

**The Politics of Assessment: A Case Study
of Policy and Political Spectacle**

CSE Technical Report 468

Mary Lee Smith, Arizona State University

Walter Heinecke, University of Virginia

Audrey J. Noble, University of Delaware

May 1997

Center for the Study of Evaluation
National Center for Research on Evaluation,
Standards, and Student Testing
Graduate School of Education & Information Studies
University of California, Los Angeles
Los Angeles, CA 90095-1522
(310) 206-1532

Copyright © 1998 The Regents of the University of California.

The work reported herein was supported under the Educational Research and Development Centers Program, PR/Award Number R305B60002, 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 positions or policies of the National Institute on Student Achievement, Curriculum, and Assessment, the Office of Educational Research and Improvement, or the U.S. Department of Education.

**THE POLITICS OF ASSESSMENT:
A CASE STUDY OF POLICY AND POLITICAL SPECTACLE¹**

Mary Lee Smith, Arizona State University

Walter Heinecke, University of Virginia

Audrey J. Noble, University of Delaware

ABSTRACT

The Arizona Student Assessment Program (ASAP) was the official state assessment policy from its inception in 1990 to its radical revision in 1995. A complex program, it included content standards, or state curriculum frameworks, a set of state tests, and various accountability mechanisms. The formal intent of the program was to increase accountability to the state's curriculum frameworks and to move schools in the direction of greater emphasis on higher-order thinking, complex problem-solving on real-world problems, integrated subject matter, and application of basic skills.

Technical problems associated with any new test are not unexpected, but politics had an unforeseen, detrimental impact on ASAP. Despite its innovations, ASAP was never implemented. This report discusses the various causes for its demise.

¹ The source for the term political spectacle is (Edelman, 1988), who wrote that (pp. 4-5), "the conventional view [assumes] that rational choice may never be optimal, but is a central influence in decision-making and policy making.... [But] the phrase rational choice is one more symbol in the process of rationalization.... [A]ny political analysis that encourages belief in a secure, rational, and cooperative world fails the test of conformity to experience and to the record of history." Instead, politics and policy are matters of symbol, myth, and spectacle constructed for and by the public.

Vignette 1: Funeral Oration for an Assessment Policy
(Can Conventional Policy Research be Far Behind?)

It is Friday, January 21, 1995, a typical brilliant winter day in Phoenix. Tired but satisfied, the policy researcher watches as the printer churns out the last pages of her report. For three years she has been studying the consequences of the state assessment policy, the Arizona Student Assessment Program—everyone calls it “ASAP.” The centerpiece of the program is a performance test, which its founder, former state Superintendent C. Diane Bishop, has touted as the cutting edge of the national movement to reform schools by imposing alternative forms of testing. ASAP tests “the way kids learn best”—that’s been the slogan, by now familiar. It “makes teachers teach” toward application of knowledge rather than toward drilling bits of disconnected basic skills, toward “higher-order thinking skills,” toward “real” reading and writing. The researcher’s findings show that this ambitious program has had mixed effects, not surprising given its short history and implementation difficulties. But definitely, these problems could be fixed, with some good collaborative work and more time.

The phone rings. Beth is calling. Beth, who is not only an informant to the research study, but a virtual poster child for ASAP. As an elementary school principal she has thrown her all into embracing its aims.

“Have you heard the news? ASAP is dead.” In rueful and ironic tones, she relates to me that at a capitol press conference, the newly elected state Superintendent, Lisa Graham Keegan, has “suspended” ASAP Form D, the performance assessment, which was scheduled to be administered the very next week. Keegan based her decision on a report—by the test publishers themselves, no less—that said that Form D did not have high enough reliability and validity to give the state accurate accountability information.

Beth muses, “How could Lisa do that? How could she just undo a policy that the legislature mandated? Does this mean the state is going backwards? Back to standardized tests and drill and kill? What about all the work we’ve done? Teachers at this school are going to freak.”

The researcher is too taken aback to provide much consolation. Long since having given up on speaking truth to power, she has had the modest expectation that research could contribute to reasoned debate about the effects of state

assessment policy on school practices. But if ASAP is no more, who cares what its consequences would be? It will take her a year to regroup her research agenda.

The spectacle is far from over. Two weeks later, Keegan will announce that ASAP is not just suspended, but due for a complete overhaul. She will commission an Academic Summit, which will develop new state standards—in one month's time—with new state assessments to follow. These will emphasize basic skills and vocational skills and the accountability that has been sorely lacking up to now; that is, under the watch of her predecessor.

Vignette 2: Governor Enters Standards Drama

From Stage Right

Fifteen months later, March 25, 1996, what we thought would be the finale to the Academic Summit turns out to be a mere complicating action. As the summiteers and other educational policy watchers take their seats in the board room, the members of the Arizona Board of Education assume their leather high-back chairs behind a long curved table raised to imposing height above the spectators. The Board meets today to consider—we expect them to vote, yea or nay—the new academic standards developed during the Academic Summit. The summit and standard-setting process has dragged on more than a year already, whereas state superintendent Keegan had originally thought that three months would be enough time to rewrite the state curriculum frameworks and develop tests to measure them. The best laid plans had also imagined that the process would be simple and managed in such a way as to feature the concerns of parents and the corporate community rather than those of bureaucrats, professionals and curriculum experts. Sweeping out the old meant rejecting the progressivism deeply embedded in the Essential Skills and ASAP, which, she believed, had mistakenly abandoned traditional drill and practice, basal and textbook mode of instruction and high-stakes testing, in favor of student-centered, integrated, higher-order thinking and problem discovery and solving, and the like. The new standards and assessments would instead be clear, measurable, and focused on basic skills. But there was still a constituency for ASAP among the Summit participants that had pushed for progressive educational values, and the standard-setting process had turned contentious and drawn out as each faction struggled for purchase. We had followed this history with interest. But now, finally, at least the Language Arts and Mathematics design teams had worked out

their final drafts. We observers thought they represented compromises between progressive and traditional values—there was something in them for both sides. So, it was now time for the Board to make them official. Members of the Board had already indicated that they wanted to go forward so that test development could begin. The agenda called for short testimonies from the floor, and then a vote. No one expected much out of the ordinary.

The buzz takes us all by surprise, as Governor Symington strides purposefully and rapidly into the room and requests permission to address the Board. Trailing him are members of his entourage, who distribute copies of his prepared address to members of the press. Among them is his educational advisor, C. Diane Bishop, the former Superintendent. She says nothing, although people beside us wonder how she could let him “trash her baby,” ASAP, without so much as blinking. Just from the looks on their faces, we can see that the Board members and Keegan are just as surprised as the rest of us at this unprecedented intrusion and at what he has to say.

We will have to look later at the text to get the details, but the gist of his remarks is this: that the draft standards show the “reckless drift toward fads and foolishness” that characterize most of professional educators’ work, that the standards fail to mention phonics, spelling, or memorization of math facts and the state capitals. The arts standards come in for particular ridicule. Taking questions from the floor, he declares that the state should reject pointy-headed elitist professional jargon and just get back to basics and standardized testing for everybody.

Keegan is clearly flustered and tries to correct what she sees as Symington’s misreading of the standards, but the damage is already done. He leaves, the Board breaks, and the arguing goes on in the audience. “This just throws everything into a cocked hat,” sighs a member of a standards design team. “All our work, all our compromises. This is just a signal to his appointees on the Board to follow through on his conservative educational agenda and not give in to the professionals. If we had any doubts before, we now know that ASAP is really dead.” Actions of the Board will later prove her prescient.

Another observer reaches a different kind of interpretation, one that highlights the confluence of Symington’s bankruptcy and criminal indictments, his plan to seek reelection anyway, Keegan’s own gubernatorial ambitions, and

her open criticisms of his administration and calls for his resignation. Word was, he had even tried to lure her out of town to a governors' meeting on education that coincided with the board meeting. But that we'll never know.

INTRODUCTION

What does this story of political intrigue have to do with assessment policy?

The conventional view of assessment policy has trouble dealing with it. If policy is defined as the rules by which a society or institution is governed, assessment policy must be a state's rules and programs for determining who and what will be tested, on what sets of content, by what instruments on what schedules, and how the results of the tests will be aggregated and counted. In addition, assessment policy sets the functions that these testing programs will serve. In the contemporary scene, the functions of state assessments are likely to include accountability for achievement as well as reform of schooling (either to increase academic performance or impose curricular coherence across schools). Most often, assessment policy will embrace multiple functions (McDonnell, 1994).

The conventional model of school reform assumes that states develop policies by consensus, state departments develop programs and policy instruments consistent with the goal consensus, and educators respond rationally and predictably to implement the programs. In the case of assessment policy, a state adopts academic standards and frameworks to bring about curricular coherence and increase achievement. Tests are chosen as the tools or instruments to attain policy goals (McDonnell & Elmore, 1987). By placing consequences (high stakes) on test results, the state expects that students, teachers, and school authorities will focus their energies on the policy targets. The conventional model proposes that schools change because of state policy, in mechanistic ways (though the precise mechanisms are unknown), rationally, with a sense of shared vision, and more or less predictably and uniformly. Even slow or variable response can be explained on rational grounds; e.g., by lack of resources or knowledge. Furthermore, a state's standards and assessments are assumed to be a fixed object, an invariant target for schools to aim their efforts.

Against this conventional model, we consider an alternative, political model. In the alternative model, assessment policy is more like a moving target

that is variously constructed by political and policy actors as well as the educational practitioners who must respond to it. Politics at both a macro and micro level influence these constructions. The process contradicts assumptions of rationality and uniformity, occasionally exemplifying the spectacular, though often hiding the ugly pushing and tugging for power, resources and political agendas behind a facade of rationality. By politics we mean not the usual view of the contest between Republicans and Democrats but the dynamic process wherein partisans contend for power, prestige, position and ideology in official (governmental or institutional) capacity. The enactment of assessment policy is as much a symbol over which the partisans contend as it is a deliberate technique to change schools. We believe that this alternative model provides the best explanation for the birth and death of ASAP—the fundamental alteration of a state’s assessment policy.

This paper rests on both empirical and theoretical foundations. Empirically, there are three sources of data. First is a policy study conducted to discover the images, ideologies, interests and tactics of policy actors at the outset of the Arizona Student Assessment Program, using interviews and documents as data sources (Noble, 1994). Second, a two-year study investigated the response of educators to ASAP employed both long-term qualitative and extensive survey data sources (Smith, 1996). Third, a set of policy actor interviews, observation of key events, and document analysis provided evidence on assessment policy at the end of ASAP and the beginning of the next phase of Arizona assessment policy. Each of these empirical projects yielded findings carefully warranted in data. The analysis of the whole is represented here as a narrative account, a series of vignettes and interpretive commentary. Our analytic choices were influenced by Rein (1976) who recommended that the policy researcher act as moral critic, adopt a historical perspective, identify and question assumptions behind policy choices, and construct narrative accounts or stories

To explain these data, we turned to several theoretical frameworks in the literature.

Marshall, Mitchell, and Wirt (1989) argued that there is reliable connection between policy and a state’s political culture. They define culture as persistent patterns of values that can predict the behavior of policy actors who contend with each other for the power to allocate these values in the form of policies. They posed three alternative political cultures—moralistic, individualistic, and

traditionalistic. Through their empirical study, the authors identified Arizona's political culture as traditionalistic, wherein the dominant values behind state policy are efficiency and choice rather than quality or equity. Traditionalistic state policy cultures emphasize "the leading role of economic elites in shaping public decisions, with a consequent fusing of private and public sectors and a limitation on citizen participation" (p. 118), a distrust of bureaucracy, labor unions, and professional (e.g., teacher and administrator) authority and concerns (e.g., professional development and certification standards). Furthermore, there are strong anti-taxation sentiment and persistent demands for accountability in political cultures such as Arizona. Localism is valued over central control of public policy. Equity issues—the recognition of disparities and injustices among groups and the use of policy to correct them—receive low priority. Government is viewed as a means of maintaining the existing order (rather than as a marketplace or as a commonwealth as it is viewed in the individualistic and moralistic cultures, respectively). Though specific policies and partisan political configurations may change, the dominant political culture persists and reasserts itself over time. For our purposes, a state political culture predicts which policies are likely to persist.

Other structural issues that transcend state political culture must also be considered. First, the national discourse of crisis due to public school failure serves as backdrop to assessment policy change, beginning with *A Nation At Risk* and so often reinforced that evidence and argument that ought to disconfirm the perspective (e.g., Berliner & Biddle, 1995) go unnoticed and even ridiculed. Second, the national discourse that relates education to the national economy (House, 1991; Ball, 1990) certainly shapes assessment policy as well. That is, the national discourse makes commodities of test scores, attributes economic prosperity to higher achievement test scores, and prioritizes the concerns of corporations.

For our study, explaining what happened to ASAP begins with these ideas of Arizona's political culture. The partisan dynamics in the state during the period of this study also refracts the data. Early on, the policy actors were split between the parties. Later, Arizona became virtually a one-party state. The governor, state superintendent, and legislative majorities were conservative Republicans, and the appointments they made to the State Board of Education, Arizona Department of Education staff, and various ad hoc groups reinforced

their perspective. The dominant discourse was union-bating and educator-bashing, federal mandate- and court order-defying. Right-wing extremists often made the news, as did religious conservatives. Assessment policy could hardly be immune from this climate, particularly because of the relationship between political and pedagogical conservatism.²

Although political culture provides a matrix for understanding policy activity, it fails to account for the spectacular events in Arizona assessment policy. For that we take the perspective that policy becomes real at the point of the interaction; at specific times and places in which particular actors encounter it. These interactions occur at many points: when a policy agenda reaches the table, when it is enacted, administered, and implemented. At each stage, the policy is interpreted by the actors involved in ways that may have little to do with the official, written policy itself. Controversial policies (Schon & Rein, 1994) become part of the political spectacle and take on the aspects of symbol, rite and myth more so than technical instruments aimed at real changes (Edelman, 1985). The garbage can theory of policy making suggests policy entrepreneurs identify problems to link to available solutions (rather than the other way round, as most rational actor models imply). The linkages are possible only in a short policy window in which the frequently contradictory images and interests of policy constituencies can be obscured long enough to get a policy on the agenda (Kingdon, 1995). Once a policy is legislated, it can go through substantial transformations from its founders' original intentions as it is reinterpreted at the levels of administration, implementation, and reaction (Hall, 1995). These frameworks suggest categories and propositions for interpreting data in this study.

² Examination of documents reveals that right-wing organizations typically extol the virtues of teaching reading by the phonics method and math by memorization of math facts. These pedagogical techniques are viewed as within the purview of the family and not the professions (Phyllis Schafley claims that any mother or grandmother can teach a child to read, for example). They repudiate the teaching of higher-order thinking, whole language and bilingual education, and other recommended approaches of progressivism and constructivism (Dewey and Vygotsky's communist leanings make these perspectives suspect). Especially, Outcomes-Based Education, which they define in an all-inclusive way, comes under fire. These preferences have the characteristics of fixed ideologies, in that they seem founded in biblical interpretation, are immune to fair debate, and tend to demonize the opposition. Refer to more extensive treatment of this subject in (Berliner & Biddle, 1995) and (Gaddy, Hall, & Marzano, 1996). We do not imply, however, that all proponents of phonics-based education are part of this ideological community, only that there is a pattern that connects political and pedagogical conservatism.

The remainder of this paper is organized this way. First, we describe the components of the ASAP program. Then we detail its history—assessment policy that antedated ASAP, its legislative and ideological origins, its initial administration and how ASAP evolved, the crucial change in assessment policy associated with a change in policy entrepreneur, the suspension and demise of state performance assessment, the development of new standards and assessments to replace ASAP, the political rivalries and struggles that explain the sequence of events. As much as possible, we let the words of the actors and the documents speak for themselves. Footnotes are used to expand the details available to the reader. Finally we synthesize the theoretical and empirical elements of the study together.

What Was ASAP?

The Arizona Student Assessment Program (ASAP) was the official state assessment policy from its inception in 1990 to its radical revision in 1995. A complex program, it included content standards, or state curriculum frameworks, a set of state tests, and various accountability mechanisms. The Arizona Department of Education derived its authority for administering the program from Arizona Revised Statutes 15-741. Analysis of state documents (Olivarez-Seck, 1994) shows that the formal intent of the program was to increase accountability to the state’s curriculum frameworks and to move schools in the direction of greater emphasis on higher-order thinking, complex problem-solving on real-world problems, integrated subject matter, and application of basic skills. In its entirety, the program included the following elements.

1. **Arizona Essential Skills.** The state curriculum frameworks in reading, writing, mathematics, science, social studies, health, foreign language, music, performing and visual arts purported to reflect high levels of expectation for all students, application of basic skills, problem-solving abilities, and higher-order thinking. They included benchmarks of what pupils ought to know at grades 3, 8, and 12.

2. **Performance Assessments.** Assessments of clusters of Essential Skills were administered to all students in grades 3, 8, and 12, in the form of ASAP Form D, the on-demand or “audit” form of the state assessment program. The state published guidelines that were intended to insure that Form D would be administered under standard conditions and scored under auspices of ADE

(Arizona Department of Education), using the “generic rubric” or 4-point scoring guide. Form D tested in integrated style (students had to write an essay or some other extended form in response to a reading assignment with an embedded math problem). Four variations of Form D (D-1 through D-4) were to be phased in over four years, each testing one-quarter of the Essential Skills. ASAP Forms A, B, and C consisted of performance assessments in reading, writing, and math to be used for preparing pupils to take the Form D, or as instructional packets and for district assessments (see item 4). Forms A, B, and C tested the content areas separately, purported to measure all the Essential Skills, and were administered and scored by teachers, also using the generic rubric, rather than at a central scoring site, as Form D was. Spanish-language versions of all forms of the performance assessment were available. There were guidelines for teachers to modify test administration conditions for handicapped students. The performance tests were developed by Riverside Publishers, which also had contracts for scoring services and technical reports.

3. **Norm-Referenced Testing.** The Iowa Test of Basic Skills and Tests of Academic Proficiency provided a means for comparing the achievement of Arizona schools with that of a national norming sample. A limited battery of tests was given to students in grades 4, 7, and 10 during the fall months.

4. **District Assessment Plans (DAP).** The DAP served as a compliance tool. Every district had to submit a DAP each year to the Arizona Department of Education, which reviewed and approved it or asked for revisions. The plan specified the method by which each Essential Skill would be measured and the grade level at which it would be measured. DAPs provided assurances that districts would measure students’ mastery of the Essential Skills by grades 3, 8, and 12 (each district set its own level of mastery). Districts could choose which of three methods to use for its DAP testing: ASAP performance assessment Forms A, B, or C or a system of portfolio assessments or criterion reference measures. Either of the latter were acceptable to the ADE if the generic rubric could be applied to the results. In response to a 1994 policy adopted by the Arizona Board of Education, ADE planned to use a revised version of Form A assessments as a graduation competency battery.

5. **Essential Skills Reporting Documents (ESRS).** Each district was required to report annually to ADE on the number and percentage of pupils that had attained mastery of Essential Skills and report results of achievement testing and

nontest indicators (e.g., amount of time spent in homework, number of books read, and the like).

6. Report Cards. In June of each year the Arizona Department of Education issued report cards for each student, school, and district, as well as for the state as a whole. The state report card reported descriptive statistics on the assessments, aggregated to the county and state levels. Demographic data and non-test indicators were also reported. School and student report cards were proprietary, with individual reports to parents and school reports to the district. Other reports were public documents.

7. District Goal-Setting. Districts were to report annually to ADE, detailing the goals for the subsequent school year, based on results of all the assessments. In addition, the report listed the strategies for reaching these goals and budgets, and timelines for implementing those strategies.

Looking back at the paragraphs above, one gets the impression that ASAP was complex but clearly delineated. This impression is pure historical fallacy, however. It is doubtful that any one individual during 1995, whether policy actor, politician, or educator, had the whole list in mind. From the vantage point of the average teacher or principal, one would only “know” ASAP as a particular set of performance tests and perhaps a set of reform ideals. What one knew was what one had experienced directly. This included statements in the press or during meetings that “ASAP was the best we know about how students learn,” or that ASAP represented authentic assessment, integrated learning, a new role for teachers, a way out of the quagmire of high-stakes standardized testing and the stale brand of teaching that seemed to follow it. The accountability aspects and intents of the program came later to awareness, and to some, not at all. Our surveys and qualitative studies (Smith, 1996) showed that ASAP was not the same thing for everyone. Later, these ambiguous and multiple meanings would explain the confused reaction to Keegan’s suspension of ASAP Form D. Those who equated Form D with ASAP found it difficult to grasp that the DAP, accountability, norm-referenced tests, and Form A provisions were still in place. The conceptual confusion also intruded on the research process. When anyone mentioned ASAP, we had persistently to deconstruct *which* ASAP they were thinking about at that moment.

Assessment Policy Prior to ASAP

You have to go back in time before the advent of ASAP to understand its place in people's consciousness as a reaction against standardized testing. Under the previous state assessment policy, Arizona students experienced one of the highest test burdens in the nation. Legislation mandated the administration of the Iowa Test of Basic Skills (ITBS) to every pupil in grades 3-8 (the Tests of Academic Progress in grades 9-12). The full battery of both tests (products of Riverside) was given in the spring of every year and the results published by school and grade. Although the state placed no consequences on the scores of students, some districts based decisions about salaries of administrators and teachers in part on test results. High stakes are often in the eye of the beholder, however, as teachers reported a high degree of embarrassment, shame, and pressure when the results (reported at the district, school, and grade level) came out in the newspapers each summer.

In addition to the norm-referenced testing, the state also required that each school district prepare and administer tests of basic skills, a list of which had been first approved by the State Board of Education in 1983. Dissatisfaction with the list resulted in the development and authorization of the Arizona Essential Skills, for which the state would hold districts accountable. In 1987, the Board of Education appointed representative groups of educators and content specialists to committees that wrote the content frameworks and revised them based on extensive hearings around the state. Staff of the Arizona Department of Education guided the work of these committees toward the newly emerging principles of constructivism and progressivism.³ Almost a decade later, an influential staff member reflected on their work:

The Language Arts Essential Skills was very different from the polyglot of little skills that had been around for awhile. The Language Arts Essential Skills looked at what people were learning about from a constructivist philosophy of education. It looked at what the writing teachers were saying and the writing professors, writing research was saying about writing as a process. It looked at new ways of reading instruction, somewhat whole-language based or literature, and it looked at integrating the language arts so that you didn't teach reading separately from writing. You looked at how they supported each other. And, finally, it looked at the possibilities of assessing

³ Like the work of all such committees, however, the finished products were compromises, so that even traditional basic skills schools could interpret the Essential Skills as accommodating their own instructional preferences.

language and other subjects directly that—instead of judging whether students can write by having them do a multiple choice test, it looked at having students actually write, and assessing that writing. So all those things came together to produce the Language Arts Essential Skills, but the conflict, the problem that had to be solved, was that the testing the state was doing was the most invasive in the nation.

Both anecdote and research studies portrayed educators' dissatisfaction with state assessment policy (Nolan, Haladyna, & Haas, 1989; Smith, Edelsky, Draper, Rottenberg, & Cherland, 1989). High-stakes uses of standardized testing stood in the way of curricular reform. Said a progressive reformer:

What became evident to everyone was that we had these curriculum frameworks out there that were representing the latest and the best thinking in the content areas, and those were mirroring what we know about the way people learn. And that the dramatic difference between what we said we wanted and what the tests were measuring, and then the new ideas circulating in the testing circles about what you test and what you get, and how you test, and how you get it, made it almost imperative at one point for the testing to be examined. I don't think testing in Arizona would have changed a whole lot if we hadn't had the curriculum framework that suggested a very different way of testing, that you can't test writing as a process with a multiple choice test.

The Birth of ASAP

One can trace the origins of ASAP to the dissatisfactions of various policy actors and constituencies. In the late eighties, the Board of Education and the Arizona Department of Education seem to have arrived at a shared definition of "the problem"⁴: that existing state tests failed to cohere with the newly developed progressive, constructivist content frameworks. C. Diane Bishop was an influential member of the Board in the late eighties and became state Superintendent and head of ADE in 1991. Then a Democrat, she taught high school mathematics and thus "understood higher-level thinking in mathematics." Bishop's ideas found support among the curriculum specialists at ADE, the Arizona English Teachers Association, the Arizona Education Association, local affiliates of the National Council of Teachers of Mathematics and the Center for Establishing Dialogue in Education, and local university professors. ADE commissioned two research studies that contributed data to this common definition of the problem. One study compared the content of the ITBS

⁴ We define "a problem" as a social construction by policy actors (Edelman, 1985) (Kingdon, 1995).

with the content in the Essential Skills and found that only 26 percent of the Skills were tested. A survey (Nolan et al., 1989) showed that most Arizona educators disputed the validity of the state norm-referenced tests, spent too much time preparing students to take them, and believed that the tests had deleterious effects of students, teachers, and the curriculum.

A powerful policy constituency at that time consisted of a group of teachers who believed strongly that students construct knowledge actively and intentionally from their interactions with texts, teachers, and other students; that reading, writing, and problem solving are parts of a whole, and teaching is best when it acknowledges this. These educators believed that one structural barrier to expanding this mode of teaching and learning was the state-mandated standardized testing program. Standardized tests, they believed, encouraged teachers to teach in ways that mimicked the form of mandated assessment. That is, since the ITBS assessed spelling by the ability of the pupils to identify which of four words happened to be misspelled, that teachers would teach spelling in precisely that way—using work sheets that looked like the test items themselves and consisted of recognition of misspellings. Authentic writing would be postponed and de-emphasized altogether while classes spent time perfecting spelling and grammar. There was empirical support for their view that state testing narrowed curriculum and restricted instructional methods (Smith et al., 1994). If the state would only stop with all the standardized testing, they reasoned, the way would open for better education. Alternatively referred to as whole language or constructivist teachers, they had collaborated with language and university professors of language arts in an organization titled the Center for Establishing Dialogue in Education (CED). A separate organization of many of the same educators had successfully lobbied for a change in legislated assessment policy to exclude first-graders from ITBS testing. Flushed with this success, the group kept up pressure on the remaining assessment policy and turned out to be a natural ally to those ADE staffers who shared their perspectives on progressive education. Common ground was also found by language educators and activists who believed that standardized tests adversely affect children whose first language is other than English. They wanted any assessment policy to allow students to be tested in their home language.

The momentum for change was building. Even within this coalition, however, one can uncover alternative views of what was problematic with

existing state assessments. One subgroup believed that the ITBS detracted from efforts to reform instruction toward progressivism and constructivism. The other believed that the ITBS did not provide adequate accountability to the Essential Skills.

It was the perspective of the latter group, which desired more accountability for particular outcomes, that resonated most with the policy constituency with the power to make change: the legislature. Asked later about the function of ASAP, a member of this group said:

This assessment is an accountability measure, because we want those Essential Skills taught. And the only way we know that it's going to be done is if you drop in and take an assessment of that ... because there really have been no accountability measures up until now.... It was a matter of here we have the Essential Skills and I think there was ample evidence that many school districts weren't getting at that. It was, "This too shall pass," and teachers were still teaching what they were teaching. They weren't focusing on those Essential Skills. I think that was a driving force to put this all under a legislative piece and put a little teeth into this thing.

Absent from the above quotation is even a nod in the direction of reform toward progressivism. From the sum of data from policy makers analyzed in the early nineties, we find little hint that the legislators involved in the birth of ASAP had concern or understanding of those principles of schooling that so motivated the policy actors at ADE and in the professional associations. Nevertheless, the actors (legislators, ADE, Superintendent, and Board of Education) came together in the Goals for Educational Excellence project to develop new assessment policy and write enabling legislation. The report the project issued in November, 1987 concentrated on accountability principles more than reform ideals. For example, "The keys to the future were ... a combination of basic skills—communication and computation—as well as skills such as citizenship, interpersonal skills, thinking skills, and developing creativity." And "education must emphasize measurement of results to be accountable for accomplishing its goals." Little progressivism and reform there.

Arizona Revised Statutes 15-741 that became effective in May, 1990 required the State Board to: (a) Adopt and implement Essential Skills tests that measure pupil achievement ... of the state board adopted Essential Skills in reading, writing, and mathematics in grades 3, 8, and 12; (b) Ensure that the tests are uniform across the state, scored in an objective manner, and yield national

comparisons; 3) Conduct a survey on “non-test indicators;” (c) Require districts to submit plans for assessment of Essential Skills at all grade levels; (d) Publish report cards at the pupil, school, district, and state levels; and (e) Require norm-referenced, standardized tests at grades 4, 7, and 10. In addition, the legislation affirmed existing (but not previously enforced) provisions for a policy of promotion from grade to grade based on achievement of the Essential Skills.

The legislation itself never mentions the Arizona Student Assessment Program nor commits schools to any principles of practice at all, nor to any particular form of testing. Performance assessment, alternative assessment, authentic assessment—these terms were not codified in the law. The only thing that supported the progressive agenda in the 1990 legislation was decreasing ITBS testing and moving its administration to the fall, when its results would serve diagnostic rather than accountability functions. Everything else about ASAP was a radical transformation by Bishop and the progressives then serving at ADE. But no one knew it at the time, which made the later suspension of Form D so much of a shock.

Policy actors would later report a high degree of consensus among the Board, legislature, and ADE at the outset of the ASAP program. Yet what they failed to see was that these agencies were agreeing to quite different things. The legislators believed that they were promoting greater accountability as a result of the legislation, while parts of the department and board believed that the state had embarked on a bold new vision of teaching and learning. This confluence of alternative, even internally contradictory perspectives reflects Kingdon’s theory that policies usually obscure underlying contradictions in values and perspectives of political actors whose various agendas come together temporarily (Kingdon, 1995). Indeed, the legislation would probably not even be passed if the contradictions were brought to the surface. The ambiguities and contradictions then send conflicting signals to those who must implement the policy and those who are supposed to react to it.

ASAP as Assessment Policy: A Moving Target

Most educators never came into contact with the legislation itself and probably would have been surprised by its wording if they had done so. What they were exposed to, in contrast, was the program implemented by ADE and the extensive communication about the function of ASAP to change curriculum and

teachers' instructional practice by changing the nature of the test. Every communication from ADE, the extensive series of meetings, workshops, newsletters, and the like, all trumpeted the merits of performance assessment and the kind of education that was consistent with it.

The rational policy model assumes that a policy has an objective and fixed reality. Interactionist theories (Hall, 1995; Lipsky, 1980) assume that the text of a policy is only a point of departure for the persons who interpret and implement it and may undergo substantial unwritten revision and transformation over time and levels of the system. The moving target of Arizona assessment policy, rife with internal contradictions and ambiguities from the outset, evolved and diversified in relation to (a) shifting power among coalitions of policy actors, (b) variations and limitations in capacity-building efforts, (c) confrontations with the tests themselves, and (d) scarcity of time and resources.

Three reform-minded officials at ADE, supported by the Board, AEA, and constructivist educators set the tone for the program at this stage. These individuals were effective and outspoken advocates for alternative assessment and for reforming instruction to be more holistic, thematic, and aimed at students' active engagement and capacity to make connections, solve complex problems, and communicate their thoughts. They believed that assessment must be authentic and integrated with instruction, subjects integrated with each other around interesting, real-life problems, that teachers should be co-learners, coaches, collaborators, facilitators of learning, and actors rather than targets of curricular reform, that instruction should follow new research on cognition, multiple intelligence, constructivist learning theory and the like. They also believed that mandating a performance test would move teachers away from reductionistic, basic skills—drill and practice—teaching of subjects in isolation, the worksheet curriculum. As one of them said later, “we saw the potential of doing something special ... the chance for the state to break the lock of the norm-referenced testing which was not serving teaching or learning very well.”

The Department public discourse rarely included the accountability aspects of ASAP in those early days, an omission that fit the reformers' agenda. Later, one declared:

... [ASAP] was never intended to be [a high-stakes test], never intended to be used as a determiner for whether or not students graduate, never intended to be a determiner of whether or not they go from grade to grade.

Initial reactions of educators to ASAP were positive, for they believed that whatever else the policy entailed, it was preferable to the widely despised ITBS, with its accountability overtones. Opposing voices cited the subjectivity of scoring the performance assessments and feared that schools would de-emphasize basic skills, but their decibel level was nowhere near the level it would attain later.

Against this reform coalition, however, was a legislature that was still more interested in holding teachers' "feet to the fire" than in changing the way they taught. As its composition became more conservative and Republican over time, one heard more complaints about, for example, the "subjective" scoring of the performance measures and the fact that "anti-business and environmental activist attitudes" had crept into the content of the tests.

Within ADE, the staff was not of one mind about assessment policy, even at the beginning. Bishop herself had always spoken of ASAP as a way to focus teachers' attention on the state's curriculum frameworks and students' attention on mastery of the Essential Skills. Staff and officials concerned with ASAP were organized in two different units, with the reform agenda represented in the ASAP Unit. Meanwhile, the Pupil Achievement Testing Unit, made up of individuals who were experienced with norm-referenced and criterion-referenced assessment (not performance assessment), aligned themselves with the accountability values of the legislature. Two factions—two different sets of values and priorities and definitions of assessment policy, each privately discounting or talking past the other.

About three years into the development of this program, however, the three key reformers left the department. Remaining staff proved to be less thoroughly grounded in progressive educational principles and practice and less effective in maintaining the direction that ADE initially took. While the ASAP Unit changed faces and voices, the Assessment Unit remained consistent. By that time, the accountability forces within the department had begun to dominate the discourse. More ADE staff time was spent standardizing and monitoring and

districts' assessment, reporting, and goal-setting procedures. Superintendent Bishop and the testing department now set the tone.

A critical element to assessment policy was missing if its intent was to change the way teachers teach. Legislation that mandated state tests had failed to authorize funds to provide for teacher training in the principles and practices of performance assessment and of holistic, thematic, and problem-solving instruction and curriculum reform. For a reform of this type, extensive time and training have proved in other states to be necessary (Flexer, 1994; Shepard et al., 1995). The report of the Goals for Educational Excellence panel had earlier promised the legislature that the new program would cost no more than the previous testing program, and this efficiency value in the state political culture foreshadowed subsequent problems. Now ADE was in a bind, able to mandate assessment policy but powerless to fund state-wide training of teachers to adapt to it. Some districts with sufficient wealth and officials who were open to the new direction suggested by ASAP invested considerable resources in local capacity development, but these were in the minority. In a state with considerable disparity in taxing ability, the already rich and poor districts reproduced disparities in staff development for ASAP as well.

The pace of transformation of the program away from its reform agenda increased under the press from key legislators to start producing data for accountability purposes, which was of course their definition and intention all along. As an ADE official reported later, the department chose to underplay the technical and administrative problems that had surfaced along with capacity building needs:

We should have gone back to [the Legislature] and said, "we're going to need some more training money, we need more field test money," but the things looked good, they had been sent out to the schools, teachers saying let's get going, we want to do this. The Legislature was saying let's get going, let's get going. At that point what should have happened is we should have said we need two more years. We need another state-wide pilot, we need more of the psychometric people in making sure the thing is ready to go, and we need additional district training budgets so when they come on line with this they could train their teachers. We had underestimated the profound training effects that this would have, clearly underestimated what it would be."

Because it was a political project, ASAP had limited time either to develop the capacities of teachers and schools or to develop sound psychometric

instruments. As a result of political pressure, it was necessary for the ADE and the test publisher to produce the various performance test forms in weeks rather than years. Form D-1 was commissioned and administered before all the psychometric and administrative kinks of Form A were worked out, D-2 was commissioned and administered before the characteristics of D-1 were corrected or even known. Nor was there an equating study to show whether Form D could function as an “audit”⁵ of Form A. An ADE official would later recall that the development of D “was done in a fairly shabby way, without adequate field testing.” But ADE “didn’t act on this information for a couple of years.” An ADE insider at the time agreed:

Now, the problem there was that the first Form D was used, we tried it out, we reported the results, but Riverside ran a concurrent field test on the form D. The concurrent form D-1 field test was returned to the department in late ‘93, or the fall of ‘93 sometime. And what it said is that the D form didn’t match the A form well enough. But, for whatever reason, the staff of the Department kind of took that report and put it on the shelf. Because what it said is we want to do the D form. Politically, the thing was developing its own momentum down there. Nobody wanted to stop the process, nobody wanted to pull it back. Riverside staff was saying you’ve got to stop this because the D now needs to be revised and re-field tested to be sure that it matches the A that it’s auditing. Wasn’t done. [The report] was shelved ... and D-2 then was commissioned, ... was developed, was not field-tested, and was ready to go as the next state-wide audit.

Nor was there time, regardless of intent, for ADE to seek independent evaluation of the performance assessments or consultation by experts in the incipient technology of performance assessment. During interviews conducted later, the contractors also noted that development time was too short and that the state had overlooked the ramifications of getting the assessments in the field on such a short timeline.

⁵ The notion of using Form D as an “audit” evolved over time, though insiders would claim that they intended that function all along. However, officials early in the Bishop administration suggested that Form D would be used as an efficient monitor of districts’ scoring and reporting of their local (DAP) assessments, which consisted of Forms A, B, and/or C, criterion-referenced measures, or portfolios. Districts could choose which assessment form they wanted, provided it could be scored by the state generic rubric and contingent on ADE approval. Although the word audit implies in the business community that an independent professional has verified that the company has used the proper *procedures* in its financial statements, ADE operationalized the concept to be the correlation of *results* of state-administered and scored Form Ds with the DAP assessments, at the individual pupil level.

Beyond ignoring the problematic technical data of early versions of the state performance assessments, Superintendent Bishop and other ADE staff reacted defensively to any criticism of ASAP. At a meeting of educators sponsored by ADE and AEA, she reacted to objections about ASAP by warning that if teachers complained too much, the conservative policy actors would likely move to reinstate universal standardized testing. At the time, the complaints seemed quite reasonable and problems for the most part correctable, having to do with glitches in administration, the burden of purchasing test materials, question wording, insufficient time limits, inadequately prepared scorers, vague scoring rubrics, and lack of time and training. The former president of AEA recalled that the superintendent “lost it” when informed of the teachers’ position, however. Open debate over assessment policy did not happen.

In June of 1993, the initial results of ASAP Form D were reported. The newspapers published the results by school and grade level and ranked them in much the same manner as they had always reported the standardized test results. *The Arizona Daily Star* headlined its report, “Tests say schools are failing.” The Superintendent called the results disturbing and distressing, but failed to note the possible technical problems associated with any new test undergoing its maiden voyage. She criticized schools and teachers for not adapting fast enough and for not teaching “the way kids learn.” Educators were shocked and dismayed at ADE and media reaction. Many had believed or were led to believe that performance assessments were to have a different function than had standardized tests. Instead, the high-stakes accountability function of ASAP was fully revealed.

Time and political capital had begun to run out for the reform faction in ADE. The State Board of Education, prompted by key legislators, demanded action on the accountability front. The use of ASAP in determining high school graduation became part of assessment policy in January 1994 through the action of state Board of Education rule R7-2-317. A Task Force on Graduation Standards was then appointed to make recommendations about proficiency levels. Its recommendations were later adopted, specifying a level of proficiency for graduation from grade 12: “A student shall demonstrate competency in reading, writing, mathematics, social studies and science ... by attaining a score of 3 or 4 on each question or item of each Form A assessment [of ASAP] ... scored with the corresponding Essential Skills (ASAP) generic rubric...” The Task Force had met

a number of times and had self-administered the Form A tests for 12th grade and used the generic rubrics to score them. A member would later report that they considered the rubric scores in terms of percentages, as if the assessment was like a competency measure. That is, four was the highest score on the rubric, and 3 of 4 was close to 75 percent, and a less than 75 percent mastery level would be taken by the public as too lenient. Therefore, a 3 would be the cut off between mastery and non-mastery, between graduation and non-graduation. There is no evidence that the Task Force examined technical data (for example: standard errors around cut- scores) or consulted experts on established procedures for setting cut-scores. It specifically ignored the Riverside technical report that warned against the use of Form A for pupil level reporting, let alone accountability. The Task Force report listed its rationale this way:

After reviewing every question/item on the Form A assessments and sample student responses at each level of the 4-point rubric, it was determined that a score of less than a "3" would not represent an adequate demonstration of competency.⁶

In addition to recommending proficiency levels for graduation, the Task Force also suggested that limited English speakers be allowed to demonstrate their proficiency in their own language and all students who initially fail should be able to take the test again.

By 1995, what schools were attempting to implement was many things to many people. Administering ASAP was more about high-stakes accountability, more about standardizing and centralizing education at the state level, and less about progressive reform than one would have predicted from the discourse of 1992. The political nature of this assessment policy was revealed in the press of time, the shifting power balance, and the subversion of balanced debate about where this program was headed. With all that, one may wonder whether any effects were possible.

⁶ A sample from the Generic Reading Rubric: "A 3-response demonstrates an adequate understanding of the text. There is evidence of understanding of both the gist and specific parts of the text. It is not as complex as a 4-response. It may include minimal extensions, such as connections to other texts, experiences, abstractions and/or generalizations. All elements of the question are addressed in the response."

Consequences of ASAP

Was ASAP effective in achieving its reform aims? The state neither conducted nor commissioned a rigorous evaluation of what happened as a result of ASAP. ADE monitored compliance through the DAPs and conducted a survey of teachers, but one would be hard-pressed to call this a serious evaluation. Our independent policy study of the consequences of ASAP, however, tracked implementation and reactions for nearly the complete life of the program. Beginning in 1992, the year in which a pilot administration of Form A was conducted, we studied policy makers as they initiated administration of the program and transformed its initial intents. Then in the first year of Form D administration, we conducted in-depth case studies of schools while they were in the throes of accommodation and reform. In the next year we continued our observation of the case study schools and conducted a representative survey of Arizona educators about response to ASAP (see Appendix A for a more complete description of methodology).

What we found (Smith, 1996) can be summarized as follows: Arizona educators were fully cognizant of ASAP, though they defined it in quite different ways: as a preferable alternative to standardized testing, as a way to change teaching in a constructivist direction, or as just another state mandate. Somewhat less than half of the educators we studied approved of ASAP as they defined it. Much of the disapproval seemed to be the result of the implementation of ASAP Form D (e.g., inadequate time limits or directions, inappropriate item content or scoring rubrics, the use of ASAP scores for high-stakes purposes, etc.) rather than the *idea* of ASAP. Change in curriculum and teaching consistent with ASAP also varied widely and depended on certain local characteristics. For example, if there were adequate financial resources to make the change and knowledgeable personnel to help, if the existing teaching patterns and beliefs were amenable to constructivism, if there was little commitment to standardized achievement testing and traditional education, then change was evident, but not otherwise. The low rate of change can be attributed in part to inadequate professional development. The state failed to provide resources for teacher training, even though the reform implied fundamental changes in teacher knowledge and skill. This left the responsibility in the hands of the districts, many of which were too strapped financially to do anything. Although a few districts devoted impressive resources to develop teachers' capacities to

implement ASAP reform ideals and invested in curricular changes as well, the average number of hours of relevant professional development reported by teachers across the state was only about eight hours over a two-year period. Still, there was enormous effort spent by teachers and administrators simply in complying with ASAP testing and reported requirements. At the least, the evidence points to wide-spread increase in students writing extended texts.

Though our surveys and case studies were made available to ADE staff, they never publicly acknowledged the results and we saw no signs that the research had any effect on subsequent decisions. Later on, when we interviewed policy actors subsequent to the demise of ASAP, we found that their interpretations of the consequences of ASAP fit their current agenda and position. For example, an ADE insider during Bishop's administration said this about what happened as a result of ASAP reform:

Well, the great strength of it was that it was changing the behavior in the classroom. We were making change with respect to teaching methodology, instruction, technique, those kinds of things. Also materials were being changed, moving from a reliance on rote memorization, teachers primarily engaged in lecture and students repeating back what they heard, to one where students were actually engaged in the application of knowledge. Teachers were engaging the students in the learning process more. In other words, they would have to solve problems, they would have to apply whatever they had learned in the classroom to a real live situation. They would have to actually write.

But a legislator reflecting back on the program noted:

[There were] constant complaints about the content of the test, about perceptions that the test wasn't valid, that it disrupted the classroom—just constant complaints, and zero—zero [support to keep it going].

A district administrator noted that the limited consequences of ASAP might be due to the slow pace that schools have of changing anything and the complexity of the particular change involved in moving from basic skills teaching to more thematic, problem-solving, constructivist teaching. Teachers in that district were finally ready for ASAP Form D just at the point when it was suspended:

They were all ready to go. They were just getting to the point where teachers were taking it seriously and gearing their teaching to it. The first year no one took it

seriously, because the district was convinced that ASAP was not going to last, and we didn't do too well. The second year teachers were just starting to get going, but they didn't have the understanding to do it well. By the third year they were ready and eager to show what their students could do. But those teachers that modified their teaching will go ahead in that direction. The instructional aspect really had an effect.

However, the pace of curricular and teaching change failed to keep pace with the pace of political change.

Cast Changes/Assessment Policy Changes

Bishop's first term of office was due to conclude in 1994. For reasons unrelated to policy, she decided not to seek a second term of office. The Democrats nominated a prominent official in the Arizona Education Association, the state's largest professional association. The Republican nominee and subsequent electee was Lisa Graham Keegan, a state legislator and chair of the Education Committee.

The 1994 election also featured the reelection bid of Governor Fife Symington, a prominent businessman who shared politically conservative views with Keegan. They were particularly in tune over the issue of vouchers and school choice. He was opposed by another businessman, Eddie Basha, who enthusiastically supported public schools and opposed voucher initiatives.

In a move that took everyone by surprise, Bishop bolted the Democratic Party and campaigned both for Symington's reelection and for the support of charter schools and vouchers. After Symington was elected, he appointed Bishop to a newly created post as education advisor to his administration. Since then, the consensus view is that she has been nearly invisible as a policy actor.

Few could be considered more visible than Keegan, a bright, attractive, articulate woman in her thirties, a Stanford-educated speech therapist. She was termed the "Voucher Queen" after two terms as a highly visible state legislator, according to a subsequent article in *Education Week* (9/25/96, p. 20, "Ariz. Chief Puts Name and Face Front and Center"). A very powerful speaker, she called her approach "populist," "emphasizing equity and access for all." That article noted that the *Phoenix Gazette* had tagged her in May 1996 as the strongest candidate for governor in 1998, a designation that may have proved significant to the subsequent events.

Keegan's Agenda. As a political conservative, Keegan supported policies of less government, less regulation, efficiency, decentralization, and choice. Most observers agreed that her central mission would be the support of the charter school movement. These values she shared with Symington. They disagreed with each other, however, on the issue of financial equalization and response to the Supreme Court order in *Roosevelt v. Bishop*, 1994 that found substantial inequity in the way Arizona funds schools and ordered the state to rectify the inequities. Keegan believed that making the educational opportunities of students more nearly equal was a fundamental responsibility of the state, while Symington fought the order at every turn, launched a court challenge, and even declared that making districts equal was tantamount to "state socialism."

Where Keegan stood on assessment policy was less clear.⁷ Judging from comments made later in the Board of Education hearings on the standards, she may have made promises to conservative supporters during the election campaign to eliminate ASAP and reinstate standardized testing to a more

⁷ A district official stated, "Although I think she puts on a really good show. I don't know how really committed she is to performance-based education and assessment. But when you hear her talk, you'd swear to god that she was. I don't really know where she stands." In Keegan's own words, "I think that the demand for writing and demonstration of confidence is one of the strongest aspects of the original ASAP. I do believe that there is a little bit of a simplistic assumption: that if I write out my answer it is just de facto a better answer than if I say "yes" or "no." And of all sudden performance examinations or answer that are written out and explained became almost cultish. I mean it is not the case that there aren't any good computations that can be made and answered than in a multiple choice fashion. You have to have gone through a computation to get there. This sort of slavish devotion to watching everything get written out, I think has taken over in large aspect of what performance assessment is really about, performance assessment being the ability to think critically and to demonstrate that you've done that. It's not the same thing as is just writing out an answer. As a matter of fact, I would suggest that some of the problems that we've got in curriculum are around a very poor understanding of performance-based assessment. So that if a student writes out an answer, even if it's flawed in its thinking, if the answer has been written out, it's somehow defended, even if it's wrong. We score that, you know, just magnificently, because look at this, you can process. That oversteps, in my opinion, what was meant by the original devotion to performance assessment, which I happen to adhere to, I think, and that is we ought not to see doing only memorization, but using memorization as a tool in the critical thinking process, not just writing things out, as though the mere act of writing them out infers some sort of intelligence. That doesn't. I see horribly written material that was scored very, very highly simply because the student went through the exercise of defending a position that was really indefensible. And unfortunately, that's one of the reasons I feel like professional development is so key here, there is a difference. But how we maintained our sort of original principle which was that we're going to require not just a collection of facts, but utilization of those facts, which I think you can do through performance assessment. That's the strongest aspect of ASAP and a lot to recommend it. It also, I think, is a more difficult testing process than say an Iowa test, or something like that. It shows up diagnostically better. So there were all sort of reasons not to want to do away with it, not the least of which was we were making progress in the direction of a different kind of teaching and a higher accountability.

prominent place in state assessment policy. Yet, before teachers' organizations and the like, she professed her support for ASAP and always used the inclusive pronouns, "we" and "us." She aligned with her corporate supporters, however, in her professed belief that the schools were underachieving, bureaucratic, and falling short of producing graduates that could plug into jobs in the corporate world. According to the corporate view, the problem with public schools was a lack of accountability, and more testing and more control over the curriculum was the solution. The corporate constituency believed that a high school diploma certified "seat time" rather than academic proficiency. Keegan seems to have incorporated that view, even though control of curriculum and academic standards might be seen as inconsistent with the libertarian, anti-government ideology (though refer to Ball's postfordist analysis) (Ball, 1990).

Her first press release upon taking office at ADE in January 1995 stated that her vision was, "To promote academic achievement and ensure the responsible, effective, and efficient use of state moneys through services provided to all constituents." Among her goals were streamlining the system to allow schools to focus on academics (including reduction of mandates and paperwork burden); create options in the system (e.g., charter schools, parental choice grants, open enrollment); provide accurate and timely information and communication (on-line school report card and financial data bank), ensure effective funding (work with governor and legislators to equalize school finance); and "emphasize high-stakes academic accountability and testing" (critique and refine ASAP, continue norm-referenced testing, emphasize Essential Skills standards, adapt Goals 2000 to Arizona use, report cards, assure relevance of education to the workplace).

Very quickly, Keegan gained visibility in educational policy in the national arena as well as the state. On the national scene, she broke ranks with the Council of Chief State School Officers, the group that comprises most state superintendents. In an *Education Week* article (Lisa Graham Keegan and John Root, "Why We Formed the Education Leaders Council, 2/21/96, p. 39), she expressed her opposition to federal assessment policy (Goals 2000) and her support of local reform and parental choice:

[We] share the belief that education initiatives, policies, practices, and standards are strongest when generated from within individual communities and weakest when handed down from on high. We also believe that true education reforms are those that center on the needs and choices of families, empower parents and teachers to work in

concert to chart the course of a child's education, increase accountability in America's schools, and restore local control over school policies and practices. While all of that separates us from the education establishment, we believe it unites us with parents and the vast majority of American teachers and school administrators who share our exasperation with the nationalized business-as-usual approach to reform, and our fear that unless we act quickly and boldly to restore excellence to all schools our nation at risk will become a nation of ruin.... We have two constituencies: parents and children. We share their views on what they want and need.... As reported by Public Agenda, a public-opinion research organization in NYC, parents' priorities for education are: safety, discipline, high standards, and a focus on the basics ... more than half ... said that if they could afford to, they'd send their children to private schools because ... those priorities were being met [there].... We don't need politically correct education standards set at the national level. It's our job to set standards and then to hold schools, educators, and students accountable ... too many [children] are stuck in failing schools. We haven't lost our zeal to free them ...

The Reorganization of ADE. One month after taking office in January 1995, Keegan announced the reorganization of ADE, "to better focus our energies on our mission of improving academic achievement" (press release, 2/9/95). Brenda Henderson was named to head the Division of Student Achievement and Assessment, wherein the Essential Skills and ASAP Unit was placed. Staff who had previously been involved in ASAP were moved to Academic Support, Teacher Certification, and Teacher Development projects. Various informants used words such as "purge," "hit list," and "litmus test," to describe the changes in the department.

Results of the reorganization on assessment policy were immediate both symbolically and in reality. "When you call for information, either no one answers or someone answers who knows nothing," said an observer of ADE. Another noted:

The word, reorganization, sounds like an oxymoron. They came in and they decimated the department, is what they did. Anybody that wasn't on permanent status was history—gone. That was 90 something people, as I understood, overnight. Anybody with a doctorate—gone, or moved to a marginal position. And then because they had people with—that you had different departments that didn't communicate with one another, but the solution was to take people from one department and put them in another department, so they would take their experience with them and so on. So it was like a salad bowl; they just tossed everybody around, and they landed all different kinds of places and they did it in (clicks fingers) a very short order. And the result was chaos.

Another official in the Bishop administration who had already left ADE reflected later:

It wasn't personal, it was political. It says more about the lack of confidence in a professional educator—that someone who has a doctorate in curriculum and instruction, someone who has a doctorate in tests and measurements, someone who has a doctorate in fine arts, for example, is not really important, it's not necessary. That someone who is a layperson or a bureaucrat in a government agency or someone of that ilk can do the job better than a professional educator can.... [The people who were committed to using ASAP as a way of improving education] are now the ones who are doing menial tasks, shuffling papers somewhere ... no longer involved in that kind of work.

Technical Report: Text or Pretext?

On 1/25/95, the *Arizona Republic* (Hal Mattern, "State's public-school test suspended," p. A1, B-8) wrote:

Arizona's new standardized [sic] test for public-school students has been suspended for one year because of concerns about whether it accurately reflects what students are learning, the state's top education official said Friday ... she decided to suspend the annual Arizona Student Assessment Program test after questions were raised by the company that developed it. "The results we have so far have been called into question," Graham said. "I can't say with confidence that it's a valid test. It hasn't been verified enough to determine whether it correlates with how much kids know." ... Graham said the suspension of the test won't affect the curriculum portion of ASAP, which required teachers to change their methods of instruction. "Instituting that program has really made a difference in the classrooms," she said. Graham said that she and other state Department of Education officials will be studying the test over the next year with representatives of Chicago-based Riverside Publishing Co., which developed the test, in an effort to improve it. Suspension of the test didn't surprise some educators, some of whom have complained in the past about inconsistencies in the way the ASAP program has been implemented and the lack of training on administering the test. "We like the concept of the test, but more training and adjustments in the way districts use it are needed to make it as successful as it should be," said an AEA spokesman. A teacher in Avondale said, "I'm major-league disappointed," and said he likes the test because it "stresses writing and thinking skills, not the memorization of facts, rules, and formulas. I love the test. It's really worth teaching to." Graham said teachers ... needn't worry. She expects the ASAP test to be back in place in 1996. In a memo to school-district officials announcing the suspension, Graham described her action as an "affirmation of ASAP and nothing less. We are not abandoning the process," she said.

The report (Arizona Student Assessment Program Assessment Development Process/Technical Report, by Riverside Publishing Company) that Keegan referred to had been available since June, 1994, but had not been made public.⁸ An official later admitted that the technical data on the performance assessments had never been very promising, but that “ideologues,” that is, those who were actively promoting ASAP reform goals and performance assessment in general and those who suppressed or ignored information that cast ASAP in a bad light paid little attention.⁹¹⁰

Upon taking office, Keegan became aware of the technical report and the direction of its evidence. She sought counsel from representatives of the test publishers, staff members, and some university and district testing experts before making her decision about suspending ASAP. Other sources of information, such as an ADE teacher survey and opinions of staff and advisors, were said to influence her decision, but by her own account, the technical report was the principal stimulus:

The technical merit of the test, probably on a weight basis weighed more with me than anything else. You can work your way around a whole lot of veracity and work your way out of a whole lot of curricula problems or pedagogy problems or any of the things that I thought were evident in the skills themselves, because you could have fixed them, but you can't get around basically an invalid test.

After consulting with the State Board, she announced the decision, just two weeks before Form D-3 was due to be administered.

⁸ The cover page of the technical report that we obtained from a district testing coordinator lacks either an author or a date, and reads, “For use 2/2/95.” It is labeled “draft,” but according to department insiders, a final copy was never submitted.

⁹ A former member of the Bishop administration commented on the calculated neglect of the weak technical data: “I think it was a staff problem. I think the staff chose to ignore it because nobody wanted to go in and say to her, hey we’ve done this now once or twice, and we’re not getting the results back we ought to. It was so public at that point, that who’s going to walk into that office and say we’re using a test that has not been properly field-tested.”

¹⁰ As a Bishop administration insider reported, “And I said, ‘Wait a second, you know, it wasn’t all rosy from the beginning; by that I mean there were some problems with it, and we have this Form D that we’re just getting ready to administer again, and the first two didn’t really work out so well. We’ve got to ask ourselves, you know, Why are we doing it? Can we reform it?’ It’s not auditing.” And, of course, that was like dismissed, out of hand, because that’s the way the whole culture was. Research was ignored. That’s the culture in this building. Because it was Diane Bishop’s baby too.”

What Was in the Technical Report? The report began by placing Form D in its historical context and asserting its purpose:

Form D was conceived and designed as the Arizona state assessment to be administered each spring beginning in 1993 and continuing through the spring of 1996 to 3rd-, 8th-, and 12th-grade students. Form D is a statewide audit of student achievement on a subset of the Essential Skills. As an audit, the content and the specific skills addressed has to be secure. Form D was developed by addressing a selection of Essential Skills in reading, mathematics, and writing each year over a four-year period, for each of the grades 3, 8, and 12 with all the Essential Skills being measured over the four-year span.

Since it was neither practical nor necessary to measure every essential skill in Form D, a cluster of Essential Skills in a given Form A assessment was selected with emphasis on those requiring higher order thinking.... Thus, in the Form D assessments, students are required to use their knowledge and skills to think critically, to solve problem [sic] and to demonstrate writing skills as related to real life scenarios. (p. 7-8)

The Form D assessments were designed to be performance based and require students to engage in authentic tasks and use critical thinking skills to arrive at an opinion, conclusion, or summary; to construct a graph, chart, or table; or otherwise solve a problem. In addition to being authentic, the assessments had to be relevant, contemporary, and developmentally appropriate. (p. 10)

The report described the development of Form D, a process that apparently took less than three months.¹¹ Fairness and content validity checks were performed by convening “focus group discussions” (p. 11) and incorporating the comments of participants in revisions of the assessments. The content group evaluated the match between the assessments and the Essential Skills. Based on this review and “informal tryouts” to determine if instructions were clear and estimate time requirements, the report declared that Form D assessments were content valid and free of bias. A thorough reading of the report fails to reveal any empirical evaluation of this try-out to check for reliability or construct validity.

Scoring of the assessments was contracted to Measurement Incorporated whose personnel managed five scoring sites in Arizona or at their own offices, according to the report. A separate technical report is referred to that contained the statistical data from the test administration itself, that is, after the assessments had actually been administered rather than on pilot data.

¹¹ The contract from ADE to Riverside was awarded in October, 1992 with print-ready copy of assessments due in January, 1993.

The report explains the validation of ASAP Form D-1 as consisting of matching items with Essential Skills (content validity checks by the focus groups) and a construct validation study designed to equate Forms A and D-1 and estimate the correlations among the relevant portions of each. The equating study was done by requesting volunteer districts to contribute to the sample. Besides the volunteer Arizona schools, the publisher solicited participation from a New Mexico school because the Arizona samples proved too small. The samples in the study then took both forms of ASAP assessment. In the end, twelfth-grade data were insufficient to provide validity estimates. For the other grades, the resulting correlations between Form D-1 and the corresponding sections of Form A are represented in Table 1.

Table 1
Correlations Between ASAP Form A and D-1

Grade Three	Pearson Correlation	Corrected for Attenuation	% Common Variance
Reading	0.37	0.47	14%
Math	0.54	0.79	29%
Writing	0.41	0.67	17%
Grade Eight			
Reading	0.30	0.41	9%
Math	0.58	0.76	34%
Writing	0.34	0.45	12%
	0.51	0.75	26%

A supplement to the technical report was published in November, 1994 based on analysis of the Form D-2 administration. Another equating study was conducted in which students took D-2 as part of the regular assessment process and then were administered Form A. This time the correlations were slightly

higher, and the alpha reliabilities were also reported for both Form A and D-2. These data are represented in Table 2.

Table 2
Alpha Reliability and Correlations Between ASAP Forms A and D-2

Grade Three	Pearson Correlation	Corrected for Attenuation	% Common Variance	Alpha Reliabilities
Reading	0.45	0.59	21%	0.80 D
				0.76 A
Math	0.52	0.84	27%	0.69 D
	0.57	0.83	33%	0.56 A
Writing	0.53	0.73	28%	0.69 A
				0.81 D
Grade Eight				
Reading	0.61	0.80	37%	0.82 D
	0.45	0.54	20%	0.71 A
Math	0.42	0.53	18%	0.84 A
	0.43	0.58	19%	0.86 D
Writing	0.41	0.52	17%	0.73 A
				0.66 A
Grade Twelve				
Reading	0.69	0.95	47%	0.77 D
	0.31	0.42	10%	0.71 A
Math	0.52	0.71	27%	0.76 A
				0.69 D
Writing	0.37	0.50	14%	0.79 A
				0.87 D
				0.65 A

The report cautioned about interpreting the results of the equating study because of the small sample sizes in certain categories and because the level of difficulty of Forms A and D were different in some cases. In addition, the twelfth

grade sample produced a substantial number of scores of 0 on the writing test, calling into question the accuracy of the data. Having offered these cautions, the report asserted the following:

The validity studies performed for Form D2 provide the documentation to demonstrate that appropriate procedures were employed in the test construction process and that the statistical correlations are at a level to demonstrate that the assessments measure the Essential Skills being tested (p. 8)... The amount of common variance between the assessments was greater than 10 percent for all assessments except in grade eight reading. Although these correlations would signify that a relationship between the two assessments exists, the amount of common variance is small compared with that usually found in studies like this involving assessments in reading, mathematics, and writing. For example, in the Iowa Test of Basic Skills (ITBS) correlations between Forms K and L (two parallel versions of the ITBS) are usually in the 70s or 80s.... [In this study] the evidence is less than compelling. (p. 12)

About the reliabilities, the report stated:

This range in reliabilities is typical of the magnitude generally found in assessments eliciting student-constructed responses, however it is lower than the reliabilities of most multiple-choice assessments used in large-scale testing programs which generally range from 0.85 to 0.95.... For accurate reporting of individual student scores, reliabilities greater than 0.85 are generally expected. However, if the intent is to examine school level results, reliabilities of 0.55 and higher can be satisfactory.... [I]ndividual student scores reported from most of these assessments contain a large amount of error and should be used with caution.

Reflections on the Report. What can one make out from this technical report? First, note that the psychometric analysis occurred after the test was administered rather than before. If there were technical weaknesses, they were already implanted in the assessment results. Furthermore, the psychometric properties of scoring were not addressed. The four-point rubric that the state had adopted was extremely general, not even differentiated by grade level. Second, the equating study was a best-case scenario of the relationships between Forms A and D. Form A as implemented in schools was administered and scored by teachers and thus would likely be less standardized than the administration and scoring undertaken the equating studies. Districts had a choice of using Forms A, a portfolio system, or other local tests as long as they could be scored by the official rubric used on A. Thus the function of D as an audit of A *in practice* was not evaluated by this study, although the data are useful in other respects. Third,

the report spins the data a particular way by comparing them with reliability and validity data from standardized tests, an unrealistic burden for any performance assessment to bear. The authors give little guidance about comparable data from other large-scale assessments that have a history with extended response formats (e.g., National Assessment of Educational Progress, Advanced Placement) a perspective that might have cast the ASAP data in more favorable light. Interestingly, although Keegan focused on the inadequacy of Form D, D-2 had higher alpha reliabilities in some cases than the comparable Form A, which emerged unscathed from the department's analysis and was destined at that time to be the graduation competency test. Fourth, the report fails to mention that Form D tests content in integrated form while A tests reading, writing, and math separately, a fact that would likely depress their correlation. Finally, the analysis fails to account for the situation that each version of D (D-1, D-2, etc.) was designed to test one-fourth of the Essential Skills, so that the content domains of the two forms were different. Thus, depending on the question asked, the evidence about the technical qualities of the performance test is either more or less positive than how it was interpreted for the public. Of course, the actual report was quite closely held, and few have seen it directly.

How Was the Report Interpreted? Comments of insiders and observers subsequent to the decision represent an interesting array of information, misinformation, and alternative definitions of ideas held dear by psychometricians. We include some of the comments here to demonstrate the tenuous hold of these ideas by policy actors.

Riverside told us that Form D will not correlate with your Form A. And so you can give it to them, the kids, but you're not going to get any valid information. There isn't any validity there. Form A, B, and C were district level assessments, practice. Form D came in to say, "Did you report that your kids knew this? Did you tell us the truth?" is basically what it meant. I'm only telling you what I know about it. And to be very honest with you, I never went back and asked anyone. This is hearsay. This is not—it's what I understand to be true.

Well, the problem was the format itself. The company who had developed ASAP, which is out of California, Riverside, in the testing, it appeared that in the Form D that we weren't getting an accurate reading of the overall assessment of ASAP. And it made it impossible to absolutely certify the results. So if we couldn't do that, then basically the overall aspect means that it was worthless from a standpoint of being able to say, "Here's what's this data is doing compared to another state." It did not

mean that the individual parts—A, etc.—were not good tests. It's just that we couldn't absolutely guarantee it.

Well, I had felt that if we were going to base graduation requirements on Form A's, which the districts were kind of reporting themselves, but then we had a statewide audit on Form D's. I felt from the very beginning that as soon as those—If or as soon as those results started to ever differ, if the districts are saying, "Yeah, 90 percent of our students are competent on all Form A Essential Skills," then we give a Form D once a year, and it's 50, I felt that something would begin to unravel. I've never been real clear—I mean I'm not an educational professional, I guess, so I'm not sure that I have a complete understanding of what about the Form D's that was declared invalid or unreliable. But once the results of those too started to be different, something was going to unravel.

The teacher comments were particularly poignant in that the teacher comments suggested that the test couldn't possibly be valid because those districts that had adopted to a large degree the ASAP curriculum, the Form A's and the State promoted Essential Skills tests did not feel that the Form D, the examination followed, covered the same material.... The validity reports, the technical reports that came from Riverside itself said the same thing. And they started in the first year, suggesting that there were concerns about validity, how well did the Form D actually track the Form A. In the second year, there was no question, but they've got in their documents there really was a worse than chance equation. The contracts that had been drawn suggested that the Form D would equate with the A.

The field test reports show that these are not valid matches of the A's that they're supposed to be auditing.... At the same time, though, there were enough districts ... [saying,] hey, there's something wrong with this. Our kids completed all the A's, got them right, and then they came in and did the D's on the same skills, and their performance levels were two different? It can't happen. If you can do it one time you ought to be able to do it the next.

Among those insiders in the Keegan administration and her allies on the State Board of Education, the evidence was considered damning and the Form D incapable of resuscitation.¹² Asked about whether she had ever considered an effort to improve the assessment rather than kill it, Keegan replied:

¹² Accepting the premise that the evidence was negative, it was nevertheless less than fatal. For example, problems with poor directions or ambiguous content could have been discovered in a field trial and fixed. A rigorous and independent evaluation by groups with expertise in performance assessment could have identified problems with test content too easy or difficult for each age level, or could have suggested ways to simplify local testing practices. Reliabilities could be increased in a number of ways—by sampling items from much more specifically defined domains of tasks (and

I don't have that kind of patience. I mean I can't fathom my representing the state exam as a valid measurement of the Essential Skills which were mandatory—are mandatory—when I knew for a fact that the test was not a representation of ability in that area. It's dishonest. So I mean, no amount of time gets you over dishonesty. I don't know how that works itself out. That is not—it was not represented to be a feature of the test, [that it] would resolve itself in time, even by the testing company or the measurement company who scored this test.

The touchstone of opinions about the technical adequacy of Form D centered on its use as an audit of district assessments and whether it could be used to certify individual student competence. ADE insiders believed that Form D, “wasn't providing honest accountability. And what you needed was a high-stakes examination that was true.” A related worry concerned the lawsuits that might follow the use of a test with low reliability to deny high school diplomas. Because of the confluence of accountability functions with weak technical evidence, the Department treated the decision to suspend as a psychometric inevitability. As Keegan rhetorically asked, “What else could I do?” But others believed the data became a pretext for political action. An observer of the department reflected later:

I don't know. But I think when she made the decision she was faced with a decision that she had to make. Knowing what she knew, to go ahead and do D-3 becomes her problem all of a sudden. It was somebody else's problem, I mean, let's face it. If she did the D-3 knowing that its prior two forms were neither valid enough nor reliable enough, she then took the problem on for herself. So I don't think she had any other choice she could have made.

Not everyone, however, thought of the reliability and validity evidence in quite so negative light. A district testing official characterized the equating study of A and D as comparing “apples to oranges. It's a wonder they correlated at all.” A Riverside representative opined that, “given the nature of the materials, it was probably about average, okay.” And that performance assessments are “notoriously unreliable.” He explained that high reliability and validity coefficients rely on a much longer development process than that which characterized ASAP:

increasing the size of the sample), by providing pupils with many practice exercises similar to the tasks that were measured, by using more specific rubrics to score the results, by training and increasing the number of raters for each test, calibrating the performance of rating teams and rating sites, and the like (personal communication, Lorrie A. Shepard, 1996).

One of the problems with so many of these State mandated programs that, you know, somebody comes out here with an RFP, and where the impetus for this is coming from either the Legislature or the Governor, policymakers have this just (clicks fingers) I mean completely unrealistic idea about the difficulties in building tests in terms of time and money.... And so inevitably these mandated programs where they're mandated, they always have too short of a startup time associated with them. Which means there's no question but what the materials suffer in quality. You just can't do things that fast. And so when I say given, what I think of the comparison of these materials to other similar kinds of materials, they're fine. They're—but they're not near as good as they would be—I mean we would develop ITBS, I mean we're talking about years.... And reliability and validity? They're just words to policymakers.

An official in the Bishop administration thought that the technical qualities were adequate even in light of the methodology of the study, which she considered suspect and expressed the widely shared belief that the technical report was a pretext for a political decision. Referring to the report, she stated:

Well, it gave people a place to stand if they didn't like ASAP. But I think nothing was so severe that would require completely scrapping the examination. I mean it could have been corrected; if there was a technical flaw, it could have been corrected. Even though the correlations were low, it still didn't answer the question about how well is it measuring the Essential Skills. I'm not privy to the discussions that went on. And like I said, I said I wasn't going to speculate, but I think I did earlier when I said I think it was a political decision more than anything, and the technical report provided some place to stand. But I wouldn't say that the action was taken to suspend was really warranted based on that technical report. I mean you could have continued to administer and fix the problem. And they, you know, just kept on with the test, could always improve, but you didn't have to scrap.

A democratic legislator reported:

I do believe that it was politically motivated. I don't believe for one second that it was this great revelation to [Keegan] that this testing didn't jive, mainly because everyone in the educational field knew that Form D was not compatible or to be compared to the Form A tests.... So this trumped-up, great revelation that this is all out of whack and we have to put a moratorium on testing and I don't know what we're going to do and we have to re-tool the instrument, I think it was done with a lot of dramatic flair. And also I think it was politically motivated because there was a certain perception in political circles that it was a slam to Superintendent (Keegan) that the Governor turned around and hired (Bishop) as his education advisor. I just thought it was a reckless move. It was not done with a lot of consideration, or forethought, or ... I don't mind people coming in and taking radical changes or doing radical things as long as

they anticipate the outcomes more and make contingencies for it, but to come in and do what I would perceive as a reckless act, I can't support.

A former official in the Bishop administration connected the action to suspend ASAP with the transformed function that the assessment had assumed over the past two years:

I don't think the reliability and validity would be a problem, I don't think those are problems, if those assessments were put to the use for which they were designed. It is when you try to use them for something for which they were not designed that those things become a problem. Which makes me wonder about two things: one, the politics of it. As a new chief, she has to make her mark, and second, I wonder about the finance, the money. Those are the two things that come to my mind as you talk, that the reliability and validity are very good words to use when you want to take an action as she did to end the test. Those are both valid words to use, they're valid for her to use, even. Because she's thinking of a different use for the test. So it's perfectly legitimate for her to say I'm concerned about these tests because they're not—I would question the word "valid," I think valid they probably still are, to the old Essential Skills. "Reliable" I think would be the more appropriate term to use at this point because the old ones haven't been completely revised, they were valid to the old Essential Skills. But, reliability only becomes a problem if you're using a test for purposes other than which that test was designed.

To advise her on what action to take after the suspension, Keegan assembled an ad hoc committee that met three times during February and March. A member of the committee, which consisted primarily of district test coordinators and ADE staff, reported later:

The committee recommended that D be fixed. We kept telling her, keep working on it, don't get rid of it because you are going to lose your credibility and *of course* your validity is not going to be too high, because you're comparing apples to oranges. And if Form D tests the Essential Skills and Form A tests the Essential Skills, that's what you're trying to do. So what if they don't correlate with each other? But she had this bee in her bonnet that the validity was not high enough, and I think it was all just rhetoric.... She had promised to get rid of it (ASAP) during her campaign, and, lo and behold, Form D is gone. I don't believe that she pulled it because it was not valid. It was her way of making a statement right away that she was going to be a strong superintendent. And it was an unpopular test anyway, so what better way to get the teachers and parents behind her, to pull an unpopular test. It was unpopular because of the way it was implemented, not because of the underlying idea. I agree that they should have fixed it rather than starting over from square one. No one really wanted to get rid of it. That was mainly because of loss of credibility and the fact that we would

be several years without any state test or state data. But she just disregarded our recommendations and disbanded the committee. She convened that committee but wasn't really interested in what the committee had to say. It was all just show.

The Demise of ASAP

Keegan had promised in January not to abandon "the ASAP process," but merely to critique and refine it. But abandon it she did, if by "the process" she meant large-scale performance assessment as the touchstone, or "audit" of the ASAP program as a whole or the reform of schools toward constructivism. Between January and May, she decided that the problems of Form D were too serious to remedy, and signaled more far-reaching problems in state assessment policy. The Arizona Essential Skills needed an overhaul as well. On a local television program, Keegan foreshadowed what was to come when she offered a jaundiced view of the existing Essential Skills, which she derisively recommended sending to the scrap heap. Most of them were not measurable, and the documents were so long, convoluted, and filled with educational jargon that parents could not possibly understand them or hold schools accountable for achieving them. In addition, the Essential Skills failed to embody world class standards, and emphasized process rather than outcomes, according to her view.

An ADE memo of 3/8/95 to schools stated that "a team of specialists from the Student Achievement and Assessment and the School to Work Divisions has been formed. This team is charged with ensuring that the Essential Skills have the following characteristics:" Encourage high-level achievement, emphasize academic content, be precisely defined and measurable, incorporate the SCANS skills considered necessary by business and industry, reflect a real-world, occupational context, have a consistent, easy-to-read format." The superintendent also assigned to the team the task of ensuring that the state assessment system incorporate both standardized and performance-based tests and incorporate high school graduation tests. Two months later, May 25, 1995, ADE announced a revision of the program as a whole, and a new name with the old acronym:

... [Keegan] announced the new direction for the Arizona Student *Achievement* Program [italics added] which integrates career education and academic proficiency for all students. "What we expect of our students is what we will get," she said. "Our expectation must be for both high academic achievement and lifetime employment.

The ASAP has undergone thorough review since the Form D statewide ASAP was suspended earlier this year. "While I was unhappy to find that our previous testing program was a problem, the discussion of the past few months has resulted in very strong revisions.... What we have heard from parents, teachers, and business about ASAP has led us to keep the foundation and vision of this excellent concept, but to be far more demanding and to ensure that the program is relevant to our students' education." The major components of the revised program include:

- new statewide and district level assessments;
- professional development for teachers;
- changing to 4th, 8th, and 10th the grades in which the statewide assessments of academic proficiency are given;
- introduction of a certificate of mastery of academic proficiency in 10th grade;
- introduction of a 12th-grade workplace-specific or higher education placement test.

Our primary emphasis at the elementary level will be mastery of foundational skills.... Where we find students not proficient in those Essential Skills by grade four, we must offer solutions immediately to increase the likelihood of that child's successful completion of his or her education. Waiting on a 12th-grade graduation test of proficiency is eight years too late.

Under Graham's proposal, the focus on academic proficiency will continue in middle school, but with career counseling and an eye toward the work force. Upon demonstrating the mastery of the Essential Skills in 10th grade, students will earn a certificate of mastery of academic proficiency. The balance of a student's high school course work will be tailored to meet his or her educational and career goals.

A final test of proficiency in the 12th grade, congruent with students' course work and career goals, will be required. These tests may be either workplace-specific tests or college placement tests. The certificate of mastery and successful completion of the appropriate tests will be a graduation requirement for the class of 2002.

The difference we've created by adding the 10th-grade academic proficiency assessment and appropriate 12th grade tests is that it ensures that 100 percent of

students will have exposure to workplace experience, and that all will be expected to master the same academic skills.

We remain committed to high-stakes graduation requirements for our students.

The release noted that norm-referenced testing would be continued, that no state ASAP testing would be conducted during the 1995-96 academic year, and that the new ASAP state test would be piloted during 1996-97. Meanwhile, districts would still be required to continue their state mandated testing and reporting according to the DAPs. Subsequent to the press release, the Board of Education approved the action to reform the assessment policy. The vote was not unanimous, as some members worried that educators would construe the action as affecting the ASAP program as a whole and that state testing would lose credibility.

Reactions to this announcement varied across the spectrum, as Keegan herself noted in an interview.

Well, they fell into two categories. Either people were pleased about [the decision] because they didn't like the test, and I guess within that category you had people who didn't like the test because they weren't on board, period, didn't want tests, didn't want to deal with that. And then the other side are those people who really had held out some hope that this would be something that would be terrific, that were on board with the Essential Skills and performance-based assessment, and they just flat out thought that the examination was not reflective of what their kids knew, and they felt like they were being misrepresented in their teaching. Now, that was the majority of the letters that I saw, not "We don't want to be tested," but, "This test isn't reflective of my students' ability," you know, "This is a horrible experience, and the kids don't respond well to it," which, you know, my own experience in looking at that exam is pretty predictable. So there was that, "Gee, thanks, we didn't like it either, it does need to be fixed," and then there was, "Thanks for getting rid of it, we don't ever want to see a test again." And then probably a third which was, "We don't like performance assessment, we're fine with yes-no, true-false, traditional what-we-understand-type testing." So those folks, all of whom were happy to have it suspended. The other side was those who were not happy to have it suspended, and primarily those were curriculum directors, sort of people who had been involved in putting it together, in pushing it, the administration superintendents who were on board with the ASAP program. And there was a whole lot of diatribe and rumor about the fact that, you know, my election was seen by a lot of people as sort of a right wing thing and "See, this is a back to basics move, and we're going to start testing with some Neanderthal examination."

Republican legislators and the State Board members played down the reaction or failed to see it at all. “Zero reaction,” said a legislator. Insiders in the Bishop administration, however, reported dissatisfaction and even protest from groups such as the Parent Teachers Association, teacher groups, and administrators in those districts that had spent time and energy adapting their programs toward constructivist education. Said one:

People were very disappointed, very disappointed. And they didn’t understand why it was stopping. People in school districts said, “geez, we’ve really invested a lot of time and effort getting the kids ready.” Everything that we had hoped for that districts and schools would do to prepare their kids to do well on these assessments was happening. Everything was happening. There were places obviously it wasn’t, but just stopping was a tremendous let-down for people. They said, “oh, geez, we’ve really worked hard.” So, I thought at that point in January of ‘95 that the program really had come a long way. Had really made a huge impact. And people here, I went in and talked to the principals about not giving the D-3, which was going to be within two weeks or something—three weeks, I don’t know what the timing was, March maybe—and they were tremendously disappointed because they had invested in the program, they began to understand the power of this. Our principals here were saying our teachers are teaching differently, they’re thinking differently about achievement, they’re using resources and their time—everything that late in ‘88 we thought we could do was being done. And so they were very disappointed.

Members of the Keegan administration were befuddled by the negative reaction in some quarters, because they claimed that this was not a wholesale change in assessment policy but only a “correction” or “clarification,” and that it only affected parts of ASAP. What they misunderstood was that, to many educators, Form D *was* ASAP—an integrated assessment that both mirrored and promoted integrated instruction (Smith, 1996). The words of the superintendent seemed to indicate a return to competency-based teaching, skills-oriented testing, and a highly tracked, test-driven school structure.

An insider in Keegan’s administration noted that her original plan had involved taking five years to revise the Essential Skills along the lines of her specifications, consulting with teachers, parents, business, experts, and the like, and carefully developing appropriate state testing. But the Board of Education rejected that plan as too slow. Board members wanted an immediate revision and no gaps in accountability data. The majority agenda on the Board was to institute, as soon as possible, a graduation competency test as a substitute for

“seat time.”¹³ So Keegan put aside her plan for a patient, collaborative process of standard-setting in favor of staging the Academic Summit, in which the Skills would be replaced with new standards in minimal time. According to a Board member,

Well, I think there had been some prior belief among board members that the Essential Skills were too voluminous, and that when the time was appropriate, it would be a good idea to streamline them, to perhaps review them, to make them higher standards. And probably mid-year, we asked the administration how quickly could we develop standards so that then assessments could be developed, so that then we could develop an appropriate testing mechanism, keeping in mind that we wanted this to happen as soon as possible, which is A-S-A-P. And we're told that they felt that they could accomplish that with our support by January of '96.

Some educators wrongly believed that the actions of the Board and Superintendent contravened state law. In fact, the legislation had never mandated or even mentioned “ASAP,” but prescribed norm-referenced testing in three grades and “Essential Skills testing,” which the districts carried on during the transition. However, the proposed change of grades for state Essential Skills testing (i.e., from grades 3, 8, and 12 to grades 4, 8, and 10, plus a graduation test) required that the legislation be changed. The necessary change in legislation again made the state legislature a powerful player in assessment policy. By this time, the Legislature had moved far to the right from where it had been at the time of the last legislative change in assessment policy, as one can see in a subsequent section.

The Academic Summit: Blitzkrieg Standard-Setting

In a 8/29/95 letter to superintendents and principals, Keegan announced that the Board of Education, having worked with the ADE staff over a period of some months, had formulated a plan for “an accelerated delivery of state standards and assessments,” including the development of state standards in nine academic areas: language arts, math, workplace skills, science, social studies, arts, foreign language, comprehensive health, and technology.

¹³ A State Board member said this: “I have always been concerned that seat time should not be a graduation requirement. Ever since I was in high school, seat time was all you really needed to get a diploma. And we all agreed that we wanted a diploma to mean something, to have some stakes to it, some risks to it, perhaps even get to the point ultimately where there could be a guarantee to the business community that if our students have a diploma that they can count on them having certain skills.”

This announcement was followed by a Press Release from ADE on 9/28/95:

[Keegan] has scheduled an Academic Summit to develop state standards for student achievement in nine key academic areas. "We need to establish standards that will tell us what our students in grades K-12 need to know and what they must be able to do as they progress through their education.... It's absolutely essential that the standards we develop provide a high academic challenge for our students, are uniform in presentation and easily understood, can be assessed, and are relevant to the workplace. I'm hopeful that we can develop standards, that, in just two to three pages, explain what we expect ... and that parents, as well as educators, can easily understand them." ... Standards for each content areas measure student achievement through five basic levels: readiness, foundations, essentials, proficiency and distinction. These levels cover students' advancement from kindergarten through 12th grade.... We've invited a broad cross-section of education professionals, students, parents, and constituents to help in this critical step.

In a letter to districts from ADE on 9/28/95, the Department announced that the standards would be developed by the summit teams, submitted for approval of State Board and that the only areas of state testing would be reading, writing, mathematics, and workplace skills. All other areas of standards would have to be tested by districts on measures that they would develop locally and submit to ADE for approval:

Training on the state standards and the new assessment process will be available for teachers, parents and students.... Local schools will work on developing curriculum packages that will include local assessments.... A competitive process will be used to select up to three model curriculum packages in each context area that will be offered as models for other schools or districts.

Plans and Realities of Conducting the Summit. The Academic Summit took place in a Scottsdale resort, October 16-18, 1995. The nine Design Teams, one team for each of the nine content areas for which standards were to be developed, had had two prior meetings to become acquainted with each other and with the task before them. The schedule would be tight, but the rhetoric of the summit planners made it clear they thought the task was feasible. They felt spending three years writing standards, as some states had done, was absurd. The summit planners believed it would be possible to form a team, write a three-page list of clear, measurable standards, present them to the other teams, get reactions during public hearings in December, and write a final draft to present to the State Board in time for its January, 1996 meeting. The Board would then approve the

standards and issue a request for proposals to test publishers. The winning bidder would then construct pilot assessments to be administered in March of 1996. That was the plan.

Even as early as the opening day of the Summit, it was apparent that the schedule could not be realized. The single most prevalent complaint about the standard-setting process was the rush, and few could understand why that rush was necessary. A former official in the Bishop administration said that the schedule was crazy even by their standards. A district administrator noted the naiveté of ADE and the summit planners about curriculum development. Over the three-day Summit, he saw their dawning realization about the complexities of the task. On the language arts Design Team, 20 planned hours became, by some estimates, approximately 200 hours of meeting time, spread over nearly a calendar year, plus countless hours spent in reading, writing, and reflection. The language arts standards would not be accepted by the Board until the summer of 1996, and even then, only parts of the standards submitted by the Design Teams (reading and writing) were approved. Other standards finally made it through in mid-1997. That was the reality.

By January, 1996, ADE had conceded that its schedule would not hold up. The ambitious plan to complete drafts of the standards during that three-day session proved impossible, and the Design Teams continued to meet thereafter. The drafts were passed back and forth from teams to ADE. In a letter from Keegan to schools (1/18/96), she announced that the drafts would not be submitted as planned to the Board for its January meeting and that the spring pilot assessment would therefore not be conducted.

Agenda for the Summit. Although officially managed by the ADE Deputy Superintendent, the summit organization bore the marks of two groups: a set of consultants from Doyle Associates and trainers and facilitators from Keegan's corporate partners. The influence of Dennis Doyle was apparent in Keegan's letters to schools prior to the summit and in her introductory remarks at the Summit itself. In those remarks she expressed the perspective that the new Standards implied a restructuring of the entire system. Thus, the standards were to be written in terms of "Levels," rather than grades. Grades would no longer be the means by which pupils progress through the system, but would be replaced by competency testing. She endorsed Doyle's argument that schools should hold

achievement constant and vary time, rather than the other way around, as we do now.

The influence of the corporate partners (e.g., Motorola and Allied Signal) was felt in choice of facilitators, in language and concepts more appropriate to the corporate world than education (“Design Teams,” clear and measurable standards, market incentives, performance equated with product, and the like), and in the prominent place of workplace skills and technology. The corporate community favored the idea of replacing “seat time” with competency tests as the criterion for high school graduation. In addition, both ADE administrators shared with the corporate partners the idea that since half of Arizona high school graduates do not go to college, that they should be given work place skills training. Keegan’s memo of 5/25/95 even hinted at an extensive tracking system wherein students would be directed by junior high into either a college program or a vocational program and receive their subsequent teaching and testing accordingly.

Participants in Standard-Setting. Each of the nine Design Teams comprised nine members plus one or more facilitators. The participants included parents, teachers, students, and laypersons who had been appointed by ADE from a list of self-nominations. By looking at the list of participants, it was difficult to figure out how *these* persons found themselves attached to *these* particular teams. That is, a math teacher was just as likely to turn up on the social studies team as on the math team. And curriculum specialists were conspicuously absent. Some of this seemed to be intentional and some the result of fortuitous circumstance. An official with ADE explained it this way: that when Keegan was on a fact-finding trip subsequent to her election, she had mentioned the possibility of convening committees to rewrite the Essential Skills. People started sending their resumes to the Department, which kept a list and invited some of these people to be on the Teams. Later, when the Summit was announced more generally and officially, people (i.e., curriculum specialists) who asked to participate were told that the teams were already full. But the Department also made clear that loading the teams with nonspecialists would have the effect of reducing educational jargon and making the standards clear and measurable. During a Board meeting, one member said about curriculum specialists, “We don’t *want* to know what they know. We deliberately cut them out of the process.” Another Board member said in an interview:

We had teachers involved, but we did not have curriculum coordinators involved, nor did we have the Department of Education employees involved. We felt those latter two, while many could contribute, just as many, if not all, would have a stake in maintaining the status quo, or in making their job easier. And that was not our intent. We don't care really whether their job is easier. We want it to be just as challenging for them as it is for everybody else.

A Department observer commented on the composition of the teams this way:

It's just another slap in the face of the current system, another way to say that, "We distrust what you're doing. We distrust your knowledge. You haven't done a good job."

Whether intentional or accidental, ADE did not pursue a strategy of representation¹⁴ on the teams, except to include laypersons on each. Among the missing as a result were not only curriculum and assessment specialists, but members of ethnic groups as well. A glance over the Summit opening day sessions revealed overwhelmingly white faces.

Organization of the Summit. The Superintendent began the Summit at an open meeting attended by the team members, a variety of policy actors, people who wanted to make their comments part of the official record, representatives of test publishing companies, media, and observers—more than 200 people. After welcoming and orienting remarks were made, people from the audience made public comments; for example, about the tight time frame, the poor representation on the teams, about rumored change in direction away from constructivism and toward basic skills teaching. Few spoke, the character of their remarks was monologic, and no one on the dais bothered to respond to their concerns.

Before the teams deployed to their separate rooms, one of the facilitators sternly admonished the observers:

We have high expectations that you will observe only and *not participate*. You will allow the teams to do their work. You will not have access to members of the team while they are working. If you have comments, please reserve them for the public comment times. You will not have access to the working papers but you can sign up to receive the draft standards when they are ready.

¹⁴ Later, during the review and revision process, ADE appointed three subcommittees (content, technical, and "special population" issues) that were selected to include teachers, content specialists, and minority representatives (see next section).

During meetings of the teams, the oversight of the facilitators, whose time the corporate partners had contributed and who came in from as far away as the East coast, was evident. Much of the teams' work was prescribed, for example, the format of the eventual standards, the levels at which they were to be written, and the demand for clarity and measurability. One participant referred to the management as "heavy-handed." The press of time made it necessary to work fast. Later, the language arts team would rebel, ask the facilitator to take a back seat so the team members could do their work. In spite of the tight grip on group process, however, there was a good deal of initial struggling to grasp the task requirements and the expectations of the Summit planners. Because the team members were diverse, most lacked experience with curriculum development, there was a period of negotiation about, for example, the difference between a curriculum and a curriculum framework and which of these they were supposed to be considering. The press of time resulted in many glosses and a failure, according to one of the participants, to consider fundamental philosophic assumptions and biases.¹⁵

In addition, we saw a tension over the role that the existing Essential Skills should play in the proposed new standards. Participants were quite aware that their task involved supplanting the Skills. After all, if the Skills were adequate, there would have been no need for the current exercise. Echoed by the facilitators, the Superintendent had spoken derisively about the Essential Skills and reminded the teams that many were not measurable, that they were too process-oriented, too jargon-laden, and not sufficiently ambitious. The Summit planners provided participants with a notebook composed of the existing Essential Skills, some of the national standards, and the standards adopted by the state of Colorado and Beaufort County, South Carolina Schools. It was to the latter, however, that the planners continually referred as the models these teams should emulate. Yet the Department came to take on a different tone as time passed and pressed. Feedback from the observers reminded ADE about the extent of professional work that had gone into the Essential Skills. Perhaps they realized that a revision of the Skills was the most that one could expect in such a tight

¹⁵ In a discussion of the language arts team, the comment was made that some of the issues about standards could not be reconciled unless they saw the form of the assessment and whether second language learners would be tested in their dominant language or in English. The facilitator reminded the team that assessment was not its problem, that someone else would take care of assessment, and the team reluctantly proceeded on that assumption. There was never (until it was too late, according to one informant) an open discussion of educational philosophy.

time frame. Later, ADE staff would deny that they ever intended to abandon the Essential Skills, a claim that does not stand up to the evidence of observation and documents.

In the language arts team, at least, other tensions surfaced. Although the summit directives attempted to guide participants away from constructivism toward the simple, clear, and measurable, resistance was evident. Two participants introduced principles from the standards of the National Council of Teachers of English, which are constructivist in philosophy and filled with just the kind of jargon that the Summit Planners had warned against. Again resisting directives, participants repeatedly tried to deal with issues of equity and quality education for second-language learners. Despite their pleas, the Superintendent resolved that, "Without question, the standards you are creating are for proficiency in the English language. Assessment of the standard will be in the English language" (1/25/95).¹⁶

According to an informant, even considering the diversity of backgrounds that characterized the language arts team, the members came to common understandings of the issues and their own agenda. They took their "work seriously and produced draft standards with integrity, focus, and balance" (though this would not be realized until several months after the Summit). They incorporated constructivist principles in several ways, one of which was (defying orders) deliberately writing standards that could not easily be tested with multiple choice tests. In addition, they attempted to inject issues of global literature, cultural comparisons, reading for pleasure, and self-assessment of student as writer, "but we knew from the beginning that that would be a tough battle." The team believed that it had incorporated constructivism by designating four varieties of standards within language arts. The team came to see language arts "as a whole," that is, embracing listening/speaking and visual representation as well as reading and writing. Many of the constructivist ideas such as content integration, projects, thematics and problem-solving were expressed in the connections among these elements.

¹⁶ Keegan's memo continued, "I know you are all well aware of the frighteningly low rate of reading proficiency as measured in the elementary years. And you know well that writing skills are seen to be in decline, especially by potential employers. Should the standards document you are completing suggest anything short of the highest degree of rigor and proficiency in the English language, I believe the public will question the sincerity of our effort."

Many of the intentions of the team in this respect were overturned later. “It was the classic dichotomy between authority and responsibility. We were responsible for generating good standards. But we didn’t have the authority to adopt them.”

Reactions and Revisions of Standards

Who did have the authority and what were the political dynamics of standard-setting? The political nature of assessment policy revealed itself in a set of public hearings, revisions to the standards, and the actions of the board to consider and approve them over the next calendar year.

During the third week of December, after the nine draft standards documents had been distributed, 11 public hearings were conducted on the Standards documents—two in Maricopa County, two in Tucson, and one each in Flagstaff, Prescott, Nogales, Show Low, Kingman, Yuma, and Chinle. ADE staff ran the meetings, some of which were attended by the Superintendent and Board of Education members as well as the Design Team participants.

Vignette 3: Claques, Catcalls, and Stage Whispers¹⁷

It’s an odd mix of types here tonight, and they’re checking each other out. I’m hearing grumbling even before things get started. A teacher nearby asks rhetorically, “How do you expect us to be serious about this when we’ve had so little time to read these standards? And they only give us three minutes to comment, are they really interested or is this just a show of soliciting input?” No one answers, but then, there would be little give and take for the formal part of the meeting, either. Just one monologue after another. Some people do come to the podium from the floor, speaking into the ears of the people up there, which makes me wonder if some people get more input than others. Who knows? A large contingent comes in and takes seats together near the front. They seem to be from out of town. They have their heads together talking and either glare or nod vigorously depending on what each speaker says.

¹⁷ This vignette is a composite of three public hearings which were observed by the researchers. The comments are direct and paraphrased quotations of what we encountered. Our intent was to convey the truth of the interactions we observed and recorded through a compression of literal verbatim remarks.

The superintendent welcomes us and provides background on the standard-setting process. As the comments begin, I realize that there is a small contingent of University people here. They are members of a coalition of public school supporters, remnants of a group that ten years ago successfully lobbied to eliminate the mandated standardized testing of first graders. That seems like a long time ago. One testifies that, standards or the lack of them “are not the problem—funding is the problem.” She admonishes the state not to mandate testing of standards until all students have adequate opportunity to learn them. Otherwise, it will be the minority students who will suffer, who will “bear the brunt of the state standards.” Another one worries that no one has thought enough about second language issues. That group over there is really glaring and whispering at this.

A person I recognize as a member of the state Bilingual Consortium fears that the emphasis on English in state assessment may mean that some non-English speakers will never graduate. But a Hispanic parent rises next to say that assessment must be in English, otherwise it is patronizing to Hispanics.

There is a series of teachers who complain about the state’s intrusion into standard-setting. “What was so wrong with the Essential Skills,” one asks, and notes the years of teachers’ and curriculum specialists’ work that went into their development. Another comments that “this is a great step backward,” from the ASAP process. Several worry that the trend toward problem-solving teaching, and integrated subject matter will reverse itself if the state retreats to basic skills testing. One cites research on the futility of teaching word recognition, spelling, and grammar in isolation from authentic tasks.

Others dispute the benchmark years and the interpretation of the developmental levels of the standards. Some believe that certain standards are placed too high or too low in terms of age and grade levels. These sound to me like characteristic teacher concerns, thinking about what is taught, what should be taught, and what kids of a certain age are likely to learn.

Then there are one or two who dispute the standard-setting process itself. One says that there hasn’t been enough time, that other states required several years to accomplish the same task. A University person says that even ASAP was inadequate, having three years of development, so that the new test could hardly be any better with only three months to develop. One notes that the Design

teams were not representative, particularly of curriculum experts. One called the process arbitrary and “done for show.”

An art educator complains that there is no museum mentioned, and that there is too much “doing” and not enough art appreciation in the fine arts standards. But another one gushes over the very fact that fine arts were included at all. A professor protests the stereotypes and gender bias in the health standards. A teacher says that foreign language should be integrated with other academic areas. Another says that language arts standards lack a literature component.

Those all seem like insider stuff to me. You can make book that tomorrow’s editorial will use the word “educrats” to describe the people who raise these issues tonight. The people at the podium listen politely but take no notes. They make little attempt to respond to the issues. But things begin to pick up, as the people in the front of the room, who represent themselves as parents, obviously organized and ready, gear up to take their turns at the microphone. Several express their opposition to sex education in the health standards, “It’s the province of the family, not the school.” Another says that the dance standard in fine arts “requires too much dancing,” and this might be offensive to parents. In general, the fine arts standards come in for quite a bit of abuse. A parent opposes the foreign language standards, stating that foreign language is not needed in America. A parent from Mesa says the whole Social Studies standards should be replaced with the Bible. The person sitting next to me has written on her notes, “The nuts have really crawled out of the woodwork.”

Then the debate that I’ve been waiting for. A woman addresses the Superintendent directly and asks her to acknowledge that the new standards are the state’s expression of compliance with Goals 2000. Isn’t it true that Arizona had to participate in Goals 2000 in order to qualify for federal funds? And aren’t these standards really Outcome-Based education? She strongly opposes OBE, because it teaches values and self-esteem, rather than basic skills. (My neighbor nudges me, “Who let the Eagle Forum in here?”) Keegan denies that the state is tied to Goals 2000 and disavows Outcomes-Based Education (“I would never put my children in a school like that”), but the parents won’t let up. They demand to know why phonics aren’t mandated in the standards. “Phonics is the key to all learning,” a parent cries out. The Superintendent agrees, at least in part. She states there is too much process-orientation in the current Essential Skills and

not enough emphasis on correct spelling, grammar, and phonics—that it was the emphasis on invented spelling and whole language that got California in trouble. That group murmurs its assent. But they wait in vain for her to go all the way to demand for phonics in the standards. My tape recorder preserves her words: “I think the importance of standards are that they say, ‘This is what you must know. This is what students must demonstrate in terms of their proficiency.’ And then how you get there, we would leave that up to the schools, because I do think schools just have to have different approaches because students are different.” The parent presses on, and the Superintendent finally admits, “I said I was a fan of phonics.” The parent, finally satisfied, says, “That’s what I wanted to hear.” But a teacher calls out from the audience, “Don’t tell me how to teach.”

The last word is yet to come. A woman from the up front group stands up, very agitated and accuses the Superintendent of having reneged on her campaign promise to repeal ASAP. She thinks the standards are just more of the same. As I recall from the pre-election statements that Keegan had supported ASAP. Was the accuser wrong or were different pledges made to different audiences? I see a Design Team member shaking his head. He’s probably wondering how to accommodate this disparate “input” into the Standards. Some things just don’t blend. Me? I’m wondering about the state of democratic deliberation, or even of its possibility. There certainly hasn’t been any here, only a bunch of people, many organized in what looks like political action groups, making bald assertions of opinion, masquerading as fact or principle.

Following the public meetings, further reactions to the draft standards came in by mail and fax. No one could say, however, just how extensive was the distribution of the design drafts or how representative were the comments sent back. If state teacher organizations had an official response, it was not reported in the newspapers. A national representative of the American Federation of Teachers compared the drafts with the AFT standards for standards, and found them deficient and the process “absolutely flawed from the beginning.” Members of the design team were strongly encouraged to pay heed to the feedback, whatever its quantity or quality. A member of the Language Arts team claimed

that it was a “testament” to the public to take the trouble to provide feedback and a testament to the team that it took it seriously.

Revisions were made and passed back and forth to ADE and the Summit facilitators over several months. Two members of the State Board also participated in the review. They went through four drafts. The evolution of the drafts reflected the tensions already evident at the Summit: the Board and ADE emphasizing the simple, brief, measurable, and ambitious; and the Team leaning toward the complex, process-oriented, holistic, integrated, and influenced by the national standards. It was the textbook example of what micropolitical theory (Ball, 1987) calls the dialectic contending over what group gets to define the situation, in this case, what the definition of language arts would be. Between drafts three and four, the Superintendent wrote a long memo to the review team, recommending a variety of substantive changes. She recommended additional components to the standards, clarifications and elaborations, and rewording. In addition, she wrote in a memo, “We must develop a sample reading list for each of the five levels to give ... a sense of the quality and complexity of text students are expected to read and master ... we may want to add a requirement that students read a certain number of books per year (e.g., 20-30) from an identified number of writers and genres.” Angered, the team ignored most of the recommendations. One change she suggested did work its way into (or more properly, out of) the writing standards. Originally, one standard at the readiness level read, “Perceiving themselves as writers.” Her response, “important, but how do you measure?,” led to deletion of that standard and the substitution of “spells simple words,” and “writes the 26 letters of the alphabet.” The department’s insistence on measurability also resulted in the deletion of standards related to developing students as life-long readers as well as all the standards written for listening, speaking, and visual representation.

According to a participant, the process reached a turning point when the Board altered the membership of the Design Team to include Board members and certain “friends of the Board.” The latter group included patrons of the Franklin Traditional School, and the loading up on one perspective had its intended effect:

The subcommittee of the Board was supposed to work with the design team to try to come up with common ground of what we could live with. This is where you see a tightening of examples and leaving out the “fluffy stuff” and the multicultural stuff

and all that. And this is where you see all the concessions about how much emphasis to give to phonics and whether phonics gets its own bullet. There was this constant refining and rewriting and paring down. "Is this tight enough? Is this clear enough? Can this be tested? Is this a one-answer thing?" that kind of hammering pressure gradually shifted. It wasn't consciously decided about how much constructivism we could safely put in there. We knew who the players were and where the pressure points were. It was a subtle—and sometimes about as subtle as throwing a brick—but it was a constant movement in that direction, a constant struggle between those who had a commitment to the standards as first written and those who had the authority to turn them into approved items and then the test.... This is where the shift from process and constructivist theory to basic skills emphasis took place. We were very conscious about putting some of those broader, more integrated more holistic aspects of language development—those went into the other two areas. So you lost the vision of language arts for the future, which we worked very hard on incorporating. We had had a broader view. There are two other strands that we thought were important. We chose what to put in reading and writing based on the assumption that the other two parts [listening/speaking and visual representation] would also be there.

It was the Board that decided to separate out reading and writing from our document that integrated language arts. Because it was taking so long. You had a time table that said you have to do this, you have to do that—do a pilot, got to do the RFP, got to get this going and this is as much as we can manage. I don't think anyone sat down to say we only need reading and writing or we don't like what the team is doing, it was more sitting down and saying this is what we have, we have to get something out, let's go with this. The department and the board don't want to be criticized for giving a partial view of language arts. [When we complained] that language arts was violated, that it was truncated, people from the department were offended by this because they said, "*We know* that language arts is more than reading and writing. We never said we weren't going to do the rest." But in practice that's what's happening. But there's this whole other agenda about getting an RFP out and getting it piloted and so on. On top of that you have this press from parent groups and certain members of the Board, including the Board president to get it to look more like the governor wanted. And so you ended up with more phonics and more disaggregated skills.

The *idea* of having kids making assumptions! Even the idea of personal experience narrative. "You shouldn't be having kids write about personal, private things. You shouldn't have kids speculate. You shouldn't have kids theorizing." It's like, "when you've mastered your skills, and when you've been educated, and when you've got to that point when you've matured, *then* you can start considering possibilities and rationale, but you don't do that until you're finished being educated." It's that common thread. "We don't like kids making comparisons about literature, and what other literature in the world would we want them to look at anyway? What other culture

would we want them to study besides this one? Why would you put that in there?" [laughs ironically]. This is what comes through in the Board's discussion. And guess what the next step is? That's when they broke off reading and writing from listening and visual representation. It was all they could do to agree on just those two. The hard stuff, the up-dated, not-by-the book stuff, the not right-answer only stuff was in listening and visual representing standards. The further you went into the document the less control you had. So you think it's an accident that they didn't get to it?

The Legislators, Governor, Superintendent, Board President:

At Play in the Fields of Assessment Policy

During the months after the Summit when the team members, ADE staff, and Board president were occupied with exchanging drafts and moderating their positions, the governor and key Republican legislators were also working on their own versions of assessment policy. If department insiders viewed the legislature as "out of the loop," and "relatively neutral" or "without a position" on the suspension of ASAP and revision of Essential Skills testing,¹⁸ they were mistaken. The interplay of power and position among these actors (the alliance of Superintendent and Board President in 1995, the governor's appointments to the Board in 1995 and 1996, the behind the scenes collaboration between the governor and key legislators in 1996) had everything to do with the shape of assessment policy in 1997 and beyond.

By an informant's perspective, over the months of sending drafts of the language arts standards back and forth, Keegan learned, accommodated, evolved in her understanding of the complexities of curriculum building and reform. Moving more to the center, she toned down much of the rhetoric of the previous fall about work place skills, tracking, and test-based grade promotion. Meanwhile, what was the legislature doing? A staffer reminded us that there had been serious opposition to ASAP (the old version) for some time, and bills had twice been introduced to kill it:

There were a number of concerns around the ASAP test. Some of them are the same kinds of concerns that have come up about what you call outcomes-based education, that it's value-driven, that you're funneling values to the kids that might be contrary to some people's family values, and so forth. We heard some of that, from very conservative people, we were hearing that from. I know that, for example, the House Speaker and a

¹⁸ In an interview, Bishop said, "The legislature is absent this time around."

number of parents were quite upset one year when the twelfth-grade test had to do with rain forests. They claimed that the test biased the children in favor of keeping rain forests at all costs [and] they come from it from an economic perspective which is—this might be oversimplified, but—“don’t harm people’s economics just to save the environment.” I’m probably misstating that, but you’ve got the economy versus the environment, those two arguments, and that was the clash as I understand it.... Then there were those who hated ASAP because they thought it was outcomes-based education and others, like the Senate majority leader hated it because he mistrusts educators in general.... He wanted to go back to what we had before, which was solely the Iowa test.... The people who feel very strongly about this know that that’s what they want. And they didn’t ask for research. That’s what they wanted.... We’ve got a state driven by very conservative people who think that private schools do it for a lot cheaper than public schools, so why don’t we just have vouchers. Get rid of all this mess, this big ugly bureaucracy that no longer responds to its clientele.

Observers of the legislature and its staff saw the influence of the governor in legislative action during 1996. House Bill 2417 was sponsored by the Chair of House Education committee.¹⁹ The bill proposed to modify the existing testing law to include Essential Skills tests in workplace skills, in addition to the existing areas of reading, writing, and math. It eliminated the requirement to do Essential Skills testing in grades 3, 8, and 12, instead permitting the Board to designate the three grades in which the tests must be administered. It broadened the existing law to include the administration of standardized norm-referenced tests to all students in grades 3-12. It required school district governing boards to administer competency tests for the graduation of students from high school. Beginning in 2002, only those students receiving a passing score could receive a diploma. There was a good bit of conflict and compromise over the bill within and between the chambers. What finally emerged was a bill that eliminated workplace skills from required Essential Skills testing. It allowed the Board to designate the grades (at least four grades) at which Essential Skills testing would be required. It eliminated the provision in the 1990 legislation that limited norm-referenced testing to the fall and required these tests be administered in at least four grades. In its final form, passed close to the end of the session, the bill required that the Board develop and implement competency tests for graduation from high school and establish passing scores on each (reading, writing, and math). Even after the passage of the legislation, debate over assessment policy

¹⁹ This portion of the report was taken from an unpublished paper written by Andrew W. Pearce, legislative intern, in July, 1996.

continued, however. The constituency that favored standardized testing voted to deny ADE funds to develop tests of the new Standards and amended the law to state that the Board was not obliged to test the Standards if funds were not available to do so. It also found enough money to fund standardized testing in additional grade levels.

The dominant values of efficiency and accountability of the Arizona political culture reasserted themselves in the 1996 legislation²⁰ and refuted the values of professionalism and progressivism that the ASAP reform agenda had briefly introduced. This is perhaps not surprising. The character of the process must also be noted, however. The versions of the bill passed privately among the political actors with virtually no public scrutiny or commentary. When the bill passed both the education committees and the full legislative bodies, it happened without a public hearing.

Symington Contra Keegan. Looking at the final legislation (April, 1996), it may be that Governor Symington aligned with the most conservative of the legislators to frame the bill and preempt the standards-setting process, which the bill seems to have ignored (by naming Essential Skills rather than the new standards as the content to be measured in state assessment policy).

By state constitution, the Arizona Department of Education does not report to the governor, as some other state departments do, and therefore he has no control over its budget and operations. The state superintendent who heads ADE is elected rather than appointed and thus is not part of the governor's cabinet. However, the governor appoints members of the Board of Education, which has statutory authority for educational policy. The Superintendent both serves as a member of the Board and holds primary responsibility for carrying out Board

²⁰ A separate bill, Senate Bill 1227, was introduced in January 1996 by Chair of the Senate committee on education, which would have eliminated both norm-referenced and Essential Skills testing in favor of a voluntary random sampling approach similar to NAEP. The bill would have allowed school districts to exclude themselves from state testing and administer any other tests of their choice. Staff of ADE testified in opposition, saying that the approach would not produce meaningful results. He also described the summit process. Senator Wettaw opposed the bill on the grounds of state educational expenditure and the need for accountability for spending. No public testimony was offered, and no information was presented on the experience of other states that use the sampling procedure. Lacking the support of key Majority members on the committee, Huppenthal withheld the bill and it was no longer considered. This bill would have produced an assessment policy that maximized information on school achievement status but eliminated the accountability and certification functions. Unlike the successful legislation, this bill would have countered the values placed on accountability that seem to have dominated.

policy through the Department. The tussle between Superintendent Keegan and Governor Symington, which reached a dramatic climax (but hardly its culmination) in the March 16, 1996 incident, seems particularly political in form and tactic and completely ideological and antiprofessional. No less than the governorship was at stake.

Fife Symington was first elected governor in 1990, campaigning on his record as a successful businessman and real estate developer and on his moderate position on social issues and fiscal conservatism. By 1995, he had moved far to the right in the political spectrum on every issue from the environment to education. Before the sixty-seventh Arizona town hall, he referred to the 1994 Supreme Court ruling that the state's method of financing public schools was inequitable and must be revised as tantamount to "state socialism." He went on to comment on assessment policy: "public schools last year took \$1.7 billion in state appropriations but failed to halt a 22-year slide in SAT scores."²¹

In a September 29, 1995 press release of a speech before the Phoenix 100 Rotary Club, he called for radical restructuring of the state school system, doing away with districts altogether and allowing site councils at individual schools to hire principals, who would in turn "negotiate individual contracts with teachers. No collective bargaining or master contracts would be allowed." He announced his plan to eliminate certification of teachers and administrators, free existing public schools from all laws and regulations, institute parental choice grants to enable parents to send their children to the school of their choice, and create a mechanism to place into receivership those schools that consistently fail to educate their students. In that speech he noted, "In our schools today, we have ample evidence that money and mediocrity are quite compatible.... Let's spend more money on the people who do the teaching and learning, and less on the programmers, policy experts, and paper shufflers. We need to replace the at-risk student with the at-risk administrator."

In addition, Symington proposed to abolish the ADE (Keegan's department), stating that since it was created by statute in 1970 it had grown from a half million in operating budget to \$9 million, with a staff of 350 and its own building:

²¹ Data show that the SAT scores in Arizona were above the national average and had been level for four years.

Today the ADE has grown into a burgeoning bureaucracy. This unconscionable growth is stifling the creativity of our public schools. The growth of our education bureaucracy has coincided remarkably with the decline of public education.

This is not a criticism of Superintendent [Keegan]. In the short time she has been in office she has shown herself to be an innovator, a revolutionary, and a visionary leader. She just has the misfortune to be the head of an agency which has outlived its usefulness, if indeed it ever had any.

The public education system spends over \$3.5 billion taxpayer dollars annually, with absolutely no accountability for results. We have a school report card that is virtually toothless because we have no independent, uniform testing system in place to evaluate our student's progress. We must restore the ITBS achievement testing of every student, every year in grades 3-12 immediately; we cannot wait two or more years for the Department of Education to revise the state testing program. We must set high graduation standards for all students. Amazingly enough, when the state Board of Education proposed to institute competency based graduation last year, the education lobby opposed the move on the grounds it hurt the feelings of those who didn't pass. This insults not only parents and taxpayers but kids themselves. No, not all students will shine in class, but all should graduate only after demonstrating a grasp of basic things like reading, writing, mathematics, and history.

A week before this speech, Symington had declared personal bankruptcy. The *Arizona Republic*, the state's largest newspaper, editorialized ("Symington's shifting priorities," 10/3/95, p. B4), "If Gov. Fife Symington had set out deliberately to divert public attention from his personal financial travails, he couldn't have picked a better strategy than getting his critics, and others, focused on something else."

Symington's 9/29/95 proposal followed shortly after Keegan's own restructuring proposal, which the *Republic* reported on 9/8/95 (Hal Mattern, "Schools chief: Scrap the districts," p. 1 and 23). Each school would become its own district, governed by site councils and funded by a pupil-based system to equalize school funding. Her department, the ADE, would remain, but in a reduced role and streamlined size. She proposed a computerized system so that individual schools would input student performance data as well as expenditures.

The *Republic* reported afterwards that Keegan tried to downplay the rift, but disagreed with Symington's proposal to eliminate ADE: "I was surprised ... I had

not heard about it.' Because Graham and Symington are both Republicans, and have been allies on education issues, his failure to consult her about his proposal was viewed as a snub by many political observers. Graham reportedly was angry after hearing about the plan." The governor also downplayed the rift although pointing out that they disagree on standards and testing.

The March 25, 1996 Incident. The spectacle depicted in Vignette Two displayed a rift between the two that clearly encompassed both policy (both assessment and finance) and raw politics. The year was a difficult one for Symington, who was engaged in a court battle over disputed claims in this bankruptcy and a criminal indictment. A press release from the governor's office preserved his remarks at the State Board meeting.

I have been following the effort by the board and Lisa Graham Keegan to develop curriculum standards for Arizona's public schools. I support the concept, but I am concerned about the direction the board may be taking.... The only known relationship between increasing government spending and student performance is an inverse relationship. No amount of money would ever satisfy the endless demands for more. Even more important, no amount of money could ever ensure a quality education.... In my travels around the state and discussions with concerned parents, the most pressing question they have is this: What are you teaching my children? This exercise of developing standards gives us a chance to consider that question. From age to age, there is very little difference in what a child must learn to become a literate, competent and rounded individual. In education, we have been making the same mistake humanity always makes again and again. We have casually cast aside the settled and true in favor of the trendy and allegedly exciting. [Other than technology,] there is almost nothing new about a high-quality primary education, and very little new in secondary education. Most of the social and academic "innovations" the so-called professional educators have brought to our classrooms are wasteful at best and insidious at worst. I stopped by today because some of this reckless drift toward fads and foolishness is evident in the standards currently under consideration. The reading standards, for instance, mention nothing about phonics for primary school students, nor, say, great works of literature for those in high school. They do, however, insist that our students learn to "use consumer information for making decisions," and to "interpret visual clues in cartoons." The mathematics standards state that students should be able to "explore, model, and describe patterns and functions involving numbers, shapes, data, and graphs, and use simulations to estimate probabilities." Educational concepts more familiar to most of us, such as multiplication and division, are unmentioned.

In recent years, American schoolchildren have been found to suffer a world-class deficiency of geographical knowledge. You would never know it from reading the

proposed geography standards—or perhaps you would. These standards ask students to “understand the nature, distribution, and migration of human population on Earth’s surface” which causes me to wonder what other planet’s surface human populations might be migrating on. The geography standards require nothing by way [of] identifying the nations of the world on a map, their capitals, or their core exports.

There are only so many hours in a school day and so many days in a school year. The claim of government schools on the time of young people is necessarily limited. When that time is spent on “dance styles,” or for another example, “participation in multicultural physical activities,” it is not spent memorizing rules of grammar, diagramming sentences or learning to use mathematics in a way that teaches reasoning skills. It is not spent learning the geographical history of the world or the development of Western Civilization. It is not spent studying an essential work of literature that adds to a child’s understanding of human nature or moral precepts.

The central purposes and elements of a quality education are unchanging. We jeopardize our future and that of our children by substituting fads and jargon for bedrock educational concepts. Second, the people who are most central to quality education are also unchanging. It is not academic professionals, not Ph.D. types from our education colleges, and not even teachers, although they are clearly a strong second place. It is parents who always have and always will be most important to children’s education. As I read these proposed standards, I wonder how we can keep parents involved in the education of their children. If education is re-defined in a lot of pointy-headed jargon that only an elitist core of “professionals” could ever understand, we will freeze parents out of the process. I believe we must move in the opposite direction. In fact, I would urge this board to adopt this simple standard for its own work: If the proposed standards you consider are not clearly understandable to the average parent in Arizona, throw them out.

Informants present at the board meeting reported that Symington had brought Bishop with him, but she said nothing. Keegan appeared taken aback by their appearance and his pointed criticism of her standard-setting program. She offered a weak defense, pointing out that math facts and operations were indeed mentioned in the standards and that more basic forms of literacy were also addressed. Insiders on the design teams were angered and befuddled. Some time later, one informant explained the event as Symington’s attempt to weaken Keegan’s political position and strengthen the conservative Board members’ resolve to renounce any progressivism remaining in the standards.

Answering questions from the floor of the board meeting, Symington held up the Benjamin Franklin Traditional Schools in Mesa as an example of the

kind of school that all Arizona students should have. That school is famous for its teaching of the “three R’s,” homework policy, its required uniforms, strict discipline, and a code of conduct. Reporting on the above event the *Republic* (3/26/96, p. 1 and 3), Hal Mattern wrote:

Symington’s comments at the board meeting caused a stir among audience members, with debates spilling into the hallways after his speech. The meeting was packed with people supporting the governor’s comments, including many parents from Ben Franklin. Many of them expressed concern not only with the proposed Arizona standards, but also with the federal Goals 2000 program, which they said is an attempt to impose national standards. “We want to have a choice, not be told what our children will be taught,” said Syd Curtis, who has a child at Ben Franklin. But critics of Symington’s comments said the governor is the one who is taking away parental choice by insisting that all public schools adopt traditional programs. “A lot of parents don’t want just the basics,” said Sue Braga, legislative chairwoman for the Arizona Parent-Teacher Association. “Yes, students need to learn basic skills. But we want to go beyond that. [Symington’s] proposal is setting the stage to put all children into the same mold. He is talking about a mandate that would put us back into the 19th century.” Symington, however, said that the current system isn’t producing the desired results and that establishing new standards isn’t the answer. “We should just start over and go back to the basics,” he said.

In a political analysis (*Arizona Republic*, 5/5/96, p. 1-13), headlined, “Symington moves to right seeking votes,” Michael Murphy wrote:

Symington, who once cast himself as a moderate Republican, is laying the groundwork for a re-election bid by courting the most extreme elements in the conservative coalition.... [P]olitical observers in Arizona agree that Symington, whose popularity ratings have spiraled downward because of his personal bankruptcy and the indictment of two close associates, has adopted a strategy of fiery neo-populism. The idea is to build a core of supporters among the state’s hard-liners who would be the backbone of a 1998 re-election campaign.... [O]ne close political ally indicated that Symington has developed a political playbook focused on picking hot-button issues that resonate among the GOP’s most conservative elements. [Although he hasn’t announced whether he will seek a third term ..., he] has formed a 1998 exploratory committee and authorized supporters to schedule campaign fund-raising events.

Symington’s interest [in the standard setting process] was spurred by Dinah Monahan, a Snowflake [Arizona] mother of five and a leader in the Eagle Forum, a conservative lobbying group founded by Phyllis Schlafly. She is mobilizing other Christian Right groups, including the Christian Coalition and the Concerned Women for America, to

fight what she calls the “humanist, globalist, New Age indoctrination” of Arizona school children.

Under the umbrella of “Citizens for Education Excellence,” the groups warn of secret conspiracies at the Department of Education, including one to establish computer files like those used in China to collect data on each pupil’s thoughts and feelings.

“Arizona’s education is being hijacked and our children will be held hostage,” Monahan wrote in the group’s newsletter. While some dismiss Monahan as a fringe activist, she boasts of a close working relationship with the governor.... “He’s committed to the death,” Monahan said.

Keegan, a conservative by most standards, believes Symington’s real interest in the issue may be in making headlines.

The reporter went on to describe Symington’s other far-right moves, such as opposing gay marriage, pushing tough legislation that would treat juveniles in the adult courts, revoking the state income tax, and joining ranks with the Wise Use organization (an anti-environmental, pro-rancher group), offering at a Town Hall meeting to “shoot a spotted owl” if necessary to help their cause.

Events in March were not the last in the contest between Keegan and Symington. In October 1996, Keegan called for the governor to resign (*Arizona Republic*, 10/18/96, p. A1 and 18, Marin Van Der Werf and Kris Mayes, “Keegan: Symington should quit”). The newspaper reported that Keegan had participated in a group that had been meeting to discuss his situation. Alone among the Republicans and putative gubernatorial nominees, she called for his resignation, saying he was no longer a productive or effective leader. At that point Symington was facing a bankruptcy proceeding that his creditors had challenged, a 23-count criminal indictment concerning alleged campaign disclosures and improprieties, a recall effort (since failed), and a public approval rating of 19 percent. “Keegan once enjoyed a close relationship with Symington. They basically ran as a team in the 1994 election, but their relationship has since soured.” “I think Lisa is just having a bad-hair day,” the reporter quoted House speaker Mark Killian as saying. “I think this is a time when cooler heads ought to prevail.” Recent survey results were reported in which 71 percent of Arizona residents polled think Symington should resign.

But neither the legal and financial problems nor opposition from fellow Republicans could slow Symington's pursuit of his policy agenda. The appointments he made to the State Board during 1996 and his working behind the scenes with the legislature further reinforced his conservative stance.

Approval of the Standards—Hooked on Phonetics

The original Summit plan forecast that the Board of Education would receive the standards in time for approval at its January, 1996 meeting. Events intervened. The teams asked for extra time. The social studies team disintegrated after one draft. There were several iterations of review between the teams, the consultants, and the ADE. The Board appointed a review team to speed things up. The Board had business that assumed a higher priority, including removing teacher certification requirements and hearing problems with various charter schools. Months dragged on. The standards were first on the Board's agenda at its January meeting. The draft standards in math and language arts (primarily the latter) were discussed at every meeting during the spring. Although the Board does not normally meet in the summer months, members decided to make exception and met three times. Standards in reading were approved in July and writing and math in August. By summer, the composition of the Board had shifted, and the alliance between the Board and the Superintendent had weakened.²² Redefinition and reduction of language arts as reading and writing (but not listening/speaking and visual representing) was complete. The form and substance of the approved standards than the standards the design team had submitted. Seven other drafts of standards (e.g., foreign language, health, science) remained out of sight.

Other significant shifts occurred during these months. The Board approved ADE's recommendation of the Stanford-9 achievement test as the state-mandated standardized test and determined that it would be administered in every grade 3-12 (exceeding even the grade coverage the legislation mandated). The Stanford-9 replaced ITBS and TAP, the previous form of mandated norm-referenced testing. The reasons for the change were never articulated for the public; neither were the criteria for selection of the replacement. An ad hoc

²² An informant remarked, "There is a feeling at the department that Diane is advising Fife on who to appoint to the board, because of animosity toward Lisa. Lisa had this whole plan ready to go, and then the membership of the Board changed and she has been stymied in everything she has tried since then."

committee had looked at the design specifications of the two batteries and decided that Stanford-9 was the better match with the state content frameworks.

The Board also imposed itself on fundamental elements of the Design Teams' work. Seemingly from out-of-nowhere, a Board member demanded a change in the developmental levels of the standards to require benchmark tests at the end of grade three instead of grade four, as had been true since before the Summit. As a design team informant noted, this changed everything: "As if third graders were not different from fourth graders!" Incensed but powerless, the design team had to go back and modify the standards to be appropriate to the new configuration of grades and levels.

As several observers and informants related to us, the Board interpreted its task of approving standards quite broadly. Indeed, they minutely inspected each standard, bullet, and level.²³ The primary bone of contention was the extent to which basic skills should be made explicit in the standards. The newly appointed conservatives insisted on explicit inclusion of rote memorization of math facts, direct instruction of spelling and phonic skills, and exclusion of anything they deemed unmeasurable or in some way progressive. Keegan, as a member of the Board, defended the drafts as submitted, not only because they were the products of extensive work and long and painful negotiation, but because she wanted to distinguish content standards (what students know) from teaching processes. The latter should be more up to the discretion of schools, according to her. But the conservatives wanted more control, more uniformity.²⁴ An informant describes the climactic moment in the July Board meeting:

All along Lisa hadn't wanted phonics mentioned as a standard because it was a process, and she didn't want instructional techniques in the standards. But Felicia, who is a teacher at Franklin Traditional school and is about as far right as you can get—she was unbelievable. She was relentless. She had been wearing people down all day. All day long they had been going back and forth, back and forth about whether to put the phonics in as a standard. Finally it came down to 5 o'clock in the afternoon and

²³ A point-by-point comparison of the approved standards and the final draft (Draft 4) submitted by the review team shows that modifications were few, except for the addition of standards related to phonetic skills at the foundations and essentials levels (e.g., R-F1, "Use phonetic skills to decode words," and "Use structural analysis skills such as identifying root words, prefixes, suffixes and word origins to decode words unfamiliar in print," R-E1). Amendments to writing standards were few. Of course, standards submitted on listening and speaking, and visual representation were omitted entirely.

²⁴ The apparent paradox of local control and heavy accountability/control over curriculum is resolved in postfordist theory (Ball, 1990).

everyone on the Board had left but five people and you need all those five votes to pass anything. And Keegan and the Board president caved in. And I had quotes from her earlier in the day that said she wouldn't go for it, but she did. Because Felicia said in so many words that she would not approve the standards unless phonics were *not only* part of K-3, but also part of K-8. So now we have phonics all the way to eighth grade. I think they caved in just to get *something* passed. Anything. All these months had gone by, and still nothing had been approved. They had already given up standards in listening and visually representing; that was gone, and I think they were desperate to get something officially approved. So what they did, they agreed to a change in wording. Instead of calling it phonics, they called it *phonetics*. And the vote was 5-0.

The Board appointed three subcommittees, and assigned them limited responsibility. They appointed 120 teachers and district content specialists to the content committee and asked them to translate each of the approved standards into performance objectives. The technical committee, 20 people from district assessment offices and local university researchers, was charged with the task of advising the content committee on whether each of the performance objectives they had written were measurable.²⁵ The special populations subcommittee had the task of alerting the content advisory subcommittee to any potential problems related to special education or language minority pupils.²⁶

As they had during the revision of draft standards, members of the Board also participated in the subcommittees. Referring to the role of a Board member, a committee informant had this to say:

He agreed to certain things in the subcommittee but when it went before the Board, he spoke up again as if he was not a member of the subcommittee and had not already had a chance to state his case. So he got to state his case twice, and guess what. In the board meeting when the decision was made, he went the other way. So you can imagine the anguish of having gone through all these different steps—compromising here, giving in there, wording things in certain ways so that it won't be overconstrued or overly tested and then when it finally gets to the Board, you have a lay board making curriculum decisions based on politics, based on the kind of feedback they get from home.

Another committee member spoke of a Board member's participation on the technical subcommittee:

²⁵ The technical committee exceeded its assignment by issuing a white paper on assessment principles, limitations, and purposes.

²⁶ This group also had a limited assignment that members attempted to expand. Membership overlapped a private group, the Central Arizona Bilingual Consortium, which attempted to express policy initiatives related to the assessment of language minority pupils.

The inclusion of phonics as a performance objective caused real problems. Then Janet Martin, a new appointee, very conservative, is on the technical committee and saw that there weren't many performance objectives stated. She didn't like anything we did. We wanted some of this to be left for district tests, but she wanted all of it covered in the state test. She's very big on accountability. Apparently over night she called all the members of the State Board and the representative from Harcourt Brace, which publishes the Stanford-9, and told him she wanted to get phonics put into the state achievement test as well as the Essential Skills test, and that she had enough support on the board to require it. But the rep. said it was not part of the contract, that third grade test doesn't even have phonics, so they'd have to give the second grade test. At the last board meeting, she asked the second grade Stanford-9 phonics test be added, which means that the test would be 48 items and 23 minutes longer and that the teachers would have to give it verbally. Keegan pointed out that using the second grade test with third graders would destroy the norming information, and that was news to the board members. But it will probably go through anyway. It also is big bucks, because it would have to be added to the contract. Felicia wants all the standards laid out for each grade. She wants to specify spelling and memorization. She is very strong in her convictions and she absolutely won't give in, and she works on everybody else and wears them down.

A person who saw the evolution of ASAP and the content standards spoke of "a broken system" and a loss of democratic debate over substance in favor of backroom bargaining over ideology and power politics:

Given that we were in agreement on many core issues and were highly committed, I attribute our team's inability to get that through the process to the fact that there were so many different actors with different levels of authority and very different philosophies. Until you resolve that, you have a broken process. While we were coming along and coming to terms with our differences, the board had a different philosophy. So when it comes to things like whether kids are going to be able to use calculators, whether you teach phonics separately, is it okay for kids to form hypotheses and make value judgments, those are fundamental issues. And if you don't have basic agreement on those issues, how are you going to make decisions about how much the test will be performance based and how much closed end items. We have people here with authority who are not willing to talk about basic philosophical issues, who are not interested in consultation and resolution. They want to appear to be part of a group deliberation, but they still want to exercise the kind of control that they've always been able to exercise. When do we get to talk about educational philosophy? We talked about it in the team, we were able to find the ground for discussion, but not in the system. We didn't have board members listening to our deliberation. They were allowed to keep their feelings to themselves until the decision was made. Then when it comes time for finalizing the draft, that is not the time to have a philosophical discussion about how constructivist or how skills-based we should be.

Another subcommittee member put his finger on the distinction between the substance of curriculum development (what he thought he had been working on) and the politics (what had been working on him):

The same people who are committed to more accountability are equally committed to local control—that is the control by individual parents. It’s a paradox. But that’s what I learned from this process, that local control turned out to be the most dominant factor in the equation. I underestimated the power of that. We knew those groups were out there, whose blood pressure goes up at the mention of Goals 2000. But we underplayed it because we didn’t know the degree to which individual board members would be willing to go to bat for those voices. We’d like to think that democracy works, and representative government, but in this structure, a small number of people can be in touch with a small number of board members and get the job done.

The New Shape of Assessment Policy—“The Same Train Wreck”?

Although by this writing it is still unclear what the next shape of assessment policy will be, it is clear that ASAP is dead. No official talks about problem-solving, integrated, thematic teaching, or reading-writing connections or testing the way teachers teach and learners learn. Discourse about progressive reform no longer occupies the wings, let alone the center stage. Talk about performance assessment as a way of authentically representing the way teachers teach and help learners learn has given way to concerns about efficient testing of basic skills, isolated by subject matter. While the constituency for progressive reform left the field, its members having lost their bureaucratic positions or renounced them for greener pastures, the constituency for standardized testing remained uniform, consistent, vocal, relentless. The test burden on Arizona school children is higher than it was before ASAP, with standardized testing for every pupil in all grades 2-12 and the testing of phonics skills extending even into junior high. The new Standards testing (reading, writing, and math) will take place in four grades as well. In addition, the districts will be required to develop and implement local assessments at the other grades and of the Standards not covered in the state battery. Student mastery will be reported to the Arizona Department of Education, which will make all these data available on the Internet. High-stakes accountability is also greater than it had been before, with graduation and promotion tests fixed on the agenda.

Though we cannot yet see the end of the line, we can see what lies along several mileposts. First, the state assessment policy has moved farther to the right both politically and pedagogically. Second, Arizona Standards have replaced the Arizona Essential Skills, at least in reading, writing, and mathematics, although some observers claim that there are more similarities than differences in the two versions of content standards. Yet to be finalized is a Request for Proposals for publishers and others who wish to bid on a contract to develop tests that match the Arizona Standards, that is, the replacements for the old performance tests, Form A and D. ADE staff predicts that the RFP will specify a mixed format of multiple-choice, short-answer and a few extended-response items. They cite the need for reliable and objective assessments to serve the function of comparability and pupil-level accountability. Their choice of format also reflects the cultural value of efficiency. When the ADE appealed to the legislature for funds to support development of Standards-based tests, it was rebuffed, the legislators noting that the Stanford-9 Achievement Test ought to be sufficient for all state assessments. ADE will now have to fund the RFP out of its other appropriations and programs (as it had had to do for ASAP). Political observers interpreted the legislators' decision as another skirmish in the Symington-Keegan fight.²⁷

What's missing from assessment policy in early 1997? The RFP specifies testing in English only, a situation that forecasts future problems for Arizona's limited English speaking children. Concerns about equity for ethnic and language minority and disadvantaged pupils are recast by policy actors. The official line is that the state must set the bar high and equally for everyone, and that it is racist to think that all children cannot vault it successfully with the available pole.²⁸ The Central Arizona Bilingual Consortium published a briefing

²⁷ The Chair of the Senate Education Committee introduced a bill that would have scrapped requirements for mandated state standards and assessments other than Stanford-9 and was reported as saying, "that the state should avoid major testing changes each time a newly elected schools chief takes office" (*Education Week*, 2/19/97).

²⁸ During an interview, Keegan responded to a concern that minorities may suffer adverse impact based on their lower scores on mandated tests: It is one of the reasons that you set the even bar for everybody well. And I'm a particular—I think that is a pernicious and quite frankly sort of a racist view that particular students don't deserve a high expectation, and that's what I read that as. I mean I've listened to that for years about school choice; you can't give poor parents or minority parents a choice because they don't know how to use it; the parents won't profit from it. I don't believe it. And I've seen too many things about students, just in general, coming into a good education system. Where there is high expectation of that student, they rise to the level of expectation. And I just don't believe that the problem with our minority kids is capacity; I think it's expectation.

paper on the possible impact of high-stakes testing on minority pupils, but their voice had little chorus from other possible advocates. Nor was there much protest by teacher organizations or the progressive educators who had played a supportive role in the early ASAP days.

What else is missing? No one debates that the main fallacy of ASAP was the absence of provision for professional development. Yet the new assessment policy also fails to provide for it. ADE policy actors give lip service to teacher training,²⁹ although they lack the power or budget to do anything else. The legislature that has the power is disinclined, because of higher priority on tax-saving and general anti-professional sentiment. Once again, the districts and individual teachers are left to their own devices about retraining, curriculum development, adjusting to the demands of the new tests, and the like. In general, teachers were left out of the process of assessment policy, although they had some limited opportunity to respond to drafts of standards. The curriculum subcommittee that worked in 1996 could only translate the standards into performance objectives, but not debate on their rationale and substance. Only the Board got to do that.

Missing as well, as it had been during ASAP, was any plan to evaluate the consequences of the new assessment policy. Current policy actors have the same theory of intervention that their predecessors had: put the test out there, attach high stakes to it, and educators will adapt to it. And if not, market forces will take over. Worse now than before perhaps, is educators' sense that a game is being played out on them. They are more cynical and less likely to speak out on policy issues and more likely to retreat behind their closed classroom doors. A district informant talked about teacher reaction:

And so everything from my philosophy, if I won out, it's going to be—what it will look like is a clear expectation for all students, and then we're going to have to figure out ways to expect better things of all students. I don't think there's any question if you've looked very hard at public education that we have a lower expectation for certain groups of students than others. That's not news to anybody. And our problem is how to get over our expectation problems, not how to lower the bar. So I've read the same stuff, and it makes me angry, because it's failing to do well by the kids and then excusing ourselves for that. So I don't buy it."

²⁹ The Superintendent said in an interview: "We have tried to put out a part of the plan that ... talks about professional education. We don't have it done because it's an expense, and we need to convince the people who hold the purse strings on this that's a good thing to do. Now, whether that's the legislature or the voter at some point, I don't know. But any credible reform effort is going to speak professional development.... I don't think there should be an expectation that you're going to get a change in standards or a change in student skill or what students learn without a significant dedication to human capital. I think that's counter intuitive. I think it's wrong."

A lot of them regard it as a joke. It was a good idea, but now politics has become more important than substance. And they are just tired of being jerked around.

Another informant predicted an outcome for Standards-testing similar to that of the ASAP:

If teachers see merit to the standards, if there is a high correlation between what teachers and administrators and parents value and what makes its way into the standards, and there's willingness to commit to that, then it could be positive, contribute to positive change. There were places, and this district is one of them, where ASAP was an engine for change, for inservice training, for tightening up some things. But the fact that the system that we're in doesn't promote long-term attention to these basic issues, minimizes the positive impact of what we already had going.

Nor has there been any concern expressed for opportunity-to-learn or delivery standards. The graduation competency requirements will kick in before any official attention to curricular offerings has been paid. This will be most serious in the consequences of the math standards, some of which mirror the National Council of Math Teachers Standards. No one knows how widespread these standards have permeated math classrooms in Arizona, but the math curriculum experts believe that the dissemination is uneven and slow overall. In those districts with sub-standard math programs, it will be students who suffer the high-stakes consequences when the tests test what they have not been taught. In interviews, the policy actors pooh-poohed OTL issues, and refused to see them as anything but a bid by the district establishment for more money or equalized capital funding.³⁰ They couched these issues this way: that *after* students fail the tests, their parents will be motivated to demand more of the schools or to choose other schools for their children to attend.

Finally, what is missing from 1997 assessment policy is respect for the psychometric demands of creating, piloting, and revising the new tests. The technical subcommittee was assigned a limited role—not to advise on technical issues, but to judge whether each performance objective the curriculum subcommittee wrote was indeed measurable. There are signs, however, that the technical subcommittee will exceed its brief. Members prepared a white paper for the Board of Education to consider such basic issues as technical standards, costs, and matching the form of the assessment with its intended function. The

³⁰ Keegan said during an interview: "I don't believe the opportunity to learn stuff. There's nothing in any literature that suggests that amount of money makes the difference ..."

subcommittee asked for the chance to review the RFP and the proposals to be submitted by test publishers, and to propose validation and evaluation studies. Against this hope for design rationality (Schon & Rein, 1994), lie several characteristics of the Arizona political scene. The first is misconceptions of basic issues and language of assessment in the minds of the most powerful political actors. Psychometric expertise has not been added to ADE to offset this problem. After presenting the white paper to the Board, a member of the technical subcommittee reported that most of the Board members had no idea what he was talking about, had never heard of professional standards for testing, and were not even sure—even after all those months—what a standard was. Second is the orthodoxy of the same actors toward norm-referenced standardized testing and mass testing of all pupil on all basic skill standards. Third is the efficiency value in the state political culture that scrimps on both time and money. Appeals of the technical committee for more validation studies and more revisions are likely to be dismissed as too extravagant. Furthermore, like ASAP performance tests before them, the new assessments will have to be developed in a matter of weeks (three months is the current estimate) rather than years, because the Board and Legislature demand for accountability. The nature of test development is such, however, that this abbreviated development time will preclude adequate pilot administrations and revisions. Like ASAP, glitches in administering assessments will not be worked out before assessments are mandated and their scores used in high-stakes decisions. An ADE staff member remarked, “if we don’t start soon, we will have the same train wreck.”

Time, money and politics: not enough of the first two and too much of the last. Just as the Bishop administration recognized a brief policy window of opportunity, the Keegan administration believed it was necessary and possible to reconstruct state content and performance standards and develop relevant instruments in less than a year. Pressure from other policy actors also constricted the time frame. But months have turned to years and enthusiasm for, and political will necessary to achieve instrumental aims have waned. As a district informant noted about the intersection of the three:

There wasn’t enough political will to really enforce those high-stakes rules. Even now, I think that is the case.... And we’re only talking about reading, writing and math. This administration has thrown in the towel on the bigger picture. How can we graduate kids or not graduate them based on reading, writing and math. Do you think

the political will is there to implement the rest of the standards? I'm telling you it is not. And think about the cycle. This is going to cycle into another administration. Are you going to have that kind of commitment [to continuing and maintaining the process now put into place by the current administration?] The structure itself is such that what we're trying to do is almost undoable, because the time it takes to develop the consensus, develop the standards, develop the test, pilot the test, make sure the tests are operable and then be able to say with some confidence that now we can say that it is fair and justified to make graduation tied to these standards. How many administrations [of superintendents] does it take? A minimum of two, maybe three. Now how are you going to keep it going? So even within *one* administration there was an inability to keep the focus on this one issue, let alone *across* administrations.

The very idea that an assessment policy would persist from one administration to the next now seems unlikely, even fanciful, if one looks at this story closely. It's difficult to imagine that the next election will not pit Keegan against Symington and put another face in the superintendency and another need to stage a new spectacle in the name of assessment policy.

Can Something General be Learned From This Story?

It is not for nothing that the confrontation at the O.K. Corral took place in Arizona. There is something of Symington's March 25, 1996 invasion of the Board meeting that resonates with that mythohistorical event. Arizona often strikes observers as an aberrant case in much of its politics and policies. In the year of ASAP's demise, two successful items on the legislature's agenda were the "veggie hate crimes bill" and the "polluter protection act," for example. But Arizona is not quite so aberrant that general lessons cannot be drawn. The dramatic suspension of the California Learning Assessment System shows remarkable empirical similarities. Even if the facts of the case do not generalize, however, propositions, ideas, categories, and moral lessons can still be drawn. We look back to our conceptual frameworks to find the general in the particular. We have distilled a set of tentative assertions about Arizona's change in assessment policy.

Assertion One: The state political culture tugs policy more often toward its dominant values of efficiency, accountability, and choice and away from contending values of effectiveness, equity, and professionalism. Thus, an assessment policy like ASAP, which emphasized the improvement of teaching and learning, was likely to be unstable and to be overrun by demands for tangible

test results for the least expense. Claims of policy to enhance fairness and equity were also unlikely to endure for long. The policy that reflects localism and the values of the existing elites were likely to win out over centralized state control and distribution of opportunities and resources.

Assertion Two: The national discourse on education filtered into the political and policy culture (Ball, 1990; House, 1991). Claims that public schools are failing were repeated as policy actors proclaimed the need for greater stringency in state assessment policy. Discourse about the link between achievement test scores and economic competitiveness reinforced this trend and an image of schools as factories manufacturing achievement test scores and producing economic prosperity. The role of corporate elite and national networks of conservative actors in the policy-making and standard-setting process further revealed the influence of national political culture and political trends. The shift to the political right that characterized national and state politics from the late eighties to the middle nineties also influenced assessment policy.

Assertion Three: Within broad structural limits, policy itself is neither unitary nor invariant in that different actors in different situations interpret and experience assessment policy differently (Hall, 1995). At the stage of policy formation in Arizona, policy actors constructed links between assessment solutions and putative problems. The “problem” as defined by some actors was underperforming public schools and their lack of accountability. Others defined the “problem” as an outmoded form of pedagogy held in place by an outmoded high-stakes standardized test. With her progressive-minded staff, Bishop as policy entrepreneur grafted constituencies together to get ASAP on the policy agenda, doing so by obscuring the underlying contradictions between the two problem definitions. Both change and the pace of change (at both the birth and death of ASAP) can be explained by garbage can theories of policy-making (Kingdon, 1995). This theory suggests that there is a narrow window of opportunity during which the various constituencies (each with different policy goals) can be brought together to get the policy on the agenda. Coalitions of constituencies with conflicting agendas and interests proved to be unstable. The constituency for progressive reform through the instrumentality of performance testing was scattered and silent by 1995.

Assertion Four: Once on the agenda, a legislated policy is still neither fixed nor invariant, as the ASAP case illustrates. Hall’s model of policy as a process of

transformations from the original goals and intentions through layers of administration and implementation (Hall, 1995) helps explain how the Bishop administration could take a piece of legislation that specified only that essential skills should be assessed and reported and turn it into a program of performance assessment and reform of teaching and curriculum. That the latter emphasis was unmentioned in the legislation marks ASAP as a radical transformation and helps to explain why the assessment policy was unstable.

Assertion Five: The changes in assessment policy are best explained as political action—political spectacle (Edelman, 1988) and political competition as opposed to design rationality (Schon & Rein, 1994) and instrumental intent. Still, it would be a mistake to declare that any policy is solely political in nature.

Assertion Six: The most important influence on changing assessment policy was the change in policy entrepreneur (Kingdon, 1995). Bishop had defined ASAP as the centerpiece of her administration. The political competition between Keegan and the new coalition of Bishop and Symington in 1995 made a change imminent, even though Keegan had previously showed less interest in the substance of assessment policy than in finance and school choice policies. The events of Arizona assessment policy fit Edelman's words (Edelman, 1988):

The term "leader" evokes an ideal type which high public officials try to construct themselves to fit. In this sense leadership is dramaturgy; for regardless of the consequences of officials' actions, which contemporaries cannot know, the ability to create oneself as the ideal type maintains followings.... The leader must be constructed as innovator, as accepting responsibility for governmental actions, as possessing qualities that followers lack, as successful in his or her strategies in contrast to the mistakes of earlier leaders, and when unsuccessful, as victim of insuperable obstacles placed there by adversaries or enemies. (p. 40)

Assertion Seven: Empirical and rational analysis played little role in Keegan's decision to change state assessment policy. The decision-making process took place out of the public eye and without recourse to expert consultation, evidence and argument. A close examination of the evidence casts doubt on the contention that ASAP was any more flawed technically than it ought to have been or that field tests and revisions could not have substantially corrected those flaws. More important influences were promises she made to her conservative supporters during the campaign, the symbolism of her own stamp on educational policy, and the spectacle that showcased her gubernatorial ambitions.

An expeditious decision symbolizes for the public a decisive and competent leader (Edelman, 1985). An extensive and open process would have lacked the dramaturgical effect achieved by Keegan's decision to change state assessment policy. Edelman argues that most politically controversial policy actions are couched in rational language to reassure the public in the competence of its leaders, whereas what counts as evidence is ambiguous and largely rationalization of self-interest.

Assessment policy is not solely political in nature. Policy actors and those who administer policy have positive intentions to do substantive good. Instrumental goals are both implicit and explicit at all levels. But to ignore its political nature; that is, to treat it solely as rational and instrumental, is to engage a cycle of confusion, optimism, frenzied activity, disappointment, and cynicism. The learning and change cycle works at a different pace than the political cycle. If the political cycle is the political term limit or the four-year election cycle, and the educational learning/adapting cycle is longer, than a major disconnect is inevitable.

At the political level, major and rapid changes occur relative to office holder (political entrepreneur) turnover, terms of office and the garbage-can notion of window of opportunity, waning political capital, and the like. At the bureaucratic level, the pace of change is longer as actors hold their position longer than politicians, and can have major influence on how formal rules and regulations are translated into programs and oversight activities. Also, the norms of profession are likely to intrude here, as bureaucrats are recruited from the ranks of professionals. At the level of practice, things change much more slowly as it takes time to phase out prior materials and influences, acquire compatible materials and learn new things, and to displace incompatible ideologies and interests. That the world of practice proceeds beneath a screen, a filament that separates it from the world of policy and politics. The screen filters in some elements (districts change their tests, teachers struggle to make sense of the dictates; they engage in self-educating, buy new materials, hire new trainers). Some elements take hold. Some are overtaken by events and the changing political scene. New policy entrepreneurs emerge with needs to demonstrate their novelty, their competence, their leadership, their responsiveness to political constituency. They impose new agendas, new programs, in this case new tests, that, from the standpoint of the practitioner, layer over the antecedent programs,

creating a kind of geology of policy. The work of practice is real work. But the world of politics is a world of symbol and myth that most of us never touch.

REFERENCES

- Ball, S. J. (1987). *The micropolitics of the school*. London: Methuen.
- Ball, S. J. (1990). *Politics and policy making in education: Explorations in policy sociology*. London: Routledge.
- Berliner, D. C., & Biddle, B. B. (1995). *The manufactured crisis: Myths, fraud, and the attack on America's public schools*. Reading, MA: Addison-Wesley.
- Edelman, M. (1985). *The symbolic uses of politics*. Urbana, Illinois: University of Illinois Press.
- Edelman, M. (1988). *Constructing the political spectacle*. Chicago, IL: University of Chicago Press.
- Erickson, F. E. (1986). Qualitative methods in research on teaching. In M. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed., pp. 119-161). New York: Macmillan.
- Flexer, R. J., Cumbo, K., Borko, H., Mayfield, V., & Marion, S. (1994). *How "messing about" with performance assessment in mathematics affects what happens in classrooms* (CSE Tech. Rep. 396). Los Angeles: University of California, Center for Research on Evaluation, Standards, and Student Testing (CRESST).
- Gaddy, G. B., Hall, T. W., & Marzano, R. J. (1996). *School wars*. San Francisco: Jossey-Bass.
- Hall, P. M. (1995). The consequences of qualitative analysis for sociological theory: Beyond the microlevel. *The Sociological Quarterly*, 36(2), 397-423.
- House, E. R. (1991). Big policy, little policy. *Educational Researcher*, 20(5), 21-26.
- Kingdon, J. W. (1995). *Agendas, alternatives, and public policies*. (2nd ed.). New York: Harper-Collins.
- Lipsky, M. (1980). *Street-level bureaucracy: Dilemmas of the individual in public services*. New York: Russell Sage Foundation.
- Marshall, C., Mitchell, D., & Wirt, F. (1989). *Culture and educational policy in the American states*. New York: Falmer Press.
- McDonnell, L., & Elmore, R. (1987). Getting the job done: Alternative policy instruments. *Educational Evaluation and Policy Analysis*, 9(2), 133-152.

- McDonnell, L. M. (1994). Assessment policy as persuasion and regulation. *American Journal of Education*, 102, 394-420.
- Miles, M., & Huberman, A. M. (1994). *Qualitative data analysis* (Second ed.). Newbury Park, CA: SAGE.
- Noble, A. J. (1994). *Measurement-driven reform: The interplay of educational policy and practice*. Unpublished doctoral dissertation, Arizona State University, Tempe, Arizona.
- Noble, A. J., & Smith, M. L. (1994). Old and new beliefs about measurement-driven reform: "Build it and they will come." *Educational Policy*, 8(2), 111-136.
- Nolan, S. B., Haladyna, T. M., & Haas, N. S. (1989). *A survey of Arizona teachers and school administrators on the uses and effects of standardized achievement testing (Technical Report 89-2)*. Phoenix, AZ: Arizona State University-West Campus, Education and Human Services.
- Olivarez-Seck, M. (1994). *Districts' capacity to respond to state-mandated reform*. Unpublished doctoral dissertation, Arizona State University, Tempe, Arizona.
- Rein, M. (1976). *Social science and public policy*. New York: Penguin Books.
- Schon, D. A., & Rein, M. (1994). *Frame reflection: Toward the resolution of intractable policy controversies*. New York: Basic Books.
- Shepard, L. A., Flexer, R., Hiebert, E., Marion, S., Mayfield, V., & Weston, T. (1995). *Effects of introducing classroom performance assessments on student learning* (CSE Tech. Rep. 394). Los Angeles: University of California, Center for Research on Educational Standards and Student Testing (CRESST).
- Smith, M. L. (1997). *Reforming Schools by Reforming Assessment: Consequences of the Arizona Student Assessment Program* (CSE Tech. Rep. 425). Los Angeles, CA: University of California, Center for the Study of Evaluation/Center for Research on Educational Standards and Student Testing (CRESST).
- Smith, M. L., Edelsky, C., Draper, K., Rottenberg, C., & Cherland, M. (1989). *The role of testing in elementary schools* (Tech. Rep. No. 321). Los Angeles, CA: University of California Graduate School of Education, Center for Research on Educational Standards and Student Testing (CRESST).
- Smith, M. L., Noble, A. J., Cabay, M., Heinecke, W., Junker, S. C., & Saffron, Y. (1994). *What happens when the test mandate changes? Results of a multiple case study* (CSE Tech. Rep. 380). Los Angeles: University of California,

Center for Research on Educational Standards and Student Testing
(CRESST).

Smith, M. L., & Rottenberg, C. (1991). Unintended consequences of external testing in elementary schools. *Educational Measurement: Issues and Practice*, 10, 7-11.

APPENDIX A

The Empirical Work

We base the validity of the narrative and analysis that make up the body of this monograph on three sets of empirical evidence gathered over the entire period of ASAP history. The first set of data was generated by examining documents and interviewing key policy actors who contributed to the enabling legislation and translation of legislation and policy goals into program administration. A detailed account of the methods of collection and analysis is given in Noble (1994) and Noble and Smith (1994). Interviews were conducted in 1993 with 13 policy actors, including legislators, the State Superintendent and involved officials at the Arizona Department of Education, an ADE advisor from the local university, and officials at the Arizona Education Association and Center for Establishing Dialogue in Education. The aim of the interviews was to reveal the intentions of policy actors and their expectations for assessment policy and its consequences, the images they held about the nature of teaching, learning, assessment, and reform, and their perceptions about early reactions to the new policy. These interviews were recorded, and their transcripts provided data for the present analysis. We gathered an extensive array of documents, including legislation, rules and regulations, ADE announcements and newsletters, the assessments and rubrics themselves, and as many of the technical reports as were made available. Direct observations of ADE and State Board open meetings and workshops rounded out the data collection of the policy study and later fed into the analysis reported herein.

The second source of data was a two-year study of initial reactions by schools to ASAP (Smith, 1996). During the first year, a multiple case study design placed researchers in four elementary schools for a full school year to understand the meanings and actions of educators in particular contexts as they came to terms with state assessment policy and tried to implement it. Classroom observation, interviews with teachers and administrators, examination of curriculum and testing documents were the modes of data collection. We analyzed data within and across sites (Miles & Huberman, 1994). We accounted for the local status of ASAP implementation by the financial and knowledge resources available, the

compatibility of local images and ideologies with those of the state policy, and the accountability culture.

During the second year, which coincided with the second year of ASAP Form D administration, we conducted focus group interviews at the four original sites as well as five other, purposively selected schools. The second-year interview agenda consisted of the following parts. An opening statement laid out the direction for the participants: "As you know, the ASAP program was intended to change schools toward more holistic, integrated instruction and to make schools more accountable. We would like to know how schools have reacted to the ASAP program. What does the reading, writing, and math curriculum look like here? How do you see it as consistent or inconsistent with ASAP? What do you think a teacher needs to be able to know and do to implement the ASAP program? How does that fit or not fit with your own knowledge and teaching skill or philosophy? In your view, what has happened at this school as a reaction to ASAP? What if anything has gone on in this school or district in terms of helping teachers teach more holistically (consultants, inservice, collaboration, etc.)? What messages do you get from administrators or the public about the importance of high ASAP scores? What if anything do you do to make sure your students score well on the ASAP?" The transcripts of recorded interview fed into the analysis of the present study.

In addition to the focus group interviews, we conducted surveys of educators sampled representatively from the state as a whole. Our research questions were as follows: What is the status of change toward ASAP policy ideals from the perspective of teachers? What is the meaning of mandated assessment and the role it plays in their practice? How do issues of resource availability, authority structures, assumptive worlds, and accountability relate to local change? What is the relationship of capacity development and equity to assessment? The questionnaire sent to a representative sample of Arizona teachers was the product of six developmental phases. In the first phase, the analysis from the case studies was used to construct items related to (a) local status with respect to change toward ASAP ideals; (b) Resources for Change; (c) Power to Change; (d) Consonant Assumptive Worlds; and (e) Role of Testing. In addition, items were constructed that would indicate the teachers' perceptions of equity issues in relation to mandated testing. Many of these items were statements taken directly from participants in the policy and case studies. Items

were also constructed to measure teachers' knowledge of the curricular content and pedagogy relevant to ASAP ideals, the amount and kinds of relevant professional development they had experienced and the opportunity their students had to learn material and tasks that ASAP measures. We also drew items and ideas from previous studies to enlarge our interpretive framework and provide a basis of comparison across time and sites. Telephone interviews of district and school administrators were conducted. Questionnaires were sent to teachers. Response rates were adequate at the teacher level and very high at the school and district level, so that we had confidence in the generality of the findings. Data were analyzed separately for the survey and synthesized across the various components of the study.

The final source of evidence for the present study mirrored the first. An extensive analysis of documents encompassed new technical reports and reports of assessment results, newspaper articles and press releases, legislation both introduced and passed, State Board of Education agenda and minutes, reports and briefing papers of advocacy groups and ADE advisors, archives of the Academic Summit that drafted standards, and the Standards themselves as considered, revised, and approved by the State Board.

Interviews were conducted with policy actors: the current and former Superintendent and several deputies, officials and staff at the Arizona Department of Education (both current and former), members of the State Board of Education, legislators on the relevant education committees and their staff, officials in the local affiliates of the American Federation of Teachers, National Education Association, School Boards Association, and Administrators Association. In the interviews we attempted to uncover the factual basis of the events around the change in assessment policy as perceived by these policy actors. For example, we asked what had led up to the Superintendent's decision to suspend ASAP Form D and the announcement of the Summit—who was involved and what evidence and argument had contributed to the actions and what reactions were noted afterwards. In addition, we aimed to uncover the intentions, ideologies, interests, and images that guided the work of these actors. We tried, for example, to ascertain whether the Superintendent's decisions were primarily political or primarily instrumental (Edelman, 1985; Rein, 1976)³¹ by

³¹ Rein distinguished political decision-making from rational problem-solving by noting (Rein, 1976, p. 100), "Rationality does not ... imply that some definable logical procedure has been

asking about perceived differences between the old Essential Skills and the new standards, the old ASAP and the plan for the new standards-based testing, by asking about the contributions of technical vs. political advisors and documents, and by asking about the dichotomy between constructivist and basic skills philosophies, as well as concerns for equity, accountability, professional involvement and training, and opportunity to learn. In all, more than 20 interviews with policy actors were conducted during late 1995 and early 1996. We have attempted to preserve the confidentiality of all these actors except for those public figures at the highest level.

In addition to interviews with policy actors, we had access to archives and information from several informants. A legislative intern assigned to the education committee documented the progress of bills related to assessment policy as well as teacher certification, finance, and charter schools that provided a broader political context and revealed the ideologies and actions of legislators and staff, and the sources of influence on their work. Another source of insider information was a videotape of a teleconference between staff and policy actors from Arizona and Delaware, during which issues of assessment policy were discussed. Several participants in the Academic Summit and the subsequent curriculum and technical subcommittees provided extensive time and insight into the standard-setting process and the influences of the ADE and State Board and made internal documents available to us.

Extensive observation supplemented document analysis and interviews. The Academic Summit was observed in all its general sessions. The open sessions of the Language Arts and Workplace Skills design teams were also observed. However, most of the work of the design teams was conducted privately, so that our only source of data on their work comes from the informants. We also observed and recorded four of the thirteen open meetings conducted to present the initial drafts of the Standards. In addition, we observed five meetings of the State Board of Education as they discussed and voted on the approval of the Standards and other aspects of assessment policy. Extensive hand-written notes were taken, and some meetings were tape recorded as well.

followed which has exhaustively scrutinized all possible options or considered all relevant information. Rather, it suggest that, at the least, the process of making a decision made use of whatever resources of knowledge, judgment, imagination and analysis were available in the circumstances.... To disregard real, scientifically discoverable risks is irrational.”

The ideas of Martin Rein (1976) influenced us as we amassed and addressed these accumulated data. He distinguished the consensual from the contentious and paradigm-challenging approaches to empirical policy research. The consensual approach (p. 126), "Proceeds from agreed-upon aims; it asks whether policies and the specific programmes that implement them, work as intended." In the contentious approach, the researcher acts as a "moral witness" or "social critic" about government's aims, actions, or nonactions regarding social needs. Rein noted the limitations of positivist science in policy research, in particular the aspiration to produce generalizable, objective, value-free and definitive propositions. Instead the aim of the policy researcher should be to develop stories that weave together values, facts, images, and tentative explanations with the setting and characters—the local context. The test of the story include the verification of its facts and the coherence of facts and explanations. The researchers, though never completely free of values and biases, can nevertheless strive to restrain and postpone their intrusion into the story and seek review and reaction from others to examine the coherence and plausibility of the story.

The data analysis for this study began with three readings of the data. Next, the data were arranged in an event chronology. A time by policy actor matrix was then developed, taking as major categories as those constituencies that had a stake in assessment policy. Examining the data within these categories suggested a narrative line with characters, settings, a plot and perhaps a moral. We tried as much as possible to represent the voices of the actors themselves and to quote extensively from their public pronouncements and reports. The major difficulty in this process was finding an end to the story, while events in Arizona assessment policy continue unfolding. We are mindful that plot lines are constructions, whereas elements of real life continue on, without inherent beginning, climax, and resolution. Three vignettes were written to portray the sense of spectacle (Edelman, 1988) that permeated the changes in assessment policy. Vignettes are artful and condensed reconstructions of actual data that particularize assertions from empirical research (Erickson, 1986; Rein, 1976).

Having written the narrative that tied the pieces of the ASAP story together, we then engaged in a process of internal, structural corroboration. That is, we sought to challenge the facts of the story, accepting only those that could be corroborated by more than one source. We worked reflexively to make sure that facts as we believe them and the explanations that we spun out fit one another.

Finally, we have subjected the report to review by two informants we believe are sympathetic to the point of view and one who is impartial.

Like the politics and policies we observe, we recognize that our own actions as researchers, our values, definitions, and categories—even our sense of what is a fact, are open to critique. We have given our best efforts to strive for a complete and faithful portrayal of what we have seen and understood and to avoid creating heroes and villains. Much of what we learned was new and surprising. But in at least four respects we brought in prior descriptive and normative categories. First, we have stated elsewhere our belief that the intervention model the state employed in ASAP (based on the Field of Dreams assumption) was naive and self-defeating (Noble & Smith, 1994). We cling to the notion that some intervening processes have to occur if a mandated test is to have any instrumental effects, and that educators must as interpretive beings come to grips with the policy, understand it, learn how to change toward it, and reflect on their actions.

Second, we have tried to raise issues of social equity in relation to policy from the background, even though they are seldom mentioned by policy actors in Arizona. Third, we have been convinced by experience and evidence that high-stakes testing has both intentional and unintentional effects, some deleterious to the processes it is supposed to address (Smith & Rottenberg, 1991). Finally, even with all its faults, there is a professional knowledge base connected to testing. Without being a slave to its technical minutia, assessment policy makers ought to have at least passing familiarity with it or develop consultative relationships, not with the test developers alone (who themselves must be considered policy actors), but with relevant and thoughtful experts (among whom we do not count ourselves). When policy makers ignore technical issues, they confirm that it is political spectacle that we have before us, and the instrumental aims are a kabuki mask.

We have been immersed in research on ASAP for as long as ASAP itself has lasted. Our study of its consequences, though comprehensive in a conventional model of policy research, failed to tell the whole story or adequately explain what happened. We offer this report to round out the picture and to suggest that there are no policies without politics, and no sensible policy research without political analysis.