreed that KERA had been one of the major issues in the campaign and that it was likely to be modified further during the 1996 legislative session, but that the essence of KERA and KIRIS would continue intact. In fact, during the 1996 legislative session, the General Assembly made few changes in KERA, while the governor fulfilled his campaign promise to appoint a task force to study KERA. The eighteen-member task force, with members appointed by the governor, the state Senate, and the House, will take 18 months to study KERA and make recommendations for the 1998 legislative session. So, KERA and the accompanying state assessment have been given something rare in education reform—at least eight years to be implemented and to show results with only modest revisions and a firm political commitment to “stay the course.”

Two strategies characterize the state response in Kentucky and distinguish it from California's. The first is that key state officials quickly responded to opponents' criticisms. Thomas Boysen, the commissioner of education, moved promptly to meet with KIRIS opponents. For example, he met with 180 members of the Greater Lexington Ministerial Fellowship whose members had been among those most critical of KERA. Boysen admitted that “it was extremely unfortunate, in bad taste” for *Transformations* to have included as a recommended activity that students join two liberal organizations. In speaking to the group, Boysen also promised to work with the ministers to change KIRIS items that they found objectionable and to consider ways that churches might participate in the character education of students (White, 1994). The SDE began to include anti-KERA leaders on SDE committees and advisory groups that were revising the state standards, previewing KIRIS, and developing the state's Goals 2000 plan.³⁸

The KERA opponents argued in public that the midcourse adjustments did not go far enough. Donna Shedd of the Eagle Forum called the adjustments “backfires”:

They're burning off some vegetation to stop the fire coming along the prairie. But the fire is too big. The back fires won't work. (as quoted in May, 1994c)

³⁸ There did, however, seem to be elements of a “good cop, bad cop” approach in the state's response to KERA opposition. While Boysen was meeting with the opponents and trying to address their concerns, David Karem, the majority leader in the state Senate and one of KERA's most vocal supporters, was calling Martin Cothran, one of the opposition leaders, “that little, cute, right-wing boy” (May, 1995a).

Nevertheless, opponents like Shedd participated in state advisory groups and regularly met with Boysen and his staff, even as they continued to seek major changes in KERA.

A second factor that distinguished the Kentucky response from California's is that groups and prominent individuals mobilized in support of KERA. The Prichard Committee, a statewide group that had promoted education reform for over a decade, launched a campaign to counteract criticism of KERA. As one activity, it sent letters to the editors of Kentucky newspapers indicating inaccuracies in the op-ed articles and letters written by KERA opponents. The Partnership for Kentucky School Reform, an organization of business leaders affiliated with the Prichard Committee and the national Business Roundtable, sent letters to all state legislators affirming their support for KERA.³⁹ It also placed advertisements in the state's newspapers countering charges made by KERA opponents. When KERA became an issue in the 1995 gubernatorial campaign, the Partnership spent \$100,000 on a television advertising campaign defending the state reforms (Associated Press, 1995).

So, unlike the case in California, political and business elites were willing to speak out in favor of the state assessment and the principles underlying it. One example is a speech by John Hall, the CEO of Ashland, Inc., that was excerpted as an op-ed article in the *Lexington Herald-Leader*:

What is vital about assessment is the principle, not the details. The principle is that we must test our children in a comprehensive way so we can measure their progress.

...

The KIRIS tests probably do need to be revised and perhaps should be accompanied by some type of national measure that enables us to evaluate how our students are performing against those in other states. I seriously question, however, whether we can accurately measure sustained progress in student performance with just a simple, multiple-choice, computer-graded test. . . .

So where do we go from here?

Speaking on behalf of my fellow employees at Ashland and my colleagues in the Partnership (for School Reform), we stay the course. We have come a long way, and we still have a long way to go. Some course corrections have already been made, and others will be necessary. (11/19/95)

³⁹ The Partnership for Kentucky Education Reform was founded by the chairmen and chief executive officers of United Parcel Service, Ashland Oil, and Humana, all Fortune 500 companies headquartered in Kentucky. The group was established in March 1991 with the intention of supporting KERA for ten years.

One critical difference between Kentucky and California largely explains the two states' differing responses to the similar opposition each faced. The scope of KERA, including its visibility and the new financial resources required, meant that the political leaders who endorsed the reform program had to be active and visible in their support to get it enacted and its funding levels maintained. Several respondents noted that the core of legislators who were in leadership positions when KERA was enacted had "basically staked their careers on education reform" (interview with legislative staffer). As a result, they had no alternative but to ensure that "objections to it [were] met very energetically" (interview with Kentucky Department of Education official).

One of the elected officials most involved in enacting KERA further attributed active support by political leaders to the process by which it was developed:

There were three viewpoints at the time of the Supreme Court decision about how the legislature would react. One was that the legislature would be involved in responding to the decision and devising a remedy. The second was that the legislature would stonewall, and the third was that we would form a blue ribbon commission composed of business leaders and the like. The first option prevailed. As a result, it gave members of the legislature pride of ownership. If the leadership is stable, we should be able to support and maintain KERA.

Consequently, even though KERA was developed by a small group, it was more directly a legislative product than in California where the APC was responsible for the basic structure of CLAS, with elected officials' involvement limited to ratifying what their representatives had designed. Not only was there a greater policy incentive for elected officials to take a strong stand in Kentucky, but the deliberations that had occurred during the enactment process meant that at least a few political leaders understood, accepted, and could defend the ideas behind KERA and KIRIS.⁴⁰

⁴⁰ Although rank-and-file legislators understood KIRIS less well than the leadership and were less vocal in their support, most endorsed the concept. A telephone survey of 102 of the 135 members of the General Assembly, conducted during the summer of 1994, found that only 11% found the assessment successful with another 59% judging it somewhat successful. Ninety percent of the legislators surveyed had fielded some negative complaints about the assessment program, with the accuracy of the assessment program and its overall appropriateness among the major concerns voiced. Just over half had received some positive reports from constituents. Nevertheless, 69% agreed that the state assessment should be primarily performance-based (Office of Education Accountability, 1994).

The Technical Becomes Political

Testing experts have warned that while performance assessments could potentially be a vast improvement over multiple-choice tests, they are not without their problems. In an analysis of what is known about performance assessments, Linn (1993) focused particularly on the generalizability problems associated with new forms of assessments. He noted that one of the major stumbling blocks to the implementation of performance-based assessment systems is “the limited degree of generalizability of performance from one task to another” (p. 9). Citing a variety of evidence from performance assessments in history, mathematics, science, and even licensure examinations in law and medicine, Linn concluded that because performance on one task has only a weak to modest relationship to performance on another, a large number of tasks (or increased testing time for more complex tasks) will be necessary to ensure that assessments produce comparable information and that results are fair to the individuals being tested.

Others have warned of the same potential problem, noting that “some students who fail on the basis of one overly limited or non-representative sample of tasks [might] have passed if given an equally defensible alternative set” (Koretz, Madaus, Haertel, & Beaton, 1992). If a test has consequences, but valid generalizations cannot be drawn from the exercises included on it, then test takers will be treated capriciously. However, the remedy of increasing the number of tasks creates its own feasibility problems in terms of cost and time burden.

Limitations on the ability to make valid generalizations and the reliability of a scoring process requiring considerable judgment on the part of multiple scorers are the major technical problems associated with performance assessments. These problems are exacerbated when parents and politicians expect individual-level student scores, which require a different and more expensive design than the matrix design typically used in assessing schools. When assessments are used for high-stakes purposes, reliability and validity problems also become more serious.

The assessment systems in Kentucky and California were evaluated by panels of testing and measurement experts, and both systems were found to have serious flaws. The report on Kentucky found that KIRIS was not reliable enough to use as the basis for rewarding or sanctioning schools. The panel concluded that the performance standards for deciding students’ level of proficiency are based on

too few items; the scoring of portfolios is too subjective and inconsistent for a high-stakes system; errors in equating across assessments make year-to-year comparisons of KIRIS results of questionable validity; and at least for fourth-grade reading, student gains on KIRIS are not matched by the scores of Kentucky students on the National Assessment of Educational Progress (Hambleton et al. 1995).⁴¹ The panel agreed with the Kentucky Department of Education's own evaluation (1994) that concluded:

The content being assessed is not well enough defined and the relationship among the learning goals, academic expectations, the curriculum frameworks, the program of studies, course outlines, released items and performance standards is not clearly described. This makes it difficult for teachers to focus their instructional efforts and to explain to parents what is being taught and assessed. (p. 10)

The panel recommended that portfolios not be used at this time for accountability purposes; that validation work be expanded; that the performance standards be re-established and full documentation of the process provided; that public reports be clear about limits on generalizability of findings to a new set of tasks; that multiple-choice items be added to increase content validity and scoring reliability; and that the state reconsider its shift toward instructional process at the expense of curriculum content.

SDE officials accepted the majority of the panel's recommendations, and moved to make changes in the assessment. Among the changes are the inclusion of multiple-choice items beginning in 1997; greater attention to the technical quality of the test through more extensive equating and validation work; greater emphasis on the testing of content knowledge; and the administration of national norm-referenced tests in reading and mathematics in Grades 6 and 9. The major disagreement with the panel was the SDE's decision, with legislative support, to continue to include writing portfolios in schools' accountability scores. However, the state has implemented routine audits to ensure reliable scoring of the portfolios by teachers. The state also decided to continue KIRIS as a high-stakes assessment, with school-level rewards and sanctions attached to test scores.

⁴¹ Although the state did report significant gains in student scores over a two-year period (e.g., the percentage of fourth graders scoring at the proficient level increased from 7% to 12%), 85% of the students who were tested were still performing at either the novice or apprentice levels. In other words, the overwhelming majority of students could not be considered proficient on the state's performance standards.

The select committee evaluating the California assessment identified fewer fundamental problems with the test and in fact, argued in its report that all the shortcomings of CLAS could be remedied and that as CLAS matured, it should be able to deliver “a highly useful product” (Cronbach et al., 1994, transmittal letter). Some of the technical problems with CLAS were logistical ones—for example, lost test booklets, and errors in matching student identification sheets to test booklets. The most serious problem was that because of resource constraints, the SDE and its test contractor did not score all the test booklets. As a result, few school-level reports in 1993 had adequate reliability. The level of standard error was also unacceptably high as a “consequence of allowing no more than one or two hours of testing time per area for examining on a range of complex intellectual tasks” (p. 4). The select committee recommended that the state not distribute individual-level scores until the technical problems could be resolved.

In both the California and Kentucky cases, the assessment designers had to balance daunting technical challenges against the political reality that elected officials expected to see performance assessments on-line within a very short time frame. There is no question that SDE officials in both states overpromised on what they could deliver. Given the constraints of having to generate individual student scores, limit testing time, report student performance in terms of a set of absolute curriculum standards, keep scoring costs within reasonable time limits, and in the case of Kentucky, make high-stakes decisions based on assessment results, considerably more time was needed for test development. However, in both states, SDE officials calculated that a flawed assessment was less costly than waiting until it met higher technical standards and risking the loss of political support for standards-based reform. In hindsight, SDE officials probably could have bought themselves some time—perhaps one year—without losing the entire endeavor, but technical problems would still have remained an issue.

In both states, opponents used the technical reviews to bolster their cases, and the reports did become part of the political debate. But in neither state were they a deciding factor in how the assessment controversy was resolved. In California, the report on the technical quality of CLAS was issued after political controversy had overwhelmed the assessment. Consequently, the report’s assumption that technical problems could be remedied over time was never verified, as CLAS was discontinued several months later. Because officials moved quickly in Kentucky to implement most of the panel’s recommendations, and the

legislature reaffirmed its support for KIRIS, the technical report became only temporary fodder in the ongoing debate over the assessment system.

IV. CONCLUSIONS

The most obvious conclusions to draw from an analysis of new state assessments in California, Kentucky, and North Carolina deal with the determinants of successful policy implementation. One might conclude that North Carolina's program was less problematic than that of the other two states because it relied heavily on proven testing technology and the resulting assessment represented only an incremental step toward an alternative format. Some of California's problems might very well have been predicted by the competing goals that key policymakers held for CLAS, with the unresolved priorities likely to constrain the implementation process. The broad scope of Kentucky's reforms meant that even a thoughtful, well-coordinated implementation strategy would tax state and local staff and generate some backlash. All three states labored under the traditional implementation constraints of short time lines, limited resources, and the need to communicate complicated, new routines down through the governmental system to street-level bureaucrats in local schools and classrooms.

Each of these conclusions is reasonable and suggests a variety of implications for other states considering the implementation of new forms of assessment. However, to see these states' experience only as lessons in the "do's and don'ts" of policy implementation is to miss the most important message about the politics of testing. Because curricular standards and new forms of assessment can only accomplish their intended purposes if they embody powerful ideas about what knowledge is most valuable for students to learn, those values will shape the politics of implementation. Kentucky's assessment has the added lever of rewards and sanctions allocated to schools on the basis of their test scores. Yet, just as in the other two states without tangible consequences attached to their testing systems, implementation was less influenced by that characteristic of KIRIS than by its curricular values and the assessment technology used to measure student performance. All of these assessments are linked to state curriculum standards that not only specify what knowledge and skills are most important, but also assume that knowledge will be conveyed to students through a pedagogy

that gives them opportunities to construct their own learning through projects, experiments, work with other students, and real-world applications.

Policymakers in the three states also assumed that the assessments would ensure school-level accountability, either through rewards and sanctions in Kentucky or through the public reporting of scores in California and North Carolina. But what most distinguished these assessments from earlier ones was the strong belief that the assessment should influence classroom instruction. Not only would their content motivate what is taught, but the test format would lead teachers to emphasize writing, have students work in groups, require them to explain the process by which they solved problems, and engage in hands-on activities. In order to influence instructional strategies as well as content, the assessment format had to include more than multiple-choice items, thus introducing the perception of greater scorer judgment, and hence variability, than in the machine-scored, multiple-choice tests.

Thus the assessments were premised on a strong belief that the state standards embodied what was most worthwhile for students to know, that the state assessment could move instruction in that direction, and that the end product would be higher student achievement. The changes assumed in North Carolina were relatively modest—viz., more student writing and stricter adherence to the state *Standard Course of Study* which had been a fixture in the state for close to a century. In contrast, what is remarkable about California and Kentucky is that not only was the state specifying a more precise curriculum through its framework and assessment than it had traditionally, but the curriculum was unfamiliar to most adults in the state.

Such a marked departure from the state's historical role, coupled with the novelty of the curriculum, meant that implementation of the new assessments was unlikely to be simply a technical, administrative task. At one level, the vocal protests against CLAS and KIRIS can be dismissed as the views of a small minority representing the “radical right”—as several policymaker respondents in both states called them. It is true that opposition leaders in both California and Kentucky resisted the state assessments because they viewed them as symptomatic of aspects of secular life that they found to be hostile to the traditional religious and cultural values they espoused. In that sense, the opposition to CLAS and KIRIS could be viewed as just another chapter in the continuing debate over the state interest in schooling versus the right of a small

minority of religious parents to protect their children from what they consider to be alien and harmful influences (Gutmann, 1995; Macedo, 1995). In the assessment cases, the courts found that the test items did not intrude into students' religious beliefs or family life because students were not required to reveal information from those domains. Rather, the items were designed to elicit analytical, comprehension, and writing abilities.

Even if the CLAS and KIRIS controversies are seen as only the attempts of a small minority of religious parents to protect their children from what they perceive to be public harm, these cases are significant because they illustrate the competing goals of home and state that the courts and the larger political system have attempted to resolve over most of the nation's history. But to see these controversies as only a minority protest is to miss some important points. The available opinion data strongly suggest that the larger public is skeptical of new curricular approaches in reading, writing, and mathematics. In addition, the controversy over whole language and phonics in reading instruction that has recently dominated the education agenda in California indicates that there is considerable merit to parents' commonsensical concerns that their children are not learning the "building block" skills needed to read well or that "invented spelling" and a lack of knowledge of grammar rules will hinder their writing ability.

Clearly, parts of the opposition leaders' agenda were not endorsed by most parents, but the opponents were also tapping critical concerns shared by a broader group of secular parents and the general public. The Kentucky opinion polls showing relatively low awareness of KERA even after several years of implementation suggests that concerns about the early use of calculators and inattention to rules of grammar expressed in the national Public Agenda surveys is representative of a diffuse unease on the part of many people, though it may not be directly linked to judgments about specific policies. But it may also be the case that particular issues, as reflected in the press for greater attention to phonics in California, represent the concerns of those parents who are most attentive to their children's education and whose support is particularly important to the public schools. But whether these concerns come from uninvolved parents or from those who are better informed and more active is less important than the fact that apprehension about new curricular approaches extends beyond the religious conservatives who spearheaded the opposition to CLAS and KIRIS.

So what can states do to assuage these concerns and still pursue a reform agenda anchored in common curricular standards and new forms of assessment? The experience of California, Kentucky, and North Carolina suggests three important lessons for other states considering new forms of assessment. The first is to decide whether the state is ready to make fundamental changes in curriculum and testing like those in California and Kentucky or whether a more incremental approach like North Carolina's is preferable. Where there is uncertainty about the effectiveness of new approaches, where key decision makers hold very different expectations about what the reform should accomplish, or where political will and resources are lacking to make the needed, long-term investments in teacher training, then an incremental approach may yield greater success than a comprehensive strategy implemented without political conviction.

Second, if a state does decide to take the California and Kentucky route to testing, strong political leadership is necessary. Fundamentally different approaches to teaching and testing need articulate spokespersons who firmly believe in the ideas and who can persuade parents and the general public that these strategies will produce positive gains for individual students and for the state as a whole. That support has to be visible, and it has to come from people whom the public feel can be held responsible for the outcomes. For that reason, the support needs to come from people who are electorally accountable, and not just from professional educators and unelected officials within the education establishment. Political support also needs to reflect a long-term commitment that recognizes that implementing a reform as complex as new curriculum and testing will take time.

However, strong political leadership does not mean that elected officials are rigid in their defense of an assessment initiative and thus closed to changes proposed by critics, whether they be technical experts, teachers, or parents. The novelty of the testing technology and the fact that these curricular innovations reflect applications of basic research that have only recently begun to be applied on a widespread basis mean that improvements and changes are inevitable. Consequently, policymakers who espouse these reforms face a daunting challenge. On the one hand, they need political courage to support a major departure from the status quo whose time frame extends beyond the constraints of two- and four-year electoral cycles. At the same time, they must enter into a dialogue with the public to persuade them of the value of these ideas, but still be willing to modify

assessment policies in the face of parental and public concerns which surface during that extended dialogue.

Another facet of strong political leadership requires that elected officials exert regular oversight over the experts developing the assessment. This recommendation is not a suggestion that politicians “micro-manage” the test development process. However, legislators who are in regular contact with constituents and who must run for office every few years bring a “real world” perspective that may be lacking in people who are primarily concerned about effective teaching or the technical quality of the test. Two comments from California respondents illustrate this lesson that they unfortunately learned too late:

[An elected official] would look at the test from the perspective of how John Q. Public would regard it. The content committee was looking at it as including wonderful literature. It’s a completely different perspective. But someone should have taken a common sense perspective and asked “How would I feel as a parent reading this?”
(Legislative staffer)

We didn’t use anyone as a political screen. There was no tradition of showing the test to policy committees. The idea was that technicians build tests, and policy committees decide things like what subjects and grades should be tested. It may have been a big or a little mistake not to show the test to them, but it never occurred to us until things blew-up. (State Department of Education official)

The need to balance a “real world” perspective with curricular and psychometric expertise leads to a final lesson: The development of new curriculum standards and assessments cannot be solely a technical process with participation limited to experts. Again, this recommendation is not a suggestion that parents become test item writers. But decisions as significant as what knowledge is most important for students to learn and how they should be tested on their mastery of it require open, public deliberation. That participation can be organized in any number of ways, including: state-level review committees; forums in local communities sponsored by the PTA or the League of Women Voters; informal “coffees” in people’s homes; and op-ed exchanges in local newspapers, radio, and television programs. Clearly, in large states as culturally diverse as California, reaching acceptable levels of consensus will be difficult. But not to try is to make a mockery of the notion of common curriculum standards for public schools. Parents and other members of the public who voice concerns about new

assessments need to be treated with respect, and not dismissed as either uninformed or politically extreme.

Above all, those responsible for implementing new state assessments need to recognize that the process is inherently political in the best sense of that word because it involves public deliberation about what skills and knowledge are most important for a productive life and active citizenship.

REFERENCES

- Associated Press. (1994, May 2). Opposition to KERA has led to forming of splintered front. *Lexington Herald-Leader*, p. B3.
- Associated Press. (1995, October 28). Business leaders' group supporting KERA through \$100,000 in TV ads. *Lexington Herald-Leader*, p. C3.
- Associated Press. (1996, February 25). Family loses bid to refuse KERA testing. *Lexington Herald-Leader*, p. B3.
- Bishop, B. (1994, March 23). "Independent" legislature still wheels and deals in secret. *Lexington Herald-Leader*, p. A9.
- Bardach, E. (1977). *The implementation game: What happens after a bill becomes a law*. Cambridge, MA: The MIT Press.
- California Reading Task Force. (1995). *Every child a reader*. Sacramento, CA: California Department of Education.
- Capitol Resource Institute. (1994). *CLAS information packet*. Sacramento, CA: Author.
- Cohen, D. K., & Peterson, P. L. (1990). Special issue of *Educational Evaluation and Policy Analysis*, 12(3).
- Colvin, R. L. (1995a, March 23). State's reading, math reforms under review as scores fall. *Los Angeles Times*, pp. A1, A21.
- Colvin, R. L. (1995b, April 6). Eastin names panels to probe education failure. *Los Angeles Times*, pp. A3, A17.
- Colvin, R. L. (1995c, July 5). School goals spelled out in "ABC Bill." *Los Angeles Times*, pp. A3, A12.
- Colvin, R. L. (1995d, October 2). Pressure mounts on Wilson to OK new test for academic assessment. *Los Angeles Times*, pp. A3, A15.
- Colvin, R. L. (1995e, November 19). Her best subject. *Los Angeles Times*, pp. E1, E2.
- Corbett, H. D., & Wilson, B. L. (1991). *Testing, reform, and rebellion*. Norwood, NJ: Ablex Publishing.
- Cronbach, L. J., Bradburn, N. M., & Horvitz, D. G. (1994). *Sampling and statistical procedures used in the California Learning Assessment System* (Report of the Select Committee). Stanford, CA: Stanford University.

- Cross, A. (1995a, October 19). KERA opposition rife, Forgy says. *The Courier-Journal* (Louisville, KY), p. B5.
- Cross A. (1995b, October 20). New Patton TV ad says he'd "fix KERA." *The Courier-Journal* (Louisville, KY), p. B6.
- Daniel, S. M. (1995, November 19). KERA teaches a radical agenda. *Lexington Herald-Leader*, p. E2.
- Delsohn, G. (1994, September 28). Christian law firm has clout in state. *Sacramento Bee*, pp. A1, A7.
- DiegmueLLer, K. (1995, September 27). More basic-skills instruction in California urged. *Education Week*, pp. 14, 17.
- Elam, S. M., & Rose, L. C. (1995). The 27th Phi Delta Kappa/Gallup Poll of the public's attitudes toward the public schools. *Phi Delta Kappan*, 77(1), 41-56.
- Elam, S. M., Rose, L. C., & Gallup, A. M. (1991). The 23rd Annual Gallup Poll of the public's attitudes toward the public schools. *Phi Delta Kappan*, 73(1), 41-56.
- Elam, S. M., Rose, L. C., & Gallup, A. M. (1992). The 24th Annual Gallup/Phi Delta Kappa Poll of the public's attitudes toward the public schools. *Phi Delta Kappan*, 74(1), 41-53.
- Fullan, M. G. (1991). *The new meaning of educational change*. New York: Teachers College Press.
- Gutmann, A. (1995). Civic education and social diversity. *Ethics*, 105, 557-579.
- Hageman, B. (1993, October 11). New state test spurs criticism. *The Press-Enterprise* (Riverside, CA), p. B1.
- Hall, J. R. (1995, November 19). The debate never ends: Business needs schools that go beyond the basics. *Lexington-Herald Reader*, p. E2.
- Hambleton, R. K., Jaeger, R. M., Koretz, D., Linn, R. L., Millman, J., & Phillips, S. E. (1995). *Review of the measurement quality of the Kentucky Instructional Results Information System, 1991-1994*. Frankfort, KY: Office of Educational Accountability, Kentucky General Assembly.
- Harp, L. (1994, May 18). The plot thickens: The real drama behind the Kentucky Education Reform Act may have just begun. *Education Week*, pp. 20-25.
- Herman, J. L., & Golan, S. (no date). *Effects of standardized testing on teachers and learning—another look* (CSE Tech. Rep. No. 334). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing.

- Ingram, C. (1994, March 12). Stories deleted from tests to be reinstated. *Los Angeles Times*, p. A1.
- Ingram, H., & Schneider, A. (1993). Constructing citizenship: The subtle messages of policy design. In H. Ingram & S. R. Smith (Eds.), *Public policy for democracy*. Washington, DC: The Brookings Institution.
- Johnson, J., & Immerwahr, J. (1994). *First things first: What Americans expect from the public schools*. New York: Public Agenda.
- Kentucky Institute for Education Research. (1994). *Statewide education reform survey*. Louisville, KY: Tom Wilkerson & Associates, Ltd.
- Kentucky Department of Education. (1993). *Transformations: Kentucky's curriculum framework (Volume I)*. Frankfort, KY: Author.
- Kentucky Department of Education. (1994). *Position paper on recommended changes in the KIRIS Assessment and Accountability Program*. Frankfort, KY: Author.
- Kingdon, J. W. (1993). Politicians, self-interest, and ideas. In G. E. Marcus & R. L. Hanson (Eds.), *Reconsidering the democratic public* (pp. 73-89). University Park: Pennsylvania State University Press.
- Kingdon, J. W. (1995). *Agendas, alternatives, and public policies*. New York: HarperCollins College Publishers.
- Koretz, D. M., Madaus, G. F., Haertel, E., & Beaton, A. E. (1992, February 19). *National educational standards and testing: A response to the recommendations of the National Council on Education Standards and Testing*. Testimony before the Subcommittee on Elementary, Secondary, and Vocational Education, Committee on Education and labor, U.S. House of Representatives.
- Koretz, D. M., Stecher, B., Klein, S., & McCaffrey, D. (1994). *The evolution of a portfolio program: The impact and quality of the Vermont portfolio program in its second year (1992-93)* (CSE Tech. Rep. No. 385). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing.
- Linn, R. L. (1993). Educational assessment: Expanded expectations and challenges. *Educational Evaluation and Policy Analysis*, 15(1), 1-16.
- Macedo, S. (1995). Liberal civic education and religious fundamentalism: The case of God v. John Rawls? *Ethics*, 105, 468-496.
- Madaus, G. F. (1988). The influence of testing on the curriculum. In L. N. Tanner (Ed.), *Critical issues in curriculum, 87th yearbook of the National Society for the Study of Education* (pp. 82-121). Chicago: The University of Chicago Press.

- Marcano, T. (1990, November 11). Year off might do CAP test some good. *Los Angeles Times*, p. T1.
- May, L. (1994a, January 16). Some see KERA changes as effort to quiet foes. *Lexington Herald-Leader*, p. A1.
- May, L. (1994b, March 6). Kentucky schools' test scores are still full of questions: 2 studies to review validity of KERA-required program. *Lexington Herald-Leader*, p. B1.
- May, L. (1994c, March 23). Education petition delivered calling for KERA repeal. *Lexington Herald-Leader*, p. A6.
- May, L. (1995a, February 7). KERA critic embraces "world of controversy." *Lexington Herald-Leader*, p. A1.
- May, L. (1995b, September 29). Education groups unite with concerns about KERA testing. *Lexington Herald-Leader*, p. A1.
- May, L. (1995c, November 9). How much, how quickly KERA will change uncertain. *Lexington Herald-Leader*, p. A16.
- May, L. (1995d, November 10). Wilkinson cites school voucher issue. *Lexington Herald-Leader*, p. B7.
- McDonnell, L. M. (1994a). Assessment policy as persuasion and regulation. *American Journal of Education*, 102, 394-420.
- McDonnell, L. M. (1994b). *Policymakers' views of student assessment*. Santa Monica, CA: RAND.
- McLaughlin, M. W. (1990). The RAND change agent study revisited: Macro perspectives and micro realities. *Educational Researcher*, 19(9), 11-16.
- Merl, J. (1994, May 11). Judge finds no privacy invasion in CLAS tests. *Los Angeles Times*, pp. B1, B8.
- National Council on Education Standards and Testing. (1992). *Raising standards for American education*. Washington, DC: Author.
- National Council of Teachers of Mathematics. (1989). *Curriculum and evaluation standards for school mathematics*. Reston, VA: Author.
- Noble, A. J., & Smith, M. L. (1994). Old and new beliefs about measurement-driven reform: "Build it and they will come." *Educational Policy*, 8, 111-136.
- Office of Education Accountability. (1994). *A survey of legislators on Kentucky Instructional Results Information System*. Frankfort, KY: Author.

- Office of Educational Research and Improvement. (1988). *Creating responsible and responsive accountability systems: Report of the OERI State Accountability Study Group*. Washington, DC: U.S. Department of Education.
- Palumbo, D. J., & Calista, D. J. (1990). Opening up the black box: Implementation and the policy process. In D. J. Palumbo & D. J. Calista (Eds.), *Implementation and the policy process: Opening up the black box* (pp. 3-17). Westport, CT: Greenwood Press.
- Picus, L. O. (1996). *Estimating the costs of student assessment in North Carolina and Kentucky: A state-level analysis* (CSE Tech. Rep. No. 408). Los Angeles: University of California, Center for Research on Evaluation, Standards, and Student Testing.
- Pitsch, M. (1994, October 19). Critics target goals 2000 in schools "war." *Education Week*, pp. 10, 21.
- Ravitch, D. (1995). *National standards in American education: A citizen's guide*. Washington, DC: The Brookings Institution.
- Resnick, L. B. (1994). Performance puzzles. *American Journal of Education*, 102, 511-526.
- Rose v. Council for Better Education, Inc. (1989). 790 S.W. 2d 186 (Ky. 1989).
- Rothman, R. (1992, September 23). Kentucky reformers await reaction to results of tough tests. *Education Week*, pp. 1, 20.
- Rothman, R. (1995). *Measuring up: Standards, assessment, and school reform*. San Francisco, CA: Jossey-Bass.
- Schneider, A., & Ingram, H. (1990). Behavioral assumptions of policy tools. *Journal of Politics*, 52, 510-529.
- Schreiner, B. (1996, January 29). Making waves family foundation trying to keep issues at forefront. *Lexington Herald-Leader*, p. B3.
- Smith, M., & O'Day, J. (1991). Systemic school reform. In S. Fuhrman & B. Malen (Eds.), *The politics of curriculum testing*. Philadelphia: Falmer Press.
- Smith, M. S., & Scoll, B. W. (1995). The Clinton human capital agenda. *Teachers College Record*, 96, 389-404.
- Sommerfeld, M. (1996, April 24). California parents target math frameworks: Critics worry schools have abandoned basics. *Education Week*, pp. 1, 11.
- Stammer, L. B. (1995, November 2). A man of millions: Broadcaster James Dobson has become a leading name in evangelical circles. *Los Angeles Times*, pp. E1, E5.

- Stone, D. H. (1988). *Policy paradox and political reason*. New York: Harper Collins.
- Walsh, M. (1994, September 28). Christian Coalition puts education at heart of election agenda. *Education Week*, p. 10.
- Ward, B. (1993, November 20). Minister says evidence shows witchcraft taught in schools. *Lexington Herald-Leader*, p. B3.
- White, S. (1994, March 3). Boysen offers concessions to ministers on KERA. *Lexington Herald-Leader*, p. A1.
- Wilgoren, J. (1994, June 16). A look at what's really there in the CLAS tests. *Los Angeles Times*, pp. A1, A24, A25.
- Wilson, P. (1994, September 27). Veto message for SB 1273 (reauthorization of state testing program).
- Zoroya, G. (1995, October 1). Flying right. *Los Angeles Times*, pp. E1, E2.