**Reforming Schools by Reforming Assessment:
Consequences of the
Arizona Student Assessment Program (ASAP):
Equity and Teacher Capacity Building**

CSE Technical Report 425

Mary Lee Smith
with Audrey J. Noble, Walter Heinecke, Margo Seck,
Carol Parish, Marilyn Cabay, Sue C. Junker, Susan Haag,
Kim Tayler, Yvonne Safran, Yolanda Penley,
and Amy Bradshaw

CRESST/Arizona State University

March 1997

REFORMING SCHOOLS BY REFORMING ASSESSMENT: CONSEQUENCES OF THE ARIZONA STUDENT ASSESSMENT PROGRAM

Mary Lee Smith

Assisted by

**Audrey J. Noble, Walter Heinecke, Margo Seck,
Carol Parish, Marilyn Cabay, Sue C. Junker, Susan Haag,
Kim Tayler, Yvonne Safran, Yolanda Penley
and Amy Bradshaw**

ARIZONA STATE UNIVERSITY

SOUTHWEST EDUCATIONAL POLICY STUDIES

1996

CONTENTS

APPENDICES

CHAPTER 1
INTRODUCTION

No other issue has occupied the discourse of educational policy so much as the idea that American public schools are failing and that such underachievement threatens the economic health of the nation. And no other solution is proposed so often as raising academic standards. The proposals of the 1989 Charlottesville Conference embodied this preoccupation and led to Goals 2000, the putative federal policy on educational reform.

The heart of this policy agenda, known as standards-based reform, can be found in the idea that ambitious standards and goals, promulgated from the federal government, will point American schools toward curricular content more rigorous than what they now pursue.

Rapidly and inevitably following calls for standards are calls for assessments to measure achievement of them. The assumption that ties these proposals together is that without assessments, the public will not be able to gauge the progress schools make toward the national goals. Furthermore, without accountability for test scores, neither school personnel nor the students themselves will be motivated to improve. Beyond this notion that schools will act to avoid punishment (public embarrassment, loss of incentive funding, loss of clientele, or takeovers), the reform agenda is silent on the means the schools should use to accomplish the standards. Implicit in the policy is the hypothesis: Since you get what you test, and you don't get what you don't test, if a government agency mandates a test that mirrors the type and level of instruction you want, then teachers will do whatever it takes to adapt their practice to it.

Coinciding with the standards-and-assessment-driven reform agenda were three related agendas in the educational policy arena. First, a largely voluntary effort by professional curriculum specialists resulted in the development of sets of curriculum standards in subject matter areas. The best known, math standards, were developed by the National Council of Teachers of Mathematics (NCTM), a group of expert teachers and professors of math and math education. They proposed a revolution in math teaching that focused on real-world problem finding

and problem solving, for example, and based their recommendations on a constructivist theory of learning.

Professional standards were also developed in other curricular areas, using NCTM as a model and also subscribing to constructivist principles. Constructivist theory assumes that students construct their own knowledge (rather than passively receiving knowledge transmitted by schools) out of intentional transactions with materials, teachers, and other pupils. Learning is more likely to happen when students can choose and become actively engaged in the tasks and materials, and when they can make their own connections across subject matter on tasks that are authentic and organized around themes. According to this theory, literacy is whole, embodying reading authentic texts and writing as a way of unifying all the subjects. For example, to be literate is to be able to explain the reasoning one uses to discover and solve math problems. Explicit in constructivist theory is the rejection of the pedagogy of worksheets and the exclusive reliance on phonics, spelling out of context, computation, isolated subject matter and the like.

The second agenda that coincided with the movement to improve schools by imposing official standards was the movement in the psychometric community to reform tests. Long dissatisfied with traditional, standardized testing, a subset of testing specialists set out to develop alternative assessments to correspond to how teachers teach and pupils learn. They foresaw models of assessment that were integrated with instruction. They proposed portfolio assessment and tasks that required students to read and write in authentic and interesting contexts and recommended rubrics for teachers to use in scoring responses to these tasks.

The third agenda was that of a loose confederation of reformers, largely part of the professional community of educators, who developed non-governmental associations and coalitions to improve school organization, teaching and learning, and teacher education. These reformers tended to focus on means as well as ends, enhancing professionalism, empowering teachers, and the like.

Although the three reform agendas described above were quite influential and no doubt intersected with government policy, they are not the focus of the present paper, which instead examines government mandates to change education through standards and assessments.

While budgetary and political problems at the federal level delayed the implementation of Goals 2000, several states (e.g., California, Vermont, and

Kentucky) progressed along separate paths toward reform of schools by their own programs of mandated standards and assessments. Arizona became one of the states ahead of the federal policy when it instituted the Arizona Student Assessment Program (ASAP).

Although ASAP has since been suspended, its four-year existence may serve as a microcosm for states and the federal efforts toward the same ends. For this reason, the consequences of the ASAP may be significant as a kind of natural experiment for what course the standards- and assessments-based reform may take.

The purpose of the research described in this report is to understand what happens in the aftermath of a change in state assessment policy that is designed to improve schools and make them more accountable to a set of common standards. Although theoretical and rhetorical works about this issue are common in the literature, empirical evidence is novel and scant. This research comprises several component studies that were conducted over a four-year period, coincident with the life of ASAP itself. A variety of research methods and approaches were used, and several interim reports based on the component studies have already been completed. This report represents an integration of evidence from all the component studies aimed to present a grounded understanding of what happened in the aftermath of change to ASAP, and how these things might have come about.

Problem Statement

What follows from the introduction of a policy designed to produce both accountability and instructional change? If the program worked as intended, what would be optimal consequences from the state reform point of view? In time, schools would produce higher achievement by focusing their attention on instruction and curriculum with the greatest likelihood of accomplishing the state curriculum frameworks (in Arizona, the Essential Skills) and would adapt toward the kinds of instructional practices suggested by the form of assessment. Embodied in the frameworks is the vision that the Essential Skills represent ambitious academic standards. It is assumed that higher order thinking and problem-solving skills represent more valuable criteria for school work and are important for the state's economic well being. If the policy is to be effective, one would see, during the early years of the policy, practitioners knowing about and

endorsing the intentions and instruments of the policy and building capacity toward and adopting practice that is coherent with the policy. One would see system administrators moving in similar directions, providing for and procuring resources for capacity development, and removing obstacles (local policies and practices that are incoherent with the policy). One would see capacity building at the system level as well, as the state education department advances its technical expertise and ensures the quality, efficiency, fairness, and continuity of the system.

The instruments (in McDonnell and Elmore's [1987] sense of the term) that drive the policy forward encompass (a) the Arizona Essential Skills; (b) the tests themselves (which suggest that integrated subject matter and lifelike problem solving ought to be the aim of instruction); (c) the associated assessment system (decisions about how to administer and score the assessments, whom to exempt, and the like); and (d) the accountability system that indicates who is to be held responsible for what performance and on what criteria. In the latter respect, the District Assessment Plan and Essential Skills Reporting Documents (see below) make the districts accountable for student mastery of the Arizona Essential Skills; the report cards make districts and schools accountable to parents and the public (assuming that educators will act to overcome low scores reported in the paper and that parents will pressure the schools to raise achievement as expressed in student report cards, or that parents will choose schools with high achievement). And (eventually), the prospect of not attaining one's diploma is thought to motivate students to work harder on standards- and assessment-related school work.

Description and Brief History of ASAP

Although by far the most visible component, the ASAP performance assessment was only one part of the seven-part Arizona Student Assessment Program. Legislation enacted this program in 1990; the Arizona Department of Education (ADE) was responsible for developing and implementing the program. ASAP consisted of seven parts:

1. Arizona Essential Skills. The state curriculum frameworks in reading, writing, mathematics, science, social studies, health, foreign language, music, and 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 and be able to do 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, under standard conditions and scored under auspices of ADE, using the "generic rubric" or scoring guide. ASAP Forms A, B, and C consisted of performance assessments in reading, writing, and math (with assessments in science and social studies to be phased in later), to be used for preparing pupils to take the Form D, or as instructional packets (and see item 4).

3. Norm-referenced testing at Grades 4, 7, and 10. The Iowa Tests of Basic Skills and Tests of Achievement and Proficiency provided a means for comparing the achievement of Arizona schools with that of a national norming sample.

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 students demonstrated 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 their DAP testing: ASAP performance assessment Forms A, B, or C, a system of portfolio assessments, or criterion-referenced measures. Either of the latter was 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 to report results of achievement testing and nontest indicators.

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 nontest

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.

Analysis of state documents (Olivares-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.

Before the advent of ASAP, Arizona had operated under a mandate to test in the spring of each year all school pupils in Grades 2-12 in reading, math, and language arts, on both standardized, norm-referenced tests and the continuous uniform evaluation system (district-based, standardized, objectives-referenced tests of the Arizona Essential Skills). At that time, there was considerable opposition to standardized testing. The Center for Effective Student Evaluation, a nonprofit organization of educators opposed to standardized testing, had successfully spearheaded legislation to remove first graders from the state testing mandate (Iowa Tests of Basic Skills). There were also critics of standardized testing at the Arizona Department of Education. C. Diane Bishop, a former high school math teacher, had been elected as superintendent of public instruction and head of ADE. Her administration included individuals who were outspoken and effective advocates for "authentic assessment," that is, assessments that fit what teachers do in classrooms, and curriculum that was more holistic and aimed toward higher order thinking and problem solving. The Arizona Department of Education had contracted with Tom Haladyna and associates of Arizona State University-West to do an evaluation of the ITBS and the Tests of Achievement and Proficiency. This group concluded that existing tests covered only 26% of the Arizona Essential Skills and confirmed the widespread discontent among educators toward the existing mandate. In 1990 this coalition mounted a campaign to convince educators to support a revision of assessment, because they believed that what gets tested is what gets taught, and that teachers would revise their methods and schools their curriculum if the state renounced

standardized testing in favor of performance testing. They also assumed that educators would play key roles in the planning, development, and monitoring of the testing program (their involvement would then spur professional and curriculum development by districts and teachers). This coalition of professionals combined with legislators and policy makers who were motivated by the belief that schools were not currently accountable to the state and needed more testing to focus attention more sharply on the Arizona Essential Skills. The result was successful legislative campaign.

Arizona Revised Statute 15-741, the enabling legislation for ASAP, took effect in July 1991. The ASAP program was then implemented by the ADE. The contract for test construction was let to Riverside, the publisher of the ITBS. Subsequently, contracts for developing scoring rubrics, and for the scoring itself, were let to Riverside, Measurement Inc., and other organizations. Although the ADE conducted many workshops and made many presentations to educators about the testing program, they provided no professional development in how to teach in ways that the performance assessment suggested. Teacher training was thus left to the districts, some of them quite able and willing and others with little knowledge, resources, or commitments to respond.

After a remarkably short development period, a pilot administration of the performance assessment was conducted in March 1992 with results of a technical analysis reported in September 1992. The form administered was Form A, which consists of a series of items that call for students to construct responses to questions *within* the content areas of reading, math, and writing. Riverside reported levels of reliability and validity for Form A that were acceptable for school-level inferences, but too low (according to their own cautions) to support inferences about individual pupils.

ASAP Form D-1 of the performance assessment was administered in March 1993. Form D differed from Form A in that the task that D entailed was integrated across reading, writing, and math. The scores of Form D, however, were disaggregated by content area. Form D-2, which offered students a different set of integrated tasks than appeared in D-1, was administered in March 1994. Although a sample of students took D-1 in 1994, there was no analysis to determine whether they were parallel. Each iteration of Form D purported to measure a different cluster of Essential Skills.

High-stakes consequences of test scores were to be phased in gradually, with the high school entering class of 1996 to be required to pass ASAP tests in order to graduate.

In March of 1995, a newly elected superintendent of public instruction, Lisa Graham, announced that ASAP D-3 would not be administered as scheduled, and that the program would be revised. The Essential Skills were to be replaced by a new set of standards. The superintendent based her decision on the technical inadequacies of ASAP Form D and the need for standards that were more ambitious, clear, and measurable than were the Arizona Essential Skills. "ASAP" was to be retained in its accountability aspects, but renamed the Arizona Student Achievement Program, its performance assessment aspects to be downplayed.

As of the completion of the present report, the new standards have not yet been approved nor have the specifications for the new state testing been released. In spring of 1996, the Arizona legislature approved a bill that mandates norm-referenced, standardized tests in each grade level, 3-12. The history of these changes is detailed elsewhere (Smith, Heinecke, & Noble, in preparation).

This report is organized by chapters. Methodology of the research is described and justified in chapter 2. Chapter 3 presents the findings and assertions of the study. Chapter 4 comprises the discussion of findings and refers to related literature.

CHAPTER 2
METHODOLOGY

Justification for Multimethodology

Because we believe that no single method of data collection produces a complete or unequivocal picture of the phenomenon, we have chosen to use multiple methods and approaches, which give many lines of sight toward understanding the reform. Qualitative approaches are necessary to understand how educators are defining and coming to terms with the reform and to look closely at what they are actually doing about it. At the same time, it is helpful to be able, through survey techniques, to gauge the range of beliefs and practices subsequent to the implementation of the reform. Each research approach and method has inherent strengths and weaknesses. Each contains certain assumptions, and each supports different kinds of inference. The strength of the analysis is the linking of data from the whole. This chapter of the report is organized in the following sections: methods (design, data collection, data analysis) of the multiple case study project, methods of the extension group study, methods of the educator surveys, supporting studies, and methods of integrating data from the separate studies.

Data Collection and Analysis of Data Streams

Multiple Case Study

To understand the meanings and actions of educators as they encounter the ASAP requires the kind of knowledge that can only be gained from direct contact between researcher and participants. The research design chosen to address this issue is the multiple case study design (Miles & Huberman, 1994). This design is based on the rationale that understanding complex organizations such as schools requires long-term and close-up examination of local practice within bounded social settings. The actions of participants faced with a new government mandate can only be understood in the specific context in which they occur and referenced to the meanings held by those participants. The researcher aiming to understand

these meanings must have access, over an extended period of time, to the classrooms and offices in which participants' definitions of the situation (mandated assessments, in this case) evolve and get worked out in actions. Do participants actually provide the type of instruction geared to the ASAP performance test? Do they have the knowledge they need to adapt, or do they have the intention to do so? What is the meaning of the ASAP to teachers and others in schools? Getting evidence to answer questions such as these requires more than snapshot observations and standard questionnaire items. Thus, the qualitative case study is the best design. The decision to do more than one case study was not made because four is closer to the population of schools than one. Nor is there any intent to evaluate the four schools comparatively. The rationale for drawing multiple cases is that one case provides interpretive context for the others. A case study researcher typically immerses herself in a single site and tries to understand everything there is to know about it. Holistic understanding, however, sometimes produces the holistic fallacy. In looking at one case at a time, the phenomena and events one notices may be mistakenly seen as causal. Seeing two case studies in parallel can alert the two researchers of features taken for granted or overlooked in one. In one of the cases in the present study, for example, the influence of the particular district's philosophical support of ASAP was overlooked by the researcher in her within-case analysis. Simply because that support was taken for granted by everyone in the site, she failed to make proper note of the potential influence of this condition. Yet when her case was held up against a second case, one in which the district administration was not supportive of the mandate, the importance of the factor in explaining the relative success of the mandate in the two sites became obvious.

Four cases were chosen for the study. The number was determined by the resources available to support four graduate students for the year. Only elementary schools were chosen, because of the need to contrast the effects of the new mandate with the previous one studied by Smith, Edelsky, Draper, Rottenberg, and Cherland (1989). The decision of which sites to select was made based on the desirability of varying cases across levels of economic and social resource and prior history of testing demand (the importance of test results historically in the district). Thus we tried to find schools with greater and lesser economic resources, serving advantaged and disadvantaged students, and schools located in urban, rural, and suburban settings. In addition, we made use of

contacts and acquaintances that would help us to access schools and districts that fit the characteristics of our purposive sample design.

All schools we contacted and requested permission to study responded positively. The four sites where we conducted case studies were (a) Valor, a rural school with low resource base, serving mostly poor and minority pupils in a K-8 district; (b) Franklin, an urban school with a relatively high resource base, serving mostly poor and minority pupils in a K-8 district; (c) Pine, a suburban school with an ethnically and economically diverse student body, in a large, K-8, resource-advantaged district with high test demand characteristics; and (d) Hilldale, a suburban school serving mostly Anglo and advantaged pupils, in a large, K-12, resource-advantaged district with moderate test demand characteristics. Additional information on the descriptive characteristics of the four sites is available in the case studies themselves and summarized in the Cross-Site Data Matrix (Appendix A). All names used in the study are pseudonyms. District and school personnel were promised confidentiality.

Four researchers at Arizona State University were selected to conduct the case studies. At the time, Audrey Noble, assigned to Valor, was a fourth-year graduate student in the doctoral program in educational leadership and policy studies. In addition to her case study, she acted as research coordinator for the others. Suzii Junker, a third-year student in the doctoral program in reading, conducted the study at Hilldale. Walt Heinecke, a third-year student in the doctoral program in educational leadership and policy studies, studied Pine. Marilyn Cabay and Yvonne Saffron collaborated on the study at Franklin. Both were fourth year-students in the doctoral program in school psychology. All five of the researchers had at least two courses in qualitative research at the time of the study and had produced independent research studies as part of their degree programs. All are highly experienced in various educational roles: classroom teacher, counselor, school administrator, school psychologist, testing coordinator. All five brought unique perspectives to their research role; yet consistency across researcher perspectives was maintained in several ways. First, a common design for data collection and common definitions of researcher roles were shared. Second, the theoretical framework focused researchers' attention on common aspects of the sites (the images held by the participants of pupil, teacher, learning, curriculum, assessment, and school structure). Third, monthly meetings of the researchers were held to address issues raised and problems at the separate sites,

share memos and working papers, and the like. Fourth, the work of the researchers was supervised by Mary Lee Smith, who monitored the adequacy of data collection and analysis procedures. Finally, drafts of the four case studies were read by all members of the research team, and reactions were incorporated into the case studies by the researchers to add to the overall fit of the cases together and to provide the interpretive context of each case in relation to the others.

Each case study involved the following data collection methods. The unit of study was defined as the classroom within the school. The four participating schools provided the researchers with access to faculty meetings and other school events, direct observation of one third-grade and one fourth-grade class (except for Hilldale Elementary, in which a combined third/fourth-grade class was the primary participating site), interviews with third-grade and fourth-grade teachers, and documents relevant to ASAP, curricula, and local testing programs. This access extended through the academic year 1992-93. Informal contact between researchers and participating teachers was maintained through 1993. The choice of third- and fourth-grade classes was based on the state-mandate of ITBS testing in fourth grade during the month of October and ASAP performance testing in third grade in March. The design of observations followed from this schedule, with observation occasions clustered in the fourth-grade classes in the fall and the third-grade classes in the spring. The working design called for researchers to be in the targeted classrooms one day each week normally and twice per week immediately before, during, and after the testing events. They deviated from the schedule when necessary to capture activities relevant to the research questions in the rest of the school or district. For example, the researcher at Hilldale accompanied the teacher whose class she usually observed when she attended a training session on scoring of the performance test. The researcher at Valor branched out to classes other than the one chosen in the design so that she could understand the relative authority of teachers, principal, and district officials in determining curriculum choices.

The researchers played the role of "more-observer-than-participant," developing cordial, non-evaluative, and trusting relationships with the teachers and school staff. No problems with access were experienced at the schools over the year's data collection. However, project policy about confidentiality and ownership of the data had to be clarified and reiterated with officials in one of the

districts. Our position was to maintain confidentiality and protection of the identity and perspectives of the participants with whom we dealt most directly; the teachers and principals. District officials would have access to only those data that shielded the identity of the participants or that the participants cleared for publication.

Observation occasions of school and classroom activities were aimed at understanding the role of testing in context, the meaning of mandated testing to teachers and school staff, test preparation for mandated tests, and the relationship of mandated testing to curriculum, pedagogy, and school structures. The conceptual framework of the study provided the focus for observations. That is, the researchers kept in mind the need to attend to, besides the normal, everyday life of the classrooms, incidents that shed light on the images held by participants of pupil, teacher, learning, assessment, and school structure. Researchers kept detailed notes of what they observed, transcribed their working notes, and submitted the write-ups on text files to the research coordinator. These were reviewed periodically to make sure the researchers were preserving the necessary level of concrete detail and recording material relevant to the research questions and conceptual framework. Monthly meetings of the researchers were held to coordinate insights and keep everyone on target.

By design, the researchers conducted formal interviews with the principal and teachers whom they observed and focus-group interviews with remaining third- and fourth-grade teachers in the school. In addition, interviews with district officials were conducted to understand the district perspectives on assessment and the organizational climate of the districts. The interview agenda and key questions and probes were developed by the research director and coordinator to generate data according to the conceptual framework. For example, teachers were asked questions such as the following: "The state believes that the new testing program will promote a new kind of instruction. Other than knowing what the test covers and how to administer it, what are the things a teacher needs to know to teach in the manner that ASAP promotes?" Since these interviews fit a qualitative approach to research, the exact wording and sequence of questions varied. It was more important to elicit the meanings the assessment has for participants than to standardize questions. The interviewees were encouraged to tell their own stories in their own words; the researchers used those words to construct probes so that the agenda could be addressed. For example, the probe

for the question stated above might attempt to elicit information on the kinds and amounts of professional development the teachers had already experienced or believed to be important precursors of ASAP-related instruction. The agenda was drawn from the conceptual framework and emerging issues in the study as a whole. Interviews were tape recorded and the tapes transcribed.

Researchers at the four sites also collected documents and artifacts. For example, some teachers voluntarily provided work samples from students in ASAP-related activities and journals in which students described their reaction to assessments. Curriculum guides, text samples, work sheets and instructional packets, detailed samples of district tests and test results, information sent to parents, notices of meetings and training sessions, and the like also supplemented the observation and interview data.

Within-site data analysis. The four researchers coded their data according to the categories in the project conceptual framework as well as categories emerging from their site. For example, every instance of data that plausibly referred to or illustrated a teacher's image of the curriculum was so coded for subsequent retrieval. A district administrator's contention that district CRTs were a more appropriate standard for achievement than ASAP results would have been coded as "Image of testing." In addition, local issues were coded, such as the conflict among third- and fourth-grade teachers at Franklin about the value of moving to ASAP-like instruction, which produced the inductively-derived category "Grade-level isolation/conflict." Researchers were encouraged to use qualitative analysis computer programs, such as Ethnograph and Hyperqual, to identify, mark, index, and retrieve data that instantiated the categories. They wrote memos periodically to define the categories and document their thinking processes as they analyzed their data. Finally, they wrote assertions and produced vignettes to support the assertions. According to Erickson (1986), assertions are statements that researchers inductively derive by reading and re-reading the record and data. These statements are inferences about the meaning of the evidence. For example, one of the assertions from the study of the Valor site follows: "Although performance assessment is meant to encourage the social nature of learning, learned attitudes and behaviors (prior knowledge) regarding testing persist. Teachers and students respond to the function of assessment rather than the form. Testing for teachers and students remains a solitary, inactive, and structured experience."

Researchers established the warrant for their assertions by looking closely for disconfirming instances, and checking that the assertions had sufficient confirming data of varying methods (e.g., observations vs. interviews). In addition, drafts of the assertions and vignettes of each case study were read by the other researchers, the coordinator, and the director. Revisions were made based on this feedback. Then, the researchers completed the case studies (Smith, Noble, Cabay, et al., 1994), providing their overall perspective about the role of mandated testing in their respective sites.

Cross-Site data analysis. The existence and use of the conceptual framework for the study as a whole, the monthly meetings, and supervision of researchers increased the likelihood that the separate case studies would have enough elements in common to enable cross-site analysis. The final meeting of the research team to discuss the case studies was tape-recorded to preserve a record of the ideas generated. This meeting served two analytic purposes. First, each case was used as interpretive context for the others. That is, elements that had been overlooked in one site became highlighted by comparing cases. For example, at Hilldale, district testing was simply not an issue, and the researcher at that site had consequently ignored it. At Franklin and Pine, however, the district testing program had profound impact on what happened to ASAP-relevant instruction. Through this comparison, a hole in the Hilldale account was readily identified and rectified. Second, in this meeting the researchers were treated as informants in the sense that, after a year of data collection, they "knew" much more about the educational and social context than they could have possibly included in the case study. The director and coordinator could then ask them to summarize information on issues of cross-site interest. For example, a quick reading of data and a few phone calls produced data on the missing element from the Hilldale account on the role of district testing.

The analysis of qualitative data is fundamentally a process of thinking and progressive problem solving (Erickson, 1986), with only a crude set of tools and procedures. The conceptual framework yielded categories such as the Image of the Pupil. Data had been gathered that allowed us to generate assertions within each site about the Image of the Pupil that seemed to be held by teachers and district officials. In addition, we had evolved a set of working hypotheses, or plausible accounts and explanations, for how the change in mandated testing was working out at each site, that is, what particular barriers and facilitating conditions

seemed to be responsible for local reactions. Furthermore, we understood that audiences for this report would be interested in the formal characteristics of each site (e.g., the degree of pupil disadvantage) and would need a variety of information to make their own interpretations of the data. From these considerations, we developed a set of dimensions for the cross-site matrix. Our aim was to provide data in the matrix that would reduce the sheer quantity of information to a manageable level without resorting to high-level abstractions or losing the sense of grounding and authenticity that case studies can provide.

Based on the above considerations, the Cross-Site Data Matrix was constructed. The elements in each cell are short summaries, paraphrases, or characterizations of the particular site on the selected dimensions. These characterizations were constructed by the research director and submitted to the case study researchers for their substantive and editorial comments. The Cross-Site Data Matrix is placed in Appendix A.

The process of arraying data in the Cross-Site Data Matrix stimulated further thinking about what elements were most salient in accounting for the differences among the cases in response to the mandate. In constructing the Analytic Matrix, we started with a working assumption (analyzed and critiqued in Noble, 1994) that the ASAP mandate promotes changes toward high standards and constructivist education. Furthermore, we knew from the findings of the policy study (Noble, 1994) that the state had made no provision for capacity building in support of the mandate. Nor had the state attended to issues such as delivery standards or opportunity to learn. Thus, this mandate was unfunded, and professional development provided by the state in support of change was meager or nonexistent. The only state mechanisms to instigate the change included the power of ADE to persuade through rhetoric (e.g., repeated reminders to district officials and teachers of the importance of the Arizona Essential Skills and of teaching "the way kids learn"), the threat of disapproval of the District Assessment Plans, and the performance test itself (which was initially perceived to be low-stakes), plus the preliminary Forms A, B, and C and workshops to train teachers how to administer and score the assessment. Therefore, we recognized that both the resources for changing toward the promoted goals and the authority and power to change had to be understood at the local rather than state level. Based on these assumptions and understandings, we chose four categories that seemed to account for the status of the site at the end of the first year. For

example, the curriculum and pedagogy at Valor was virtually unchanged after one year of the program. No resources were available to direct toward ASAP-consonant activities, and thus no capacity was developed. School personnel acquiesced to the ASAP requirements, and ASAP merely added to the accountability load. Some resistance was evident in the departure of one of the constructivist teachers who experienced this burden. The status of change can be attributed in part to resources issues, knowledge, assumptive worlds, and organizational culture there.

The categories in the Analytic Matrix (Appendix B) are listed and defined as follows.

- **Resources for Change:** *Material resources* refers to the district's financial capacity to purchase or develop curriculum and to offer teachers professional development activities consistent with ASAP goals. Where financial resources are available, we ask whether they are directed at activities consonant or dissonant with ASAP aims. *Knowledge resources* refers to the presence in the district and school of officials and teachers with knowledge of and commitment to constructivist education and performance testing. Each site was characterized according to whether there was some gatekeeper, such as a principal, coordinator, consultant, or other person, who could interpret ASAP procedures and help teachers make changes consistent with ASAP aims. In some sites, a coordinator had been named by the district, but the person lacked knowledge, was unavailable to teachers, or soon left the district, and thus failed to help teachers make consonant changes.

- **Power to Change:** We characterized each site according to its organizational culture and where the power exists to make changes at the classroom level. For example, a centralized and hierarchical district vests control over change at the district level, leaving teachers and principals with little discretion to change in contrary directions. Local options remaining include acquiescence, accommodations (e.g., dis-integrating integrated curriculum or inappropriate test preparation), resistance, and marginalization.

- **Assumptive Worlds:** In this category we condensed the images of the pupil, learning, teacher, and curriculum that seemed to characterize both the district and the teachers at each site and the extent to which the dominant philosophy was either consonant (i.e., constructivist) or dissonant (behaviorist or concrete-sequential) with ASAP aims (assuming that ASAP is in fact constructivist). The constructivist assumptive world views the pupil as an active meaning-maker, the teacher as a coach or partner in meaning-making, and the curriculum as thematic, integrated, and negotiated, consistent with pupil interests and prior knowledge. The concrete-sequential assumptive world views the pupil as an empty

receptacle, the teacher as conduit of curriculum and imparter of skills, and the curriculum as a hierarchical set of standard skills for the pupils to master.

- **Role of Testing:** This category reflects our characterization of the test demand or degree of testing stakes imposed on classrooms at each site. We distinguish (a) the perceived function of tests as accountability devices (performed for external audiences) rather than as integral parts of instruction and (b) whether there is a strong demand for high scores or measured change at the site. We also note the degree of test burden (proportion of time consumed by various testing functions), the expectations at the site for high or low scores based on past history, and where ASAP fits into the testing scheme.

- **Year-End Status:** This row in the matrix reflects our perspective of where each site stood with respect to reactions to the ASAP mandate at the end of one year of the reform.

Extension Study

The nature of the ASAP program as mandate without provision for capacity building suggests that accommodation to the program might logically be slow. Because of this delay, a study of first-year consequences alone would be suspect. The primary purpose of this data stream was to return to those schools that had participated in the multiple case study design to determine what had happened by the end of the second year of participation. What was the status of those particular schools at the end of the second year, as interviews with educators can reveal? The secondary purpose of this data stream was to learn the status after the second year of the ASAP experiment of other schools similar to those of our case study schools. Lacking the resources to conduct a second round of case studies, we chose the more efficient design of group interviews with educators at both sets of schools. A more deductive design than typifies qualitative research is justified when most of the questions and hypotheses are already known and when the methods are relatively standard from case to case (Miles & Huberman, 1994).

Methods of data collection and analysis. Principals and focal teachers at each of the four original case study sites were contacted by the researcher who conducted the case study there. We arranged for a group interview with the focal teachers and other interested staff. Each participating school was paid a small honorarium of $50. The interviews, ranging from one to two hours, were conducted by the original researchers, using a common interview agenda.

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. We have some hypotheses based on our earlier studies, but every school is different. We want to know what has happened here, if anything. Reasonable people have different views on ASAP. We have no stake in it one way or the other. We just want to know your experience. After two years, what is your impression of ASAP? How does it fit into what you are trying to do here? 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 (changes in textbooks, materials, local tests, adding to the accountability demands or time needed for testing, changes in how teachers think about teaching and learning, etc., reactions both positive and negative or null)? What if anything has gone on in this school or district in terms of helping teachers teach more holistically (consultants, in-service, collaboration, etc.)? Are there people at this school available to help you to adapt instruction in ways consistent with ASAP? 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 interviewers were directed to probe for specificity, to maintain neutrality, and to try to cover the agenda. After the interview, they reviewed the tape and attempted to fill in the Analytic Matrix for their school. The researchers had been trained in the matrix and interview protocol and supervised by Smith. Interviews were tape recorded and transcribed. Responses were arrayed according to the Analytic Matrix. In particular, the analysis looked for ways that Year 1 and Year 2 effects were similar or different and how the original cases were similar to or different from the new cases.

In addition to the four original cases, we interviewed educators at five other sites: one rural and one urban school serving primarily disadvantaged pupils, one suburban school with high resources, advantaged pupils, and pedagogy predisposed to constructivism, and two schools in a district with centralized

organization and an emphasis on using test scores for accountability. The latter two, Hamilton and Jefferson, were the subjects of an earlier study of the role of mandated standardized tests (Smith et al., 1989).

Educator Survey

The purpose of the educator survey data stream was twofold. First, we aimed to understand ASAP in relation to capacity and equity. Second, we aimed to gauge the effects of the program after two years of implementation, extending this analysis to a more general population than one can access in a qualitative design.

Research questions. 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?

Questionnaire development. 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 multiple case study 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 multiple case studies. For example, policy makers stated on several occasions that ASAP represents "the best we know about how children learn." One item asks that teachers state their relative agreement or disagreement with that statement. 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 have 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. For example, Madaus, West, Harmon, Lomax, and Viator (1992) measured the time spent in test preparation, the accountability pressures and the tendency to spend time on remediation of skills rather than holistic education. They found that

schools with different rates of minority pupils differed reliably on these dimensions. We included related items that would measure time spent on testing, extent of test preparation, and adherence to a basic skills pedagogy so that we could compare teachers at schools with different ethnic minority composition. From Haladyna's study (Nolen, Haladyna, & Haas, 1989) of Arizona teachers' reactions to an earlier testing mandate, we adapted items related to perceived validity and utility of the new test mandate. Corbett and Wilson's (1991) study of state mandated testing yielded items related to impact of the program and the uses of testing. In subsequent stages of instrument development, some of the borrowed items were reworded, merged, or eliminated. Nevertheless, these sources were valuable to us in the initial conceptualization of the instrument.

The instrument prototype was a self-administered questionnaire consisting primarily of statements to which teachers would respond on a 5-point scale; for example, *strongly agree* to *strongly disagree*, or *very true of this school* to *very untrue of this school*. A *don't know/no opinion* option was provided for each statement. Some other items were open-ended. Others asked for information about the professional development teachers had received as a result of ASAP.

In the second phase, the instrument prototype was reviewed by a panel of experts: teachers, administrators, university faculty, and researchers who had conducted the multiple case study described above. Two professional survey research methodologists also participated in the review. The expert reviewers were paid an honorarium for their participation. Members of the panel included Native Americans, African Americans, Mexican Americans, and European Americans, both males and females, teachers of elementary, middle, and high schools, teachers with various levels of experience with ASAP, and ASAP school and district coordinators. Panel members were asked to judge the clarity of the items and identify any ambiguities in wording. They were asked to critique the items as a whole for any indication of researcher bias, inadequate coverage (given the intent of the study), insensitivity, or burden.

In Phase 3, we revised the instrument based on the feedback of the panel of experts, modifying or deleting items panel members felt were loaded, ambiguous, or insensitive to cultural and ethnic groups. We also added items on issues that panel members felt were inadequately covered. One instance of this was the addition of items on cultural bias and gender fairness.

In Phase 4 we conducted a pilot administration of the revised instrument, asking the participants to respond as if they were members of the survey sample. In addition, we asked them to report the time required to complete the survey. The reported average time they took was 17 minutes. They were also asked to critique the burden, coverage, bias, and insensitivity of the questionnaire as a whole. We were looking particularly for evidence that the instrument could be construed as either promoting or opposing ASAP. According to the cover letter attached to the pilot instrument, "The questionnaire is part of a study of the Arizona Student Assessment Program. Items in the questionnaire were grounded in a multiple case study on ASAP that we conducted last year. Four researchers spent a full academic year in four Arizona schools to try to understand the role ASAP plays in classrooms. For example, the statement 'Using ASAP means that there are just more things that teachers at this school have to cover' came directly from teachers' statements in the earlier study. The items represent a range of viewpoints that we discovered." The 16 participants in the pilot administration were practicing public school teachers in nearby elementary, middle, and high schools.

In Phase 5, we revised the instrument based on pilot participant responses and conducted an item analysis to identify anomalous and badly functioning items. The final form was then sent to the sample of teachers. A copy is included in Appendix C.

In the final phase of development, we conducted various data analyses to reduce the amount of information so that hypotheses could be tested more efficiently. The tables for the results of this analysis can be inspected in Appendix D. Using data from the sample, we constructed a single variable from 18 items in which respondents rated their knowledge on a 4-point scale (*expert*, *good understanding, awareness only,* and *none*). The items were related to ASAP ideals, such as "Problem solving or critical thinking" or "The Writing Process." Correlation analysis revealed that all 18 items were positively correlated, which justified a simple additive model for the variable, which was titled Self-Rated Knowledge (Table 1, Appendix D). An exploratory factor analysis using the principal components method was conducted on all items that could be considered belief statements (excluding those items in which teachers estimated conditions such as time spent in testing and professional development). Examination of the scree plot and list of eigenvalues suggested further exploration with 7 or 9 factors.

Using a maximum-likelihood method of extraction and oblimin method of rotation, a 7-factor solution seemed to converge on the best use of 7 variables within the set of questionnaire items. SPSS for Windows (Norusis, 1993) was the statistical package used for all statistical analyses. Further item analyses were conducted within these 7 variables. In one case, some items were deleted from the variables to make them more internally consistent and meaningful. All then were found to have acceptable reliabilities as indicated by Cronbach alpha statistics. Tables 2 through 8 in Appendix D present the items that comprise each of these variables: Belief in the Validity of ASAP and Essential Skills, Opportunity to Learn, Accountability/Authority Culture, Belief in the Benefits of ASAP, Belief in the Disequities of ASAP, Opportunity for Teachers to Learn, and Rejection of ASAP Ideals. These tables also present descriptive statistics from the sample.

Sampling plan. The measurement units of the study were teachers and teachers within schools. However, no sampling frame for teachers exists. We therefore selected a statistically representative sample of schools and surveyed all teachers within those schools. For high schools, only teachers likely to have contact with ASAP were surveyed. English, math, and language arts teachers ordinarily fall into that category. A proportionate stratified sample design was chosen because of its efficiency. Data available in the Arizona Education Profile (Arizona Department of Education, 1992) for stratifying a sample, however, were available only at the district level and not at the school level. Therefore, we began by constructing a matrix consisting of the selected stratifying variables, then listed all the districts within each cell, and then listed all the schools comprising the districts within the cell. We determined the proportionality of each cell in the matrix for the population to determine the correct sample proportion for that cell. Then, schools were selected at random within each cell. These steps are explained in detail in subsequent paragraphs.

Because of the structural arrangement of districts (either unified, elementary, or high school) and the dramatic differences in Arizona district size, we decided to select the sample to be proportionate to district type (unified and non-unified) and size. Small districts were defined as those with fewer than 1000 pupils. Mid-size districts were those with between 1000 and 8000 pupils. Large districts were those with more than 8000 pupils.

Because of the centrality of capacity and equity issues in the study questions, it was our original intent to stratify the sample on both pupil

composition and financial resources. However, no satisfactory indicator of financial capacity could be culled from available data. Instead, resources functioned as an independent variable measured in interviews conducted with principals when the schools were solicited for participation. Balance of resource-advantaged and resource-disadvantaged schools was then gauged after the fact from the sample data. In addition, state census data were accessed to determine the rate of property tax capacity per pupil. Although this statistic is imperfect as an indicator of financial resources available to adapt to a mandate, it nevertheless figured prominently in recent judicial rulings on the financial disparities among Arizona districts. Therefore, its use is justified. Districts were assigned to four categories based on this statistic: (a) below $500; (b) $500–$1000; (c) $1000–$2500; and (d) above $2500 per pupil. Subsequent examination of the proportions of districts falling into each category confirmed that the sample was indeed representative of the population with respect to the variable of district financial resources as indicated by property taxation capacity per pupil.

To capture variation on pupil composition, we found data on the percent of minorities served by each district in the population. We formed three categories for this stratification variable: high minority composition (more than 70%, across all groups), mid-minority composition (between 30 and 70%), and low minority (less than 30%) composition. We examined the relationship of percent minority with percent disadvantaged. Since the correlation between the two indicators was so high ($r = .68$), it was unnecessary to stratify on both.

The number of schools to be selected within each cell was based on the proportionality of the particular matrix cell in the total population and the total number needed for the sample. Working with some hypothetical data, we calculated the number of teachers that would be sufficient to test hypotheses of difference in capacity and equity (approximately 1250). The budget of the study permitted a more generous allocation. We set the number for the original sample at 3000. We knew or could estimate the number of teachers at each school and selected enough schools to reach the number of teachers needed per cell. We rounded up in each cell, adding schools until we were confident we would have enough teachers in each.

To achieve proportionality with respect to size, the following procedures were used. We determined the number in the pupil population attending schools in unified districts ($N = 404,025$) and in the non-unified districts ($N = 250,481$) in the

most recent census report available. Then we calculated the population percentages within each district type that fell into each cell of the minority status-by-size matrix. Table 9 (see Appendix E) shows the numbers and proportions for the population and sample allocation for proportionate random sampling by district size and type.

The actual number of questionnaires distributed differed from the allocation sample for three reasons. First, since the number of teachers per school could only be estimated, and since rounding up was necessary, we had to choose extra schools to make sure we eventually obtained the numbers of teachers necessary for inferring the characteristics of the population and testing hypotheses of interest. Second, because we had to rely on school principals to distribute surveys to teachers, we adjusted numbers upward. Third, since we also wanted to make inferences about districts having high and low rates of disadvantaged and minority pupils, we deliberately over-sampled schools in high minority districts (particularly reservation schools), adding extra schools at the sampling stage, with the plan of reweighting at the analysis stage if necessary. Table 10 (Appendix E) contains data on numbers of questionnaires actually distributed to teachers through their principals and the proportions within stratification cells.

Having determined the number of teachers needed within schools of the different categories, we then arrayed schools within each category. Using a table of random numbers with a blind start, schools were selected. This procedure resulted in a sampling fraction of approximately 38% of the school districts in Arizona, 12% of schools, and 9% of teachers.

Study procedures. We identified the principal in each sampled school and mailed a letter explaining the study and promising an honorarium for each participating school. We promised that their responses would be confidential and that they would have access to the results. Then, we attempted to contact the principal by telephone to solicit that school's participation in the survey. Any school that refused participation was replaced randomly within its stratification category. Whether agreeing or disagreeing to take part in the survey, the principal was asked three questions: What is the proportion of students at the school who qualify for free or reduced lunch? Is this proportion consistent with that of the school district? How would you rate the extent of resources you have available for staff and curriculum development? The first two questions functioned as a check on the accuracy of our method of selecting schools because of their minority

status when the data we had available was the minority status of the district. We had already found a high correlation in the school census figures between percent minority and percent on free or reduced lunch, so that one variable could act as surrogate for the other. These assumptions could later be examined in the data analysis. The third question served as a measure of the independent variable of district financial resource. We wanted to make sure that schools of differing degrees of wealth were adequately represented in the sample. In addition, the data from the phone interview, together with the status of the sampled districts known from the Arizona school survey, would allow us later to conduct nonrespondent bias checks (differences between original sample, replacement sample, respondents and nonrespondents on key variables and status characteristics). For those principals who agreed to participate, we asked them the number of teachers that would be eligible to be measured. This was defined as all regular education teachers in elementary and middle schools and all teachers of English, language arts, and math in the chosen high schools. Finally, we asked whether the school's participation was contingent on district approval.

Principals were instructed to distribute questionnaires to faculty. A cover letter to teachers instructed them to complete the questionnaires, seal them in enclosed envelopes, and return them to the principal. Principals were to collect them and return them in the package provided. Teachers in three schools returned questionnaires separately. Questionnaires had school codes but no codes for individuals. Teachers were informed about the purpose of the study and promised confidentiality and reports of the survey results. Schools with adequate return rates were promised a $50 honorarium, to be spent according to the wishes of the faculty.

Besides the representative sample, we conducted a solicitation sample. At a meeting to describe the study to the Greater Phoenix Urban Council of School Administrators, several administrators requested to participate. Based on the number requested by volunteering administrators, an additional 360 questionnaires were sent. These data were flagged so that they would not be merged with the representative sample results. After examining the initial returns for underrepresented categories, we solicited the participation of schools with known status characteristics among practitioners we knew. These returns were also flagged and evaluated separately. The solicitation sample consisted of 86 returns.

Follow-up procedures. The first wave of data collection took place in April and May 1994. Analysis of response rates suggested that a second wave be conducted in September. Phone calls were made to principals in the schools originally sampled. In addition, replacement schools were drawn at random from appropriate stratification cells for those schools that refused participation or who failed to return questionnaires. Replacement sampling is the method employed by the National Assessment of Educational Progress and has been found to be unbiased.

In the case of eight schools, a member of the research team went to the schools and collected the questionnaires in person. Follow-up was more intense for schools falling into those stratification categories poorly represented in the first wave of respondents as well as schools in districts with high rates of minority and disadvantaged pupils The final sample (see Table 10 in Appendix E) was counted as those schools that were contacted and, if refused, were successfully replaced.

District-level analysis of response rates. As already explained, schools were the sampling unit and were selected at random from the population of Arizona schools. The schools selected for that sample were members of 81 school districts. Three sources of data informed a district-level analysis: document analysis of the District Assessment Plan (DAP), phone interviews with district administrators, and teacher surveys aggregated to the district level. Of the 81 districts sampled, three districts refused to participate in any phase of the study. Ten districts failed to participate in the district interviews, either because of outright refusal or because contact was never successfully made. Six districts failed to participate in the teacher survey but did participate in the administrator interviews.

The estimation of response rate, therefore, is 95% for the DAP analysis and 84% for the administrator interview analysis. The response rate at the district level for the teacher survey is 89%.

Inspection of the differences between participating and nonparticipating districts shows that no response bias exists in terms of district type, district size, or district level of resources. Analysis by pupil composition shows a small disproportionality. That is, districts with high rates of minority pupil composition are slightly underrepresented in the district sample.

With that caveat in mind, it is safe to argue that the respondent sample is highly representative of the population of Arizona districts. The descriptive results from the sampled districts may be generalized to the population with good confidence.

School-level analysis of response rates. Two sources of data contributed to school-level analysis: (a) administrator interviews for the schools selected in the sample; and (b) teacher questionnaire data aggregated at the school level. Administrator interviews consisted of phone interviews of school principals ($n = 53$) and phone interviews of district administrators in those districts that consisted of only one or two schools ($n = 24$). The response rate for administrator interview data is 69% at the school level.

As Table 11 (Appendix E) shows, questionnaires were sent ultimately to teachers at 111 schools. Of these, usable responses were received from 97 schools, resulting in a response rate of 87%. The responding and nonresponding schools were equivalent in terms of district type, district size, and pupil composition. Therefore, inferences from the school sample to the population of Arizona schools can be made with good confidence.

Teacher-level analysis of response rates. Computation of the response rates of teachers was based on the numbers of teachers the principals reported during the initial phone interviews we conducted to solicit participation. We discounted from that number the estimated number of teachers in schools we were never able to contact by phone. We also discounted the estimated or known number of teachers in schools that refused participation but which were successfully replaced at random from within their relevant stratification category. The numbers in the replacement schools were then added to the account. As a result of the follow-up procedures and second wave of data collection, the number of questionnaires sent to teachers was 2696 and the number responding was 1360, for a response rate for teachers of 50%. As already explained, the respondents did not differ from nonrespondents on any district characteristic. The responding sample was balanced with respect to gender, experience, and grade level taught and therefore can be considered broadly representative of the population of teachers.

However, the question of whether the teachers who took the time to fill out the questionnaires did so because of a set of beliefs and values different from those

who failed to participate can never be completely ruled out in studies of this kind with response rates of 50%. A partial answer may be found in the juxtaposition of data sets from multiple perspectives and generated by multiple methods.

Analysis of data. Data from questionnaires were cleaned and coded. Open-ended items were coded using content analysis procedures. Data were entered in a data file in SPSS format. Exploratory data analysis techniques were used to identify misclassifications and variables out of range. Factor analysis and reliability analyses were conducted as described above. An analysis of means of key variables differentiated by type of sample revealed that the solicitation sample was substantially biased in comparison to the original sample. Thus, data from the solicitation sample was filtered from the rest and excluded from subsequent analysis. Analysis of the replacement sample showed no such bias and these cases were therefore merged with the original sample for all subsequent analyses. Five cases initially counted in the original respondent group were found to be misclassified. Thus, 1355 cases comprised the analytic sample.

Descriptive data were computed on the percentage responding to each option of each item. Item-level data are reported in Appendix F. Three a priori hypotheses were tested: that pupil composition, financial resources, and type of community would influence the responses of practitioners to the policy. To test these hypotheses, analyses of variance, with Scheffé contrasts, were computed on the means of the following variables: Self-Rated Knowledge, Belief in the Validity of ASAP and Essential Skills, Accountability/Authority Culture, Student Opportunity to Learn, Belief in the Beneficial Effects of ASAP, Belief in the Dis-equities of ASAP, Opportunity for Teachers to Learn, and Rejection of ASAP ideals. Simple comparisons were computed on key categories of interest (e.g., high minority vs. low minority schools) on items, and effect sizes (e.g., the difference between means of high and low minority samples divided by the common standard deviation) calculated for the differences. These data are listed in Appendix G. The reader may consult Appendix F for more detailed, item-level data. Findings based on variables refer to the characteristics in Tables 1-8 (Appendix D). With respect to the interpretation of group differences (documented more fully in Appendix C), only differences of effect size > .20 and statistical probabilities of $p < .001$ are reported.

Open-ended comments. Approximately 30% of the teachers surveyed elected to complete the free response item. Data from these responses were set

aside for a qualitative content analysis. The codes that were used corresponded to the categories in the conceptual framework and the Analytic Matrix.

Administrator survey. The sampling plan for the teacher survey was also used to sample administrators, which we sorted into two groups. The first group consisted of district administrators; the second, school principals. A telephone interview protocol was constructed to correspond to the same variables as those in the teacher questionnaire. A letter was sent to all members of the sample to inform them of the study and request their participation. A preliminary call was placed to schedule the interview. The interviews themselves lasted approximately 15 minutes. Of the district administrators, 85% agreed to participate. Data from the interviews were recorded in handwritten notes, transferred to coding forms, and analyzed. The methods and findings are reported in Olivares-Seck (1994). Findings then were fed into the trans-method analysis described below.

Supporting Studies

District Assessment Plan (DAP) analysis. As part of the Olivares-Seck (1994) study, we also collected the DAP plans from each district sampled. These were content-analyzed and contributed to the trans-method analysis.

Policy study. During the initial year of the study, Noble (1994) completed a study of policy makers' beliefs and images and the discrepancies between them and those of practitioners affected by ASAP. These data were also available for the trans-method analysis.

Parish study. In the original design of the multiple case study, we intended to examine the hypothesis that schools within a district may differ from each other in how they respond to ASAP. We approached a colleague and doctoral student, Carol Parish, who was also an administrator in a nearby suburban district. Our idea was to study two schools, one to be studied by a member of our research team and the other by Parish, so that she could use it to complete her dissertation project and contribute to our design as well. Although this plan was then initiated, Parish's time commitments made it impossible to complete her part of the study on our schedule. The multiple case study report, therefore, was finished without benefit of her case. However, she did complete her study (Parish, 1996) in time to contribute it to the trans-method analysis. She was able to follow several schools in her district over a two-year period and helped to address the original issue of variance of schools within districts.

Trans-Method Analysis

Justification

Despite many calls in the methodological literature for studies that encompass multiple research approaches and methods, there are few prescriptions and even fewer examples to follow. Behrens and Smith (1996) noted the similarity, and hence the compatibility, between a exploratory variety of statistical analysis and qualitative analysis. Their argument for compatibility rests on a particular epistemology that views knowledge as constructed rather than discovered, as provisional rather than definitive, and as both personal (the analyst is implicated in the process) and social (the norms of the community influence the process). Methods are based on assumptions that one can deconstruct. Using a single method of inquiry leaves one open to "method error." Using more than one method allows one to juxtapose one set of data against another, sequentially posing and discarding alternative hypotheses, thus leaving a clear trail for readers to audit and, one hopes, a plausible account of evidence and reasoning about it. In this model of investigation, the *methods* of data collection (whether participant observation or questionnaire, as long as the collection was done thoughtfully and with integrity) and the *form* of the data (numerical or verbal) matter less than the progressive problem-solving model and the attitude one takes toward the results.

Although methodologists commonly identify Erickson solely with qualitative inquiry, the method of modified analytic induction he described in his 1986 work describes the above model of working with either qualitative or quantitative data, or even with both kinds. Readers who are familiar with Erickson's 1986 chapter will recognize in the following description the features of the analysis he recommended. I refer to the procedures as a trans-method analysis to distinguish the approach from the multimethod, multiperspective model of Campbell and others, who associate valid inferences with triangulation across independent sources of method and perspectives (e.g., Campbell & Fiske, 1959). For Erickson, there is no acid test of valid inferences, and one at best works toward coherence of assumptions, data, and assertions.

Description of Methods

In terms of the aggregation of studies, data collection extended over nearly four years. Throughout this time, preliminary ideas emerged as the team examined and discussed them. Memos documented this process. It is characteristic of qualitative studies that problem setting, design choices, data collection, and analysis overlap. Nevertheless, a phase primarily devoted to data analysis can be demarked. The phase consisted of repeated readings of the data as a whole, generating preliminary assertions, establishing the warrant for the assertions, developing representations from the analysis, and documenting the process in memos.

First, the data from all parts of the study were assembled and read at least three times. Memos kept track of the themes and concepts that arose during this reading. Two working assumptions guided this phase of the work. The data sources had to be treated at the level at which they were least processed. In other words, when interview transcripts were available, as was usually true, they became the data to be analyzed. The case study reports themselves were treated as data when the researchers' write-ups were not themselves available. Similarly, the descriptive statistics from the survey were treated as data, rather than any inferential statistics such as regression coefficients that had been computed during the survey research study. The second working assumption had to do with orientation to the different kinds of data. Qualitative data and quantitative data were treated as equivalent, and neither kind was privileged over the other in terms of its potential to inform.

From the repeated readings through the record of data (which ran to more than 2500 pages), the analyst generated a set of preliminary assertions. Assertions are statements that the researcher believes to be true based on an understanding of the data as a whole, arrived at largely through induction. For example, the data seemed to suggest that most teachers knew about ASAP, although exactly how they understood it varied dramatically from person to person, almost as if people were talking about a completely different entity, but using the same label. At this early stage there were a dozen ambiguous and complex sentences like this on the list of preliminary assertions, and some overlapped others.

Following Erickson's lead, I next began a process of refining and establishing the warrant for each assertion, a process he describes as skeptically looking for reasons to hold the work suspect. Warranting proceeds one assertion at a time and involves a systematic search through the data record for segments that support or confirm the assertion. Data segments, such as passages from interview transcripts, observation write-ups, or descriptive statistics from survey responses, then became "instances" or "indicators" that were organized and indexed (many analysts appreciate qualitative data analysis packages for this part, but I used low-tech methods such as sticky notes and text markers). Particular attention was paid to confirming instances that were generated by alternative data collection methods. This practice follows the assumption that assertions based on a sole form of data (e.g., questionnaire responses only or focus group interview only) are less robust and persuasive than assertions based on multiple data sources.

Having identified and catalogued the confirming instances, the next step in analysis was to search for disconfirming instances, the discovery of which provoked rethinking, recasting or revising the assertion. If disconfirming evidence had been sufficiently weighty, an assertion would have been discarded outright, but this did not happen. One preliminary assertion that was revised had been stated, "Exclusive of the sites that were moving on their own toward constructivism, ASAP has produced little coherent change." Parish (1996) presented an account of Desert School, where changes she attributed to ASAP did in fact take place. This account persuaded me to rethink my unduly pessimistic interpretation of the evidence. Thereafter, I looked for types of change rather than a simple change/no change category and eventually stated an assertion more consistent with the data.

Part of the warranting process involved looking for negative or discrepant cases that violated patterns that had been found. I had asserted that local conditions—ideologies and images of learning and curriculum, resources, and accountability/authority structures—all have to be working in concert for ASAP to take hold. The data from the Extension Study interviews at Hamilton challenged that pattern. I examined that school carefully and developed the notion of a micropolitical process that undercuts those structural situations and has an effect on change toward or away from ASAP ideals.

Following through with the warranting process for each preliminary assertion and sorting through redundancies and overlaps, I eventually arrived at a final list of eight assertions, which are used to organize chapter 3.

The next step was to select data excerpts that would best depict the central ideas, themes, and patterns in each assertion. The selection aimed not for statistical generalization but for analytic generalization (Glaser, 1978), that is, an obvious tie between concepts and the referents on which the concepts rest.

To serve the same purpose, I constructed analytic vignettes (Erickson, 1986). Both the quoted excerpts and the vignettes should demonstrate for the reader the truth of the assertion and provide concrete particulars in a vivid slice of life. Through these details, the reader should be able to judge the process by which the researcher arrived at the assertions and gain understanding though vicarious experience. Finally, interpretive commentary linked the assertions and data and pointed to links with extant literature and theory.

Following Erickson, my claim to the validity of the work lies in the extensiveness of data and the comprehensiveness of views on the phenomenon, on the coherence of assumptions, frameworks, data, and inferences, and on the plausibility of the account.

CHAPTER 3
FINDINGS AND ASSERTIONS

What happens when a state government implements a program to change education and make schools accountable? Four years of research produced a comprehensive set of evidence that only begins to answer this question. In this chapter we advance a set of analytic assertions about the consequences of ASAP policy in Arizona in the four years of its existence. In the section to follow, we restate each assertion and a portion of the warranting evidence in support of it. Since this is not an evaluation study, we avoid using words such as beneficial effects, effective outcomes, or positive response. Instead, we use the terminology "beliefs and practices coherent with and consequent to the policy."

ASSERTION 1: **Most educators were aware of ASAP although their definitions of "ASAP" varied.**

ASSERTION 2: **Approval of ASAP was far from universal.**

ASSERTION 3: **Action coherent with policy intents had begun to be realized in some places. Categories of response were: Compliance Only, Compromise, Coherent Action, Drag.**

ASSERTION 4: **Responses coherent with ASAP intentions were centered in a few places where circumstances were auspicious.**

ASSERTION 5: **Inadequate capacity and capacity building impeded coherent response to ASAP intentions.**

ASSERTION 6: **State inattention to the technical and administrative adequacy of the assessment and accountability system impeded coherent responses to ASAP intentions.**

ASSERTION 7:   **The reform intention and the accountability intention of ASAP conflicted with each other, and the conflict impeded coherent action.**

ASSERTION 8:   **The lack of attention at the state level to concerns for equity and fairness inhibited coherent local response to the policy.**

In the next section, we repeat each assertion and present portions of the evidence that substantiate it. We focus on what happened subsequent to the implementation of policy (Assertions 1-3) and how these things might have happened as they did (Assertions 4-8). As noted in the Methods chapter, we identify the source of each supporting exemplar as originating in the case studies, focus group extension study, survey open-ended comment (SOC), or teacher or administrator survey. In addition to the quoted exemplars, we employ the analytic vignette (Erickson, 1986) to demonstrate the warrant for the assertion. Note that no assertion we state here depends on a single source of data or method of data collection. For example, even though the quoted exemplars come from open-ended survey items, other excerpts—say from extension group interviews—also fed into the construction of the assertion.

ASSERTION 1: Most Educators Were Aware of ASAP Although Their Definitions of "ASAP" Varied

According to Blumer (1969), in order to act toward something, a person must first be aware of it and also form a definitional object of it. One's definitions or interpretations are constructed through local social interactions; therefore, definitions vary across contexts and persons. Applying this notion to the ASAP reform, this means that, to respond to ASAP intentions, a teacher must be aware of ASAP and must have come to a particular understanding of it: One's action rests on one's interpretation.

The weight of the evidence of this study supports the assertion that Arizona educators were aware of ASAP. During the course of the case studies and extension study, we encountered no one who had not heard of it. Moreover, on the survey, less than one quarter of the teachers agreed with the statement "I don't understand ASAP well enough to adapt my teaching to it" (65% disagreed).

Although awareness of ASAP was widespread, the interpretations or definitions that teachers had constructed varied significantly, according to the following typology: (a) "ASAP" is the test that one gives; (b) "ASAP" is a test that is different from ITBS; (c) "ASAP" is a state program to change schooling; (d) "ASAP" is a weapon in the battle over curriculum; and (e) "ASAP" is a weapon in a political battle. Of course, these definitions are not mutually exclusive, nor are they exhaustive of all possible interpretations of ASAP. Nevertheless, it is important to distinguish them, because the action that a particular teacher takes (i.e., whether to adapt one's pedagogy toward ASAP intentions) depends in large measure on the interpretation that teacher constructs. "ASAP" was not a singular or standard phenomenon.

"ASAP" is the test that one gives. Individual teachers formed definitions of ASAP based on social interaction in their particular contexts. Teachers were more likely to experience "ASAP" as an achievement testing instrument than they were as an instrument of reform or as a political instrument, because they had to unpack the boxes, read the directions, struggle with the scoring rubrics, and the like. When one hears teachers talking (or reads their comments on the questionnaire) about "giving the ASAP" or "scoring the ASAP," it suggests that they are not thinking about a state program but about a particular test.

Teachers in the benchmark grades (3, 8, and 12) and subject areas (e.g., secondary school English and math) tended to define "ASAP" as ASAP Form D. And, their propensity to adapt to ASAP related to their beliefs that Form D was a good thing.

Teachers in nonbenchmark grades and subjects, however, tended to form ambiguous definitions of Form D and regarded it as someone else's problem. As one third-grade teacher said:

> It is still viewed as a third grade test at my school. (SOC)

Instead, teachers in nonbenchmark grades associated "ASAP" with the tests they themselves had to give, which were not Form D but their district's testing program as determined by their District Assessment Plan (DAP). Since there was so much diversity in DAPs from district to district, teachers' definitions varied accordingly. The characteristics and problems encountered in the course of the

DAP testing were projected onto the definitions of "ASAP" that the teachers in each district constructed.

The District Assessment Plan (DAP) and the Essential Skills Reporting Document (ESRS) were the mechanisms designed for making the district accountable to the state for *all the Essential Skills* (rather than the sample that was measured by the Form D tests in any particular year). On its DAP, a district had to list the means for measuring mastery of each skill at each level. On its ESRS, the district would later state the percentage of its pupils at each grade that had mastered each skill. To accomplish this mandate, some districts used ASAP Forms A, B, and C to test in nonbenchmark grades *and* benchmark grades. Other districts used the criterion-referenced testing program already in place from the previous state testing mandate. Still other districts used portfolio systems that were scored according to the state generic rubrics. Because there were so many Essential Skills for which districts had to report achievement, the DAP assessments resembled a mastery testing system: multiple assessments on small pieces of achievement, often measured two or three times per school year. The comments of a high school teacher reflect this problem:

> We are required as [high school] English teachers to give both the Reading and Writing ASAP. This requires at least 3 weeks of classroom time since we are required to administer different forms of the ASAP 3 times! As a teacher who already does a great deal of writing and reading instruction, this addition interferes rather than enhances my curriculum. (SOC)

The DAP testing created enormous frustration and often attached negative valence to teachers' definitions of "ASAP." This phenomenon came across most strongly in the open-ended survey comments volunteered by teachers. For example, many complained that they were required to give tests designed for third graders to their second graders, and the resulting difficulties contributed to children's frustration and sense of failure.

> ASAP skills are not developmentally appropriate. I believe that we are asking students to perform tasks that they are not capable of performing and to grasp concepts they are not developmentally ready for. All the oral directions on the ASAP put students with language processing problems at a distinct disadvantage. Most of ASAP tests the student's ability with written expression, which again leaves out many other abilities. (SOC)

Responding to the local DAP testing requirements, teachers objected to the excessive number of tests, to the instructional time lost to test preparation and scoring, and to their planning time lost to test scoring. For example:

> Our district gives three writing and three reading ASAP tests per year. English teachers are to administer and grade them. They took four and a half weeks to administer this year. I did not prepare my students for these tests by studying similar material [as the district recommends]. If I had, it would have taken another four and a half weeks to prepare them. That's one-fourth of the year devoted to state testing that has to come out of instruction, drill, etc. We have material that we are to cover in English classes each year. This year I eliminated several required elements to do the state testing. I would not mind it if the tail wagged the dog—that is, if the ASAP tests determined instruction. I do mind when the tail replaces the dog. (SOC)

Having defined "ASAP" as their DAP testing requirements, teachers objected to poor-quality rubrics and to the subjectivity of the scoring process. They objected to the system requirement that the schools reproduce the testing materials and procure manipulatives and other materials necessary to do the testing. They objected to the errors and inconsistencies in the test materials. They pointed out that Forms A, B, and C (wherein reading, math, and writing were measured separately) were not consistent with D (wherein the content areas were measured in integrated tasks). Exemplars concerning the poor qualities of the various tests are included later under Assertion 6.

"ASAP" is a test that is different from ITBS. The reactions of many teachers rested on the interpretation of "ASAP" as a test different from the standardized tests that had formed the prior state testing mandate. One might say that "ASAP" was defined as much by what it was not, as by what it was. The following quotation is typical of many similar remarks.

> The ASAP test is of high interest. It challenges the students. It's much better than ITBS. (SOC)

"ASAP" is a state program to change schools. Although the dominant object identification of "ASAP" was the tests personally given, some teachers also understood that "ASAP" was more than the tests and included a comprehensive program with a reform agenda. We found evidence that some teachers defined

"ASAP" as a state program with intentions to change curriculum and pedagogy toward constructivism and away from sole reliance on basic skills.

> The ASAP program has one purpose. To change the nature of instruction. I believe this change is needed. What we have done in the past in schools is not working any more. Changes are necessary. ASAP is a good idea. Getting it to work is another question, but the state is struggling to do just that right now. (SOC)

> It is a good thing that the state is trying to do in moving towards ASAP and away from ITBS. However, it appears to me that the ASAPs are a long way from truly reflecting what students are capable of inside and outside the classroom. In many ways, the tests are actually a step down from what we are doing regularly in the classroom. (SOC)

> You know there are other schools that need the ASAP to make them teach holistically. We didn't need it. We were already teaching that way. (Extension Study, Hilldale)

"ASAP" is a weapon in the battle over curriculum. Among those educators who defined "ASAP" in terms of its reform ideals, however, orientation was markedly divided. Some defined "ASAP" as representing an unfortunate and even dangerous de-emphasis of foundational skills, whereas others welcomed the change or saw the new emphasis as encompassing both skills and problem solving. We looked at this dichotomy as instances of curricular politics: between holistic reading and writing, and problem solving in math on the one hand, and basic skills, phonics, and the like on the other. This dichotomy is not merely a rational choice point but full-blown ideology that one can see both in these data and in the national media.

> ASAP is a step forward—a good step, too. That's why so many "traditional" teachers are reacting so much against it. ASAP should naturally cover what is being taught if the curriculum is up-to-date. (SOC)

> I believe the "idea" of the ASAP tests is excellent. Students need to know that reading, writing, and mathematics are related to each other in the "real world." (SOC)

> I began teaching when I was 19 years old. I have had a wide-range of experience, including working many years with the so-called disadvantaged kids. They, and

many of the students we are coming across nowadays, cannot do well with ASAP because of its requirements. Four of my closest friends teach in the inner city where whole language is being thrown at them and they have been forbidden . . . to practice and drill. Yet these very students they are teaching are kids who will be lucky to hold jobs in the service industries that pay little. They need to be able to count, add, multiply, divide, subtract, read, follow directions, know how to make change, etc. ASAP is not fitting their needs at all. I am sincerely appalled, heartbroken, sick, upset, frustrated, etc. by the lack of skills people in general have today. Nobody cares about basics . . . I blame the public schools where I see so many important integral parts of instruction going down the tubes. The young teachers coming out of college will just perpetuate the problem since they are learning whole language instruction and student-centered classroom. Certainly these concepts have their merits, but not at the expense of basics on which education is based. They are not learning how to teach phonics so our children will not be able to read; they do not know how to spell themselves, so our children will not learn to spell. etc. . . . Isn't that pathetic? (SOC)

The ASAP (and all outcome-based education) is designed to do away with "skills" because kids today don't relate to skills, because they are boring. By pandering to this we are weakening our society, not strengthening it. *It is wrong!* I was told by a state official that teachers would be more like coaches under ASAP. Ask any coach if they teach skills in isolation before they integrate it into their game plan. They will all tell you yes. I rest my case. (SOC)

I don't like, one, professors at ASU, and two, the state department, telling me how exactly I need to form my philosophy of education. You know, if I'm more of a whole-group, basic skills type teachers, which I am, and I'll admit to that, my kids sure can add, subtract, multiply, divided, and write pretty well. I hate it when they say, "You're using a basal, you use text books, you're wrong, you're a horrible teacher." Because I'm not. We have to educate everyone, and we have to get the kids ready to go over to the high school. I don't think you can jump into abstract thinking when you don't have the basics of two plus two. (Extension Study)

On the survey, nearly two thirds of the teachers expressed the belief that "pupils at this school need to master basic skills before they can progress to higher order thinking and problem-solving." Nearly one third believe that some pupils will never be able to profit from integrated, thematic instruction. These survey results confirm the notion that many teachers disagreed with the curricular assumptions built into ASAP reform intentions and were on the opposite side of a political and ideological fence.

"ASAP" is a weapon in a political battle. Another common definition of "ASAP" was as much a political and symbolic tool for various powerful interest groups as it was an authentic mechanism for changing education. This macropolitical definition was also reflected in an objection to "ASAP" as impermanent—that the "ASAP" program was only the latest in a long series of state mandates that would peter out in time. The macropolitical definition is illustrated in the following remarks of teachers.

Those who have had to administer it should be consulted to evaluate its worth. Also, *classroom teachers*, not bureaucrats and ivory-tower agenda-pushing consultants, should be involved in reworking and advising. (SOC)

GET REAL. GO BACK INTO THE CLASSROOM AND SEE WHAT TODAY'S TEACHERS ARE DEALING WITH. I sincerely resent all the outsider influences of educated fools and politicians who don't know the real school world and yet are dictating what teachers do. (SOC)

I feel the concept of ASAP is good but . . . we were told the tests were to be used for assessment not comparison yet you see scores compared for districts and schools by the newspapers. What happened! (SOC)

If we thought ASAP would make students think more clearly about mathematics problems we would not mind all the additional time we are now putting into this program. If the people who designed it, so that they could claim that our schools are truly accountable for student learning, and (thereby reap political benefits) could see how these assessments are being administered, they would abandon ASAP as quickly as is politically feasible. Any validity in the results of these assessments is purely coincidental. (SOC)

I believe strongly that any testing program should be heavily tied into promotion and graduation. Until this is done, students will not take them seriously enough to make them worth the time, effort, and money! (SOC)

If the great education gods in the Arizona Department of Education determine this is the way to go, spend the tens of millions required to train teachers and purchase appropriate texts. (SOC)

I read the test and feel that whoever wrote it must have been paid by the number of syllables. The children don't even have a clue about the topics and have no reason to

want to know. The designers should be locked in a room with 27 typical first graders and administer the math assessment. (SOC)

I hear they're already having problems with ASAP as far as manpower to grade them and the money to pay for it. So how can you get motivated and want to learn more when you think, "Okay, wait a year down the road and they may do away with this thing." It's pointless. (Extension Study, Pine)

Over 40% of the teachers surveyed agreed with the statement that "ASAP is just a fad that will disappear." Concurring with this survey evidence was this comment from a teacher at Franklin:

Because you get a new administration there at the state level and they might not agree with it at all. It's just a matter of opinion. You know how the pendulum, the swing of the pendulum is in education, right? So this administration goes out, a new one comes in and then the ASAP goes out and something else comes in and you're jerked around, you know. (Franklin case study, Year 1)

Frustrations with the technical aspects of the test (also see Assertion 6) were also expressed in political terms. For example, teachers in the survey volunteered comments about the tests having been written by people who have never been in a classroom, about how students have changed and test writers have little idea of the conditions of teachers' work. Several suggested that legislators or state department personnel be required to take the tests themselves or try to administer it to students. One referred to ASAP as Another Stupid Aggravating Problem. Others pointed out that ASAP had originally been presented to teachers as a tool for instructional change, but it had since turned into a complex and hellish accountability tool.

It took so many hours to cut out and prepare the manipulatives for the Form A test in second grade that it left us resentful. Then it took us over 47 hours to score the tests in my class. (SOC)

These data excerpts make it obvious that when people write or speak about ASAP, they are not all talking about the same thing. Variations in definition of ASAP are significant in understanding the consequences of the program we describe later in the report.

ASSERTION 2:   Approval of ASAP Was Far From Universal

In a loosely coupled system such as education, dictating change is unrealistic; the state must win the hearts and minds of those who deliver the services. Yet after four years of the program and three years of the assessment, a majority of teachers failed to endorse ASAP, however they defined it. A recurring comment in the qualitative data was that ASAP was "a step in the right direction," or "a good idea" that fell down in execution. There was substantial variance in approval, however, as the following quotations illustrate:

> The children and I have enjoyed the assessment. We have fun while taking it. It is a positive learning experience. (SOC)

> Process based instruction allows the students greater ownership of the learning. Basic skills must be acquired but when learning is approached holistically, the desire for success comes from the learner's motivation. (SOC)

On the survey, only a small percentage of teachers (19%) agreed with the statement "ASAP represents the best we know about how students learn." And 43% believed that "ASAP takes away from instructional time we should be spending on something more important." Nearly two thirds of the teachers believed that "many teachers are fed up with outside demands and just hope ASAP will go away." Only 27% believed that the "benefits of ASAP testing are worth the investment of effort, time and money it takes to administer the program." Nearly half the teachers surveyed believed that "many teachers disagree with the philosophy of curriculum, instruction, and assessment that ASAP represents."

What seemed to generate approval for ASAP was the simple fact that it was seen as improvement over the ITBS.

> I feel that ASAP is a good concept as I don't believe the ITBS is a good indicator of student achievement. However, I'm not comfortable with the process yet as every year there are different kinks to address and monitor.

> I think that the ASAP has been a definite improvement over multiple choice which we had before. In the days that we did nothing but multiple choice, we didn't have any way of measuring what students learn, particularly in the area of writing, and in reading too. I think it's a step in the right direction. At the same time, from my

experience in the last two years, ASAP is still limiting in terms of our [whole language] program and measuring what our kids can do. It's still flawed, in terms of our program, but it certainly comes closer to the way we instruct children. (Extension Study)

On the survey, teachers approved of the Arizona Essential Skills at higher rates than they approved of ASAP generally. Fifty-five percent of the teachers agreed or strongly agreed with the statement "If the district curriculum is aligned with the Arizona Essential Skills, then every student will have equal opportunity to learn challenging and important material." Over half believed that "the Arizona Essential Skills represents high standards of achievement."

ASSERTION 3:   Action Coherent With Policy Intents Had Begun To Be Realized in Some Places. Categories of Response Were: Compliance Only, Compromise, Coherent Action, Drag

What happened to the educational enterprise as a consequence of ASAP? We examined this question from as many perspectives as possible, looking for traces in classrooms and materials, and questioning educators about the effects they believed that the state mandated reforms may have had. The results presented a picture of low rates of change overall and considerable variability from place to place. Some schools had made near-Herculean efforts to change in the direction of ASAP reform ideals. Others had done nothing.

On the survey of teachers, for example, we found that only a small percentage (20%) of the teachers believed that "as a result of ASAP, many teachers at this school have changed the way they think about teaching and learning." Over 50% reported that "ASAP has had little or no effect on my teaching." Only about 30% agreed with the statement "As a result of ASAP, major changes in curriculum have been made at this school."

The administrators whom we surveyed presented a more positive picture of ASAP effects than did the teachers. Fifty-eight percent of administrators reported that ASAP has had a few or many positive effects. Fifty-three percent agreed with the statement that the benefits of implementing ASAP were worth the costs of money, time, and energy that had been invested (although few were able to report or even estimate the costs of implementing the program). Thirty-one percent disagreed. Fourteen percent said the reform has had a mixture of positive

and negative effects. Eleven percent of the district administrators reported that ASAP had had "a few" or "many" negative effects. Twenty percent reported that the prospects for change in the future toward ASAP ideals were unlikely.

Both the teacher and the administrator surveys provided evidence that changes had been made in district curriculum. About 40% of the teachers reported that district scope and sequences had been aligned with ASAP. Two thirds of the administrators surveyed reported such change. Changes in scope and sequence, however, were imperfectly translated into changes in the curriculum offered to students. About one third of the administrators reported that major changes had been made in curriculum as a result of ASAP. Only 37% of the teachers surveyed reported that the textbooks and materials they used were compatible with ASAP.

Although the survey results show that the effects of ASAP reform were low overall and highly variable across the state, there can be little doubt that action coherent with the policy intents was evident in a number of schools. By examining the variations of survey responses across schools and districts and reflecting on the case study and extension studies, we attempted to characterize the kinds of changes we saw. Changes consequent to ASAP seemed to fall into a typology that we characterized as "coherent action," "compliance only," "compromise," and "drag." The following exemplars illustrate this typology.

Coherent action. The case of Desert School, in Peak Valley District, represented practitioner action and change that was consistent with the ASAP reform intentions. The following vignette was taken from the study by Parish (1996), who asserted that ASAP was "critical" in causing fundamental change in beliefs about curriculum and practices of instruction and assessment at that school. One of the teachers she followed over a three-year period she named Ms. Wonder.

Wonder at Desert: Consequences Coherent With ASAP Intentions

Ms. Wonder teaches third grade in a new school, with all the amenities one could ask for, in a new suburb. The school serves mostly white, middle-class families and in every way is an advantageous place to be. The principal has been selective in assembling a teaching staff that works energetically together, cares about kids, and shares certain views about teaching and learning.

Parish describes Ms. Wonder as philosophically inclined toward ASAP reform ideas even before the program came into play, already emphasizing meaningful reading and writing as opposed to "instruction by ditto sheets." Ms. Wonder described herself this way during the first year of the study (the first full year of ASAP implementation): "I want learning to be interesting. It has to connect in some way for an outcome, and the more process involvement that I can plan into it, the better. . . . We are here to help them develop thinking processes, not just pouring information into their heads" (Parish, 1986, p. 53). Ms. Wonder felt at odds with the district's emphasis on CRTs and tried to find ways to integrate the piecemeal district curriculum She stated, "Learning needs to be connected, and if that learning can be related to something personal—then it becomes meaningful" (Parish, 1986, p. 59).

These philosophical predispositions augured well for Ms. Wonder's receptivity to ASAP, when she initially encountered it, even though the implementation was "disjointed and nightmarish. . . . When I received four boxes of ASAP test booklets representing 18 different assessments, I was overwhelmed! I wasn't quite clear about how to use them [Form As], how to score them, or what to do with the results. But I learned about ASAP right along with everyone else, our team saw value in the reading and writing measures right away, and we used them more for learning lessons than assessment measures. I like how students were asked their opinion and then asked to support their opinions with reasoning. That's high level and it gives us, as teachers, more information about their thinking processes. . . . I'm still learning how to assimilate them into our curriculum, though" (Parish, 1986, p. 72).

In the next two years, Ms. Wonder continued to align her teaching practices with the ASAP reform ideals. She accomplished this in several ways, according to Parish. First, she sought out training for herself in principles and methods of performance assessments. The district brought in experts such as Rick Stiggins (e.g., Stiggins, 1994), and Ms. Wonder made use of every opportunity. In addition she read books and articles on the topic of performance assessment and held many discussions with other members of her third-grade team. Parish credits this team-building, at the school level, and constructing a shared vision of assessment as crucial to the successful

change toward ASAP reform ideals. In addition, the principal's beliefs and assumptions about teaching and learning were consistent with those of the team. The principal allowed the team time to experiment and collaborate and supported their capacity-building efforts. Ms. Wonder described it this way: "Of the few teachers in the district who had a broad understanding of the vision of the ASAP, together we committed to implement performance-based assessments as professional goals. We were on a mission to teach the State Essential Skills and integrate ASAP as much as possible." Not only did Ms. Wonder build her capacity along with that of the team, she sought out the role of teacher-mentor so as to expand the capacity of other teachers in the district.

By the end of the ASAP program, Ms. Wonder's class revealed a complete integration of reading, writing, math, science, and technology. Moreover, all her teaching seemed to include assessment through the use of pre-established criteria and rubrics (which, she said, provide clear targets for students to pursue), portfolios, and student self-assessment. She looked back over the three years of her development with the ASAP program in a final interview with Parish.

> I value ASAP tremendously. I was one of the teachers that bought into it. Once I learned more about the why, it was much easier for me to implement the how. I just went into it then with all my heart. . . . Also, I was giving in-class assessment that mirrored very closely the format of ASAP. We [our third-grade team] tried to use similar language and similar forms and a log of what-do-you-think kind of analysis where the kids could write their own thinking and then support that with reasoning. We try to practice that kind of thinking and reasoning all through the year.
>
> I also think I've learned more about different tools of assessment by being an assessment mentor teacher for our school. I've done research and reading, and I've been learning about Stiggins, and I've taken some of his training. All of those pieces seemed to connect for me. Additional training also gave me knowledge about assessment tools and how to use them to measure learning.
>
> Once I knew what performance criteria were, I agreed very much with Stiggins that you need to share what the standard is with the student. You need to instruct them so that they know how. I like to let my students know what the standards are, what we want the product to look like and give them the rubrics before they start so that they have a map and a plan.

As far as curriculum alignment, I think in my early years I was more concerned as a teacher about "covering" the curriculum. Now I don't even like that word—cover. I just wanted to get my content out there, cover it as well as could be expected, and do the end-of-year assessment and hope my students performed well. There was no real integration. That's something I think I've evolved towards. There wasn't even a plan really. I hate to admit it. I think . . . I have that now. . . . It feels very comfortable now for me as a teacher. Our team's year-long theme was "Connecting the Natural World" and our philosophy was to connect as many pieces as you can. That gave us a focus. Then you let them [kids] know where you are beginning and where you want them to end and take them all the way. I have a lot more classroom discussions now. A lot more modeling on my part, and then, once they get the hang of it, they model it for each other. In math, particularly, a lot more explaining of strategies to solve the problem. "Talk through your thinking"—those kinds of prompts. I say, "What do you think?" a lot. And we listen to their answers more and respect opinions more.

The principal at Ms. Wonder's school described her orientation this way, using language quite similar to that of Ms. Wonder herself: "What pleases me about ASAP in terms of what we're doing now is seeing how the rubrics are being used, the criteria, the clear targets, the standards. It has all resulted in teachers working together and saying, what is the criteria for success here? What does quality look like instead of using their own judgment. It has forced our staff to really look at that and we've done that by grade level. . . . I'm seeing a great deal more articulation from one grade to another. . . . I have had the good fortune of hiring such strong people for my staff and then giving them the license to do what they do best and encouraging them. . . . The ASAP program is one we took very seriously. Teachers embraced it for various reasons and we made sure the students had opportunities to practice. Our students did very well as a result of making it a high priority. . . . It's my personal mission to learn more about student assessment and keep up with the changes" (Parish, 1986, p. 114). The principal made use of a district site-based management initiative as well as the district career ladder to concentrate efforts at her school toward building the capacity of her teachers toward competence in performance assessment and away from the district emphasis on criterion-referenced testing of basic skills.

Compliance only. By the second year of ASAP implementation, every district had complied with the formal requirements of ASAP. Each had filed a

District Assessment Plan and reported its early data on its ESRS relative to mastery of Essential Skills. None had refused to administer the Form D assessments. As for adapting the educational experiences they provided their students however, a significant number of districts had not gone further than compliance. The responses of district administrators to the telephone interviews were categorized as having either (a) embraced the ideals of ASAP or (b) conformed to its formal requirements. In the latter category, about one third of the districts had changed their scope and sequence and made some changes in language to appear consistent with ASAP, but had otherwise retained behaviorist curricula, texts, and pedagogy, retained traditional high-stakes testing, and failed to provide professional development coherent with ASAP. Such passive and conforming response seemed to be the result of lack of knowledge, lack of capacity, contrary beliefs and practices, or merely waiting out what some educators believed would be a short-lived government mandate. There was no evidence, however, of active resistance on the part of the educational community.

Compromise. Although we saw no instances of passive compliance in the case study or extension study data, we saw considerable evidence of educators reacting to ASAP in ways that compromised it. In his critique of top-down reform mandates, Combs (1991) wrote, "Things don't change people; people change things." In application to the ASAP policy, this adage implies that the policy is not translated into practice in a standard and straightforward way, but altered by people at the local level (Lipsky, 1980). The transformation of ASAP intents can be linked with misconceptions, imperfect understanding, and lack of fit with values and practices of higher priority in local districts and schools.

We found many instances of compromise in all parts of the data. For example, scores on the first year ASAP Form D were unacceptably low in one school whose teachers we surveyed. Reacting to the low scores, the principal proposed several means for raising them. One method was the creation of special "ASAP classes," in which teachers drilled students in ASAP Forms A, B, and C and materials that approximated Form D. By so doing, he created separation out of a policy meant to encourage integration. Thus, ASAP reform intentions were actually reversed in this school, supporting Combs's notion that, when it comes to centrally mandated policy, "people change things."

Another way in which educators compromised the reform intentions was through misunderstanding its rationale or through reinterpretation of its tenets

through alternative beliefs and value screens. During the first-year case studies, we encountered instances of what we called "dis-integration," in which teachers took an integrated unit, decomposed it into its basic parts, then taught each part separately until the students mastered it. We believe that this type of event illustrates that teachers failed to understand the principles underlying content integration or imposed on ASAP their own assumptions and practices. For example, teachers trained and experienced in behavioral or task analysis would believe that breaking down a complex whole into its component parts was the correct thing to do. But in so doing, they would deny their students the opportunity to make their own sense of the material, a fundamental tenet of constructivism.

On the survey, teachers gave further evidence of dis-integration and misconceptions. Forty percent of the teachers surveyed agreed with the statement "When I get an integrated unit or test such as ASAP Form A, I have to break it into its separate parts and teach the skills in each part so that students will be able to get the right answers."

> I think that the Arizona State Department of Education put the cart before the horse—testing holistically before many teachers were trained to teach that way. Because I have spent the last several years in school, I'm perhaps more aware of current holistic trends and more flexible in my teaching style. Many traditional teachers find ASAP very stressful because they aren't used to teaching this way. Although ASAP has forced a change in style for many, I've seen some teachers who are only changing long enough to get through the testing. (SOC)

In addition to compromise through misconception, the intentions of the reform were compromised in schools and districts that tried to implement ASAP while at the same time retaining local policies and practices that are inconsistent with it. In trying to respond to both, teachers not only compromise the reform agenda, but wear themselves out with their good-faith efforts.

> I believe that the ASAP assessment as a whole is more meaningful than our district skills test. However, we are required to do both, and our students are expected to perform well, which pulls me as a teacher in two different directions. I would much rather base my instruction on tests like ASAP. (SOC)

> Our curriculum is very much over-articulated. I mean, you've seen the course of study? You could stop a semi with it. (Extension Study, Pine)

One of the things a lot of the quality schools research has said is that it's really better to do a few things, really, really well than many things, you know, haphazardly. And our district seems to be interested in that kind of philosophy. But it doesn't come out in their decision making and their curriculum design practices. I think they need to commit to one test or the other. That's one thing I need to see. I don't really think it's a statewide thing that creates the kind of difficulty. At the district decision making at the top, we are constantly being asked to learn more and more and more things. But we don't necessarily learn to do those things very well. And it does take a toll. So that's really has less to do with the state and more of our district philosophy. Somewhere, you've got to say, "Ooh, wait a minute. Which of these things can we do? There's five things. We can do two. Which would be the best two?" And then let the professionals at each level decide. (Extension Study, Pine)

On the survey, nearly half of the teachers agreed with the statement "Using ASAP means that there are just more things that teachers at this school have to cover." And about half of the teachers also agreed with the statement "I feel like I am struggling to do BOTH a skills-oriented type of instruction AND a more holistic type of instruction."

Pine Compromises ASAP

Across the sprawling Peak Valley district from Desert School (home of Ms. Wonder) lies Pine Elementary, which is older and serves a less-advantaged clientele. Looking inside the two schools, one finds it hard to imagine that they are part of the same organization. In Pine, there is no classroom that fails to show the effects of what seems to be a centralized, standardized, top-down form of curriculum and pedagogy, held in place by a system of district criterion-referenced testing. The CRTs are the basis of student-grouping decisions, promotion decisions, and judgments about teacher merit. But unlike Desert, Pine is at the bottom of the Peak Valley test score distribution. The principal reports an encounter with the district assessment coordinator in which the latter asked, "Pete, is there life at Pine? Your school's scores have been the lowest in the district." The message was not lost on the principal, who subsequently began posting each teacher's CRT scores next to his or her name, emphasizing the importance of skills and scores, and building programs such as Success for All that promised to increase scores, even of disadvantaged pupils.

Also facing the unpleasant scrutiny of the public about the school's achievement test performance, teachers focused their attention on the district skills. In one of the classrooms we observed, most instruction consisted of worksheets and quizzes (teachers called them quizitos, and they were designed to make sure the pupils were keeping up with their skills), with the worksheets having been keyed to the CRTs and to specific textbook passages. Grammar, vocabulary, spelling, and writing (a standard, 5-step process was required) were taught separately—from separate texts, during separate time periods, with students arranged in different ability groups, and tested by different CRTs. The teachers believed that their reading program was "holistic" simply because they used trade literature books. But an observer well versed in constructivism labeled their reading program as "basalized," because real works of literature were selected from a prescribed set, and then the students moved (at a pace set to match their ability group) through a set series of activities (e.g., comprehension questions, vocabulary drill, etc., that corresponded to the CRT skills). Instead of the students making meaning from text, as one desires if one is a constructivist, the teachers and pupils had a prescribed and standard meaning thrust upon them in the form of the district course of study and tests.

One can easily imagine that ASAP would not drop into this culture comfortably, and so it went at Pine. No training on ASAP teaching or testing was provided to the teachers, and even the boxes of ASAP tests remained unopened until the last-minute scramble. During the first year of ASAP, no one paid it much attention, because the focus was elsewhere. During the second year, when the district began to align its assessments to state requirements, Pine still focused on the CRTs. It was commonly understood that the disadvantaged, minority pupils could perform well only on the CRTs and on skills-based curriculum, and that only the smart and well-motivated kids could do well on tests like ASAP or on holistic curriculum.

Misconceptions about performance assessment were common, with no training provided so that teachers could overcome them. The principal limited responsibility for the ASAP to the benchmark grade. As a result, third-grade teachers felt isolated and doubly burdened. They had to work toward both— two rather contradictory sets of expectations.

The consequences of ASAP reform intentions at Pine followed logically. Teachers reinterpreted the meanings of ASAP to fit their own experience and necessity. A young teacher relates the story of personal and pedagogical compromise:

> My first year I didn't know a whole lot about the CRT and I didn't really care. I just came in and was trying to keep my head above water. And I was teaching very whole language and very math manipulative-based. When the kids' scores came back, they didn't do very well at all, but I knew we had learned and I knew we had grown but the test didn't show it. So I was thinking, gosh I've got to start playing the game. I have to start teaching to the test. So the second year, I added a little bit more subskills attention and went with that. I got a little better. . . . and I still stuck with my main stream philosophy of teaching. Then in the third year, three of us teachers got together and said that we have to come up with some quizzes so that it just keeps abreast of their progress towards mastering those skills, and it's just a snapshot again for us to say that, okay, everyone's mastering multiplication so that we can move on. And they're a review thing too, because if we don't keep reviewing, then they're not going to remember everything.

Beginning with a philosophical receptivity to ASAP and holistic education, this teacher moved toward isolated teaching and testing under the pressure of district CRTs. She believed in the value of experiential learning and problem-solving math, but she realized that such teaching was not showing up on the math CRTs. "They didn't know how to do the paper and pencil stuff, they didn't know how to multiply on paper. They could draw an array, but they couldn't show you what 8 x 2 is, which was scary." As a result, she broke down all the math lessons into a series of work sheets to get them ready for the CRTs. She experienced dissonance between what she had to do and what she felt was the right kind of pedagogy, and gradually even that view was compromised.

Franklin Compromises ASAP

Through both misconceptions and conflict with existing beliefs and practices, Franklin School changed ASAP more than ASAP changed it.

Franklin is a large school in a small district in a part of the city that is simultaneously socially impoverished and wealthy in property. Because of its

industrial tax base, the school can offer substantial amenities: It pays its teachers well and can provide a computer for almost every pupil. Yet living conditions in its neighborhood are substandard and even sleazy. The pupils it serves are almost all poor and minority, and many are native Spanish-speakers. For a number of years, what happens in Franklin classrooms has been tightly controlled by the central district office. The curriculum and instructional practices are centralized and standardized. The official curriculum is unabashedly behaviorist and concrete-sequential. District administrators subscribe to an image of curriculum that is content broken down, bit by bit, sequenced and hierarchical, with each piece repeated until the pupil masters it. The official view of curriculum is held in place by the district criterion-referenced testing system (the principal says, "If it isn't on the CRTs, you shouldn't be teaching it"). These tests are given three times per year in reading, language, and math, with the scores of each teacher posted next to her name and used in teacher evaluation and merit decisions. The principal designates as "Master Teacher" those teachers with the highest rates of success. Furthermore, pupils' promotion from grade to grade is tied to their CRT performance. District educators attribute the need for such a structured curriculum to the degree of deficit and disadvantage of its pupils.

It could not be more clear that existing beliefs and practices at Franklin are not consistent with ASAP reform ideals, a condition that failed to change throughout the life of ASAP. For example, according to the dominant image of pupils at Franklin, the pupils were too academically deficient and "at risk" to profit from an integrated curriculum or teaching toward higher order thinking skills. When asked about the value of teaching in ways consistent with ASAP, a fourth-grade teacher snorted that whole language was impossible because "these kids can't write *anything* if they don't have skills. They can't transfer what they don't know. They don't know because they don't have life experience and they don't have language."

Although the teachers rejected ASAP reform ideals on the whole, the third-grade teachers were responsible for administering ASAP Form D and for getting the pupils ready to take it. But the principal was little help, even telling the teachers not to worry about the ASAP. As he claimed, only the CRTs were important because they were developed locally with the

particular population in mind. Furthermore, Franklin pupils had so many deficits in language and background that they could not be expected to do well on ASAP, in his view. Naturally, no workshops or other training was provided to the staff.

Nevertheless, some of the third-grade teachers had received training in holistic education under a previous administration and saw some value in performance assessment. When the state department of education informed the district that the CRTs were out of alignment, the district turned to these teachers to redesign district DAP testing. They spent all summer developing this assessment, which included problem solving and constructed response reading comprehension, and writing exercises. The items resembled those of ASAP but were meant to measure the district curriculum as well as the State Essential Skills. This assessment, given as a pretest in the fall, yielded scores that were unacceptably low. So the principal and fourth-grade teacher rewrote it, scrapped all that was like ASAP and returned to the familiar items measuring word attack skills, spelling, grammar, and computation.

The third-grade teachers at Franklin reported, "We tried to align the district test as best we could to ASAP, but they told us we had to do it with multiple-choice items, and that is kind of hard to do."

Drag. Although the ASAP reform intended to move schools in the direction of integrated curriculum and thinking pedagogy, it had contrary effects in some schools where such pedagogy is already practiced with great success. In part, this can be explained by the combination of aims within the single reform; that is, ASAP had accountability aims as well as reform aims. As explained and illustrated in Assertion 8, these aims sometimes conflicted with each other. Second, ASAP acted as a drag on constructivist pedagogy because the constructivist rationale in ASAP did not run particularly deep. That is, constructivists believe that reading and writing ought to be student-centered; yet ASAP provided a standard prompt for writing and a common story for everyone. Furthermore, in ASAP, the test developers did the subject matter integration, whereas constructivists believe that it is the student who must make the connections. As a constructivist reported:

Let's say if the state department assumes that how children learn is thematically and through connections, I think probably that's valid. Most people would agree with that. The assumption comes in, who makes the connections? Is it the child or is it someone else who sets it up in a thematic curriculum? And the assumption is that, "Oh, that's how kids learn, well, we'll put together units." They missed the point. There's an assumption that someone else can predetermine those connections and program the learning. And I don't think it happens that way. (quoted in Noble, 1994)

ASAP Drags Jackson

Jackson Elementary, which serves predominantly poor and minority pupils, was the subject of our earlier analysis of the role of mandated testing (Smith et al., 1989; also see the Hamilton vignette in a subsequent section) and agreed to participate in the Extension Study. At the time of the earlier study, Jackson suffered under high-stakes, standardized testing. Not only were its instructional practices misaligned with the ITBS and mastery testing models, but one could see a perfect inverse correlation in the ranking of test scores and the percent of pupils eligible for free lunch among the schools in the district. Alone in a high-stakes testing district that elsewhere tries to enforce a behaviorist pedagogy with criterion-referenced testing, Jackson is a fully-functioning whole language school that relies on portfolio and teacher assessment and attempts to minimize the deleterious consequences of mandated standardized testing on curriculum, teachers, and students. Now as before, Jackson exemplifies the ASAP reform ideals of integrated, thematic curriculum and pedagogy that stresses authentic problem solving, reading and writing. Without exception, the teachers are philosophically committed and exceptionally well trained. The principal exerts a powerful professional and moral force and procures professional development opportunities on a near constant basis. Teachers reflect and collaborate and indeed act as leaders to other professionals from other schools who seek their counsel. So far has Jackson progressed down this progressive education path, however, that the ASAP realities actually function as a drag. Responses of teachers in multiage primary classes to a group interview demonstrated this reaction.

> In our combined first, second, and third grades, ASAP has kind of been both a plus and a minus. I think it's a plus because the youngsters are brought up

through the three-year span, and I think become more familiar with the format. I think I like ASAP because it goes along with the way I teach. Our math is mostly problem solving, and most of the things we do are very integrated. We do a lot of writing; we do a lot of reading. ASAP fits more than any standardized test certainly did. My students write a lot, so ASAP was easy for them in that regard. But the one thing that was hard for them was the editing part, because they're really used to editing with an adult and a peer. And the time span on ASAP was unlike what we do here. Typically, when they're reading and writing process type of situations, it extends over, you know, a couple of weeks. So the ASAP was unnatural, it wasn't authentic at all. And one thing my kids found difficult was the time limit. Because they are accustomed to writing as long as they want to on a piece. And on ASAP, they had to write a story from beginning to finish, more or less in one sitting, and that was difficult for them to do, because they're accustomed to writing longer, really more involved. And I had to keep saying, you know, "keep it simple, have a beginning, middle, and end," because that what the rubric pays attention to.

And most of our kids are used to deciding on their own subject matter. We don't give them a story-starter like they get on ASAP, which really bothers them. They have a little less ownership over the story when you tell them, "you have to write about spiders." Plus, if you look at the rubrics, although they may help kids getting clarity to their pieces, the rubrics aren't really fair, because they kind of squish out a lot of individual voice style, and developing a voice is one of the main things we emphasize here. Just by its nature, a rubric moves you toward standardization and away from individual style. When you use the rubric, instead of really listening to what that person is getting at, you look for specific things, to give it a one, two, three, or four, a score like that.

So overall I would say that ASAP has had very little impact on our teaching and curriculum here at Jackson. Maybe it has other places, like at my son's school where they never had students write any essays, but not here. Even there, it hasn't really changed curriculum, because teachers don't fully understand how to teach that way. Instead it has just overlaid some more things to do, another thing to do in the system. But here, it just adds to the burden of testing and compounds the level of paperwork, particularly at third grade, which is an amazing amount.

ASAP Drags Hilldale

In many ways Hilldale, one of our case study sites, resembles Jackson, in that its educational practice already exemplifies ASAP reform ideals. Its

teachers are knowledgeable, its principal dedicated and determined to provide as much training and support as possible for holistic pedagogy, integrated curriculum, and authentic assessment. Unlike Jackson, Hilldale serves upper-middle-class and mostly white families and traditionally tops out the district test score rankings. The principal has used ASAP opportunistically to buck the tendency of the district to enforce a more traditional, skills-oriented style of teaching and testing. Hilldale is a paragon of capacity building: Teachers seek out their own continuing education through university courses in higher order thinking skills, writing process, portfolio assessment, conceptual math, using literature to enhance reading and writing, and the like. Teachers have sought out opportunities provided by the state and district to understand the ASAP rationale and procedures, so that they can be test-wise when it comes to administering and scoring the assessments at their school. Collaboration, reflection, and support are obvious in the relations of teachers and the principal. This capacity development, indeed, had started some five years prior to the advent of ASAP. As one teacher reported, "When the new principal came, we started ordering lots of literature (good quality novels, mostly), doing a lot of writing process, so our curriculum has kind of changed prior to the ASAP. Our thrust has been away from maybe so much textbook kinds of things, and more of an open kind of classroom curriculum."

During the first year of the study, Hilldale third-grade teachers expressed enthusiasm for ASAP, because they saw it as consistent with what they were already doing and a spur to move more traditional teachers to their own way of thinking. They believed that ASAP provided a better match with their view of curriculum than did the ITBS. Teachers believed the state officials when they said that ASAP would be process-oriented, that the students would be able to work together, and could work on the exercises as long as they needed, and that the results would not be used to compare schools and evaluate teachers. Some teachers didn't even use the word "test" when they discussed ASAP.

But during the second year, reactions changed and disappointment set in. The district had aligned its district testing to the ASAP, requiring that Forms A-C be used to report on school and district mastery of the Arizona Essential Skills. Although this might seem to be coherent change toward the ideals of ASAP, Hilldale teachers saw it as just the opposite.

My frustration this year has been the time commitment. Teaching a third-fourth combination like I do, I have to give 15 different assessments. It's ridiculous. Third graders are bombarded with Form D *and* the tests we give for district DAP. It takes the time of a whole unit of work from the year. You have to throw away everything for the entire month.

ASAP is supposed to be integrated units, and that's the way Hilldale teaches, but because we have to give all these separate tests, you know, Form As for all these separate skills, that one day we're talking about penguins and the next day we're talking about tarantulas. It's impossible to hold a unit together. It winds up being the opposite of integration. Teaching and testing for ASAPs is just teaching in isolation. Only our whole philosophy at Hilldale is *not* teaching in isolation. Form D is set up to be integrated, but the rest—they're forcing us to teach in isolation, which is combatting the whole theory behind ASAP.

You know, there are other schools that needed the ASAP to make them teach holistic. We didn't need it. We were already teaching that way. So I look at it now more as a hindrance because I was already doing the right thing. By taking up so much instructional time, I end up doing less of the right thing and doing it less well.

ASSERTION 4:   Responses Coherent With ASAP Intentions Were Centered in a Few Places Where Circumstances Were Auspicious

Is it possible to understand the variation in the local consequences of ASAP across Arizona? Are there patterns of response? The weight of the evidence of this study shows that the ground on which ASAP landed was not equally fertile and that certain characteristics of districts and schools could explain the particular consequences achieved. These characteristics included the *material and knowledge resources* available locally, the existing *assumptive worlds* of local educators and whether they were consonant or dissonant with ASAP, and the existing *culture of accountability and authority*. These dimensions were evident as a result of the multiple case study analysis during the first year of the study and were verified with the second-year data.

Material and knowledge resources. Although cost estimates for implementing ASAP proved impossible to obtain from either the state or the sampled districts, one can safely assume that the program did not come cheap. To purchase new texts and materials consistent with ASAP and to train teachers in a novel model of pedagogy required considerable investment. With no financial

resources forthcoming from the state, allocation of funds from the separate districts was an important influence on local response to ASAP ideals. Many district budgets were strained to the limit even without ASAP. In Roosevelt vs. Bishop, a federal court ruled that there were substantial disparities in the wealth of Arizona districts, contributing to unequal educational opportunities. Some property-rich districts were able to provide computers for every pupil and amenities such as covered athletic fields. Others were unable even to meet their payroll or repair falling ceilings. To the latter group, adopting new texts consistent with ASAP or training teachers in the writing process or thematic curriculum must have seemed an impossible dream. Valor, one of our case study sites, was one of them.

Valor Fails the Financial Test

In a district in one of the lowest categories of property wealth in all of Arizona, Valor serves mostly poor and minority pupils, including those of a nearby Indian reservation. A measure of its financial resources can be seen in the copyright of its reading texts: 1969. Funding for professional and curriculum development, toward any end including ASAP, is just a dream for teachers at Valor. In the initial year of our study, the assumptive worlds of the district were contrary to ASAP as well. SRA, a direct instruction basal reading series, was the prescribed curriculum. The district subscribed to a countywide mastery testing program. During the second year of the study, a new principal brought in new ideas and support for more constructivist teaching. The district abandoned the mastery testing program, but had not yet devised a new district testing plan more aligned with ASAP, nor could it afford to purchase one off the shelf as some other districts had. Having no funds for training or books, the teachers had to "teach" each other how to teach in a manner they thought might be consistent with performance assessment. Valor provided early release time to several of its teachers to revise the curriculum, and thus, as one teacher reported, "ASAP certainly has been a big influence in our school," particularly in promoting more writing and reading.

Yet, even in a more supportive climate, these teachers were on their own. There were no funds for hiring consultants, visiting more advanced sites, or

purchasing materials for training. There were no workshops offered, although one teacher had gone to an ASAP scoring workshop offered by the ADE.

> I think if we had a mentoring program, or much more of a release time—I mean a possibility for us to visit another teacher's classroom, or even a minute to just sit and discuss ideas. Then you see how much there is to offer the children. But that's not happening. I've enjoyed these Mondays for what release time we've gotten, but it's not enough to really work on units and how to teach things. But money is what drives it, whether you get to do these things or not. Because as soon as you start releasing time, it costs money. So when you're ill, it's very difficult to even get a sub, because the money for subs comes from the same pot as the money for release time to work on the curriculum. We just have totally limited resources, period. And it takes money to teach so that kids learn by doing. It's cheaper to teach out of a book and workbook. To get the materials to teach experientially, teachers have to spend their own money.

In addition to material resources, we noted that some schools and districts had access to no individuals with the expertise to train and mentor staff or select new materials consistent with ASAP. Lacking local expertise, even willing and financially able districts stumbled along the path toward performance assessment and integrated curriculum.

Assumptive worlds. The term *assumptive worlds* has been used to summarize the patterns of beliefs that characterize a particular site. The pattern that exists prior to the implementation of a mandate makes teachers receptive or resistive to it. Schools such as Desert, Jackson, and Hilldale were fertile ground for ASAP because they shared beliefs, ideologies, and values with the ASAP reform ideals. On the other hand, the case of Franklin showed how competing demands and ideology worked against coherent response to ASAP.

Franklin Fails the Test of Beliefs

In districts and schools like Franklin, with assumptions and practices quite discordant with ASAP reform policy, capacity building occurred only when individual teachers sought it out by taking classes and workshops independently. Few of them did. Although individual teachers may have held alternative beliefs, the dominant ideology about teaching and learning could not have been further from constructivist assumptions. The dominant view held that pupils are deficient and must be trained through a centrally

controlled, hierarchically arranged set of mastery learning experiences. Teachers' role is to transmit the curriculum and be accountable to authority, not to learn or explore. The district decides what teachers should know. Neither teachers nor students have much power to choose or to make meanings and associations for themselves. Franklin, it should be noted, is an anomaly in the sense that it exists in a community with much student poverty but much industrial wealth, providing nearly gaudy taxing capacity. These financial resources, however, were directed entirely to pursuits other than developing local capacity to implement ASAP intents. Instead, the money was spent on technology, and the aim was to provide one computer for every pupil. The dominant image of the pupil there was as deficit. Administrators and many teachers believed that such poor, ethnic and language minority pupils need basic skills instruction exclusively. Hence, the computers were used consistent with this image—as one teacher said, as electronic dittos. "Is there in-service training in this district? I'll say, but it's not about ASAP or integrated assessment. It's about computers. Technology. There's a computer on every desk. The desks are actually computer desks. And the training we receive is just the opposite direction from ASAP."

> For our children to do well on the ASAP, we have to change the way teachers in our district are currently teaching, and that hasn't happened. Teachers in our district should be using a more thematic approach. We should be using more critical thinking skills for our kids to survive. But the teachers don't know how to do that. Other than the third-grade teachers, our teachers don't have a clue.
>
> *And there's been nothing done to help you?*
>
> No.
>
> *No workshops?*
>
> No workshops.
>
> *Nothing.*
>
> Nothing at all that would help us on curriculum. It's all spent on computers and training us how to use them. But the basic curriculum hasn't changed. There aren't any new materials—we've just computerized the worksheets, so that the curriculum fits the CRTs and what's on the computer fits the CRTs and the training fits the CRTs.
>
> I mean, we're now moving into electronic dittos. We're moving away from worksheets on paper, but now we've got them on the screen. It's still just practicing on prefixes and stuff. Talking skill and drill. It's burps and bits. I felt like I was a marionette on a string. Do I have an explanation for why ASAP has

not had more of an effect in this district? It's very traditional here. People here think that whole language is too hard for the kids. But I don't know why they're so opposed to it. That has always remained a mystery. Anything that has the word "whole" in it, they don't want it. And CRTs are everything. We do them beginning, middle, and end of year. No one even mentions the ASAP. I never heard about it until two weeks before we had to give it. We had no training at all. And then we also have a principal who almost denies the existence of the ASAP. He even told me not to worry about it. He says the only thing that's important is the CRTs. And that is really important. He uses it for teacher evaluation and posts your kids' scores. There they are, listed under your name, and that does something to you.

Even in climates as philosophically hostile to ASAP as Franklin's, the teachers in benchmark grades were held responsible for ASAP Form D administration, and many for the quality of their students' results as well. These teachers had to fulfill competing sets of expectations. On the survey, more than half of the teachers agreed with the statement "I feel like I am struggling to do BOTH a skills-oriented type of instruction AND a more holistic type of instruction."

Culture of accountability and authority. A school was more likely to move toward practice coherent with ASAP ideals if it was not in the grip of high-stakes accountability. Districts with major investments in testing as a means of comparing schools, evaluating teachers, determining pupil tracking and progress through grades were not as likely to produce meaningful change toward ASAP ideals. Districts with authority over management and curriculum concentrated at the top, allowing little teacher autonomy, repressed local efforts to move toward ASAP ideals.

I believe that the ASAP assessment as a whole is more meaningful than our district skills test. However, we are required to do both, and our students are expected to perform well, which pulls me as a teacher in two different directions. I would much rather base my instruction on tests like ASAP. (SOC)

We are very fortunate that our district has developed practice tests that integrate a similar format to ASAP. These practice tests are very instrumental in preparing our students to know how "to connect" all the parts together. (SOC)

It was the DAP testing where ASAP intentions were most likely to encounter local practices that conflicted with them. State-encouraged performance-based

assessment was overlaid on local district assessment systems. Although some of these systems changed over the three years of ASAP implementation, local district assessment systems tended to be standardized (either norm-referenced or criterion-referenced) in nature.

Pine's lack of coherent response can be explained by its dissonant assumptive worlds and relative absence of knowledge resources. The district's top-down organizational climate and a powerful testing program that pointed teachers in the opposite direction from ASAP must also bear the blame.

Pine Fails the Accountability and Authority Culture Test

> No, the curriculum and materials that we used did not change in response to ASAP. The only thing we did was the practice tests, and that was just two weeks before the test was given. But, the district did change the CRTs to be more like ASAP, though they weren't nearly as good, at least they made a move in that direction. But it's clear that we have several different systems going on here, determining what we do—the district course of study curriculum, which emphasizes specific skills, the old CRTs, which grew out of that, the state essential skills, and now the ASAP. But if ASAP is going to have an impact at changing curriculum, then our district has to say, "Yes, we embrace this philosophy, and this is what we're going to do to help you. This is where we want the learning to head." I don't think that's there right now. We've had no in-service, only in how to administer the test. And I wouldn't call it really an in-service. We read through the thing ourselves. The person who gave it was just another teacher from a different school who had given it once before. No consultant, no nothing. None of the philosophy of teaching that way. And, it's not that we didn't get in-services on some things—there's a big district push on quality schools, computers, how to build a mission for the school. And those were mostly the choice of the principal. Is there someone here that can help us change our teaching so it will be more consistent with ASAP? Definitely not at the school. Possibly in the district, but getting that whole process activated I think would take a lot of time and effort. I think maybe that needs to be a component that if the state is going to use this [mandate this new test], they have to provide support for us to change. Because you're talking about a shift in teaching styles and testing styles. And that doesn't just happen by reading and trying and doing it. You need to have somebody actually give you feedback and kind of teach you and that sort of thing.

Micropolitics. We have argued in this section that coherent response to ASAP reforms could be explained by a pattern of local characteristics: adequate

financial and knowledge resources, already existing assumptions and beliefs about the nature of teaching, learning, and testing, and a relative absence of high-stakes accountability and top-down authority. Even when all the characteristics worked against the coherent response to ASAP, however, there were cases that defied this pattern. Having studied these negative cases, we discovered active processes of social construction at the local level that undid the top-down influences. These we identified as micropolitics.

The theory of micropolitics (e.g., Ball, 1987) views an organization as composed of several factions that contend with each other over distribution of goods, influence, and definitions of the situation. These groups, which may cross formal organization lines, are often in conflict over different ideologies and images of, for example, the right curriculum, the best explanation of pupil learning, the favorite way to assess achievement. This theory turned out to explain some apparent anomalies in our data. Our model of how it might work in a school or district considering a reform based on standards and assessments looks as follows: In the garbage can of innovations, something like the reform is swimming around in the minds of people in the field. The announcement of the reform becomes an opportunity to push some of these agendas at the local level. For example, one expects to encounter at most schools one teacher who believes in and silently practices content integration or higher order thinking skills. Maybe there is even a cell (a grade-level team, a partnership of teachers, or a teacher and a principal) in which the members think alike. With the announcement of the reform, this cell comes out of the closet to proclaim: "Here is what I've been doing all along. This is what I mean. Here is what you should be doing too." The object acquires valence, substance. Someone thinks about ordering materials that are conducive to teaching this way. Someone invites in a guest expert. Someone commissions an evaluation of the traditional practice that might be overturned. Competing ideas are more openly contested. Micropolitical maneuvering occurs, as group members attempt to win committee membership or positions of influence from which they can advance their agenda.

The evidence shows that some educators used ASAP opportunistically to advance a holistic agenda in a behaviorist climate. About a third of the teachers surveyed agreed with the statement "ASAP gives me moral support to do the kind of holistic teaching that I have been doing already."

Hamilton Elementary Wins the Micropolitical War for ASAP

Hamilton School was the one of the sites of an extensive qualitative study on the role of state-mandated testing in Arizona conducted prior to the ASAP policy change. At the time of the study, such testing consisted of the Iowa Tests of Basic Skills (Smith et al., 1989). Members of our team revisited Hamilton during the second year of the study presented here as part of the Extension Case Study. A group of third-grade teachers consented to a group interview. It had been five years since any of us had visited Hamilton, but from our extensive experience there and knowledge accumulated from over a year's intensive observations and interviews, we imagined that the curriculum, pedagogy, and school organization of that particular school was about as far removed from ASAP ideals as a school could get. Hamilton serves a predominantly economically disadvantaged community with high proportions of ethnic and language minority pupils and a substantial rate of pupil turnover, all of which were defined by teachers and administrators as problematic. Reacting to a prior reputation of a school in chaos, the district appointed a charismatic and forceful principal to Hamilton, one with strong allegiances to the Direct Instruction Model. Coincident with our earlier study, two years into his term, the principal had mandated Direct Instruction in every classroom in every subject, had procured substantial resources and staff development to support that model, and had hired an assistant principal whose only responsibility was training and supervision of teachers (he was once heard telling a teacher that "real men don't do whole language"). The public and the district perceived that the principal had turned the school around, and indeed, it won a coveted A+ award for being one of the top ten schools in the state.

Though we were alert for resistance and discrepancies among the teachers, we rarely found any such indication. Phonics training was evident everywhere, and authentic reading and writing instruction, if they were practiced at all, were relegated to afternoon down-times. During a faculty meeting when some teachers tried to introduce the topic of "thinking education" by showing a videotape of Glaser, Resnick, and others, other teachers ridiculed the notion that their particular student population could master the basic skills sufficiently to go on to higher order thinking skills. Another teacher, who said she had taken a lot of abuse from the principal and

other teachers for attending whole language training workshops, finally admitted to us in confidence that she tried to work in real literature to her classroom on those days when she knew the "phonics police" would not be visiting.

The district of which Hamilton is a part was then (and so remains) a high-stakes testing district. Scores were important, and the test burden made particularly heavy because of Chapter 1 requirements. Messages came from the Board and Superintendent to the school principals, and from them to the teachers, that test scores must go up or stay up. In that context, pedagogy followed test format, and drilling on isolated skills dominated teaching practice.

In contrast to this historical portrait, however, the teachers interviewed in the present study demonstrated the role of micropolitics in local adaptation to the ASAP policy intentions. Their interview is presented here as a conversation.

Did we change the way we teach as a result of ASAP? We restructured our reading and language program to go with how ASAP has taught, so using literature, having kids writing, working on higher level thinking skills, doing lots of problem solving. And *no workbooks*. Using a reading log, getting the kids to write complete sentences, watching proper capitalization, punctuation versus just fill-in-the-blank, one-word answers. Making connections.

Yeah. We have in mind all the time throughout our reading and language instruction the rubrics and the types of things the students are asked to do in responding to what they read and in doing their writing.

But it is interesting how it came about. The third-grade team, we were trained by the district. When ASAP was first coming along, we all received the Form As which were supposed to be the practice-use in the classroom, and so that we could see the process that ASAP was suppose to be modeled on, the different models for instruction as far as integration of instruction. We were all trained in those materials a couple of years before the testing actually began. We were also all trained in the rubrics and how they were to be scored so that we knew what the end products we wanted from our students, what they would look like. And then as far as right before the testing was to be done, we were all trained at the district level in exactly how the tests was to be administered. And over the years anything that came from the state on ASAP, from the time that it was first

talked about, I received, and then I passed on to the rest of the third-grade teachers. Then, over the summer, we put together this whole program. We wrote a proposal to change from our basal reader which was SRA [Direct Instruction] and to restructure the reading, language, and integrate writings and thematic units. And we worked during the summer on this and put together a packet, a notebook, on time lines and stories that we would choose from the basal reader, and put together kits that would match up with that and started looking around for resources. And then we did away with the workbooks and started using questions pertinent to the basic skills that they needed on district level and for ASAP. Like on the reading log, a question might be "Write a short paragraph on why you think he felt this way," or the main character felt this way. "What is the main idea?" "What's the setting of this story." And all those kinds of things. List five contractions and what they mean. From there, we developed that a little further by having a lot of grade-level meetings, lunch, after school. And we just, all the time, when we ordered new materials, we had these things in mind as a grade level.

So it turns out that third grade has kind of become the trail blazer for changing the instructional model at our school; that the way in which we've been instructing for many years with the Direct Instruction programs does not meet the needs of what's now being asked by the state. Third grade has trail blazed that, and we have worked on it as she said. But if the goal is to have all classrooms change their instructional models, it is not going to happen if they're only testing third and eighth grade, unless there is a real push either from the state, from the district administration, or from the school-level administration to change it.

I mean in all honesty, I don't think our principal knows exactly what is involved. How can I say this? First, thematic teaching, literature-based reading—these aren't his thing. I mean we've had to work hard and do a lot of talking and a lot of showing and demonstrating "this is what the students are being asked to do; this is the way we're being asked to teach; they don't fit; we have to make some changes." We were allowed to do that, but only through our efforts. If we had not fought for that and done the work to document why it's needed, how it's needed—I mean we documented item by item how things didn't fit and why we had to change our instruction. If we hadn't done that, we wouldn't have changed it.

There were so many things working against it. Because we're a schoolwide Chapter 1 program, and because we are a Direct Instruction demonstration site, we had to get an exemption from the manner in which we're evaluated at this school, we had to get exempted from that. And that took a little doing.

That's one, I think, one plus for ASAP, is that we, I think, as a grade level recognized the need to change the way we were teaching reading and language. And ASAP gave us kind of a leg to stand on. It became a powerful tool. It helped us. And I think that we were allowed to make changes because of it. I think we were given more leeway because of the ASAP requirement for third grade. I think maybe some of the other grade levels are having, you know, some trouble with it, if they really did, in fact, want to go to a more integrated way of teaching, because of the fact that they don't have ASAP. So ASAP for us has really kind of been a stepping stone to do some things that we wanted to do a long time ago, but couldn't because of the restrictions imposed by the Direct Instruction format.

Because of our going to thematic units, because of our going to year-around-school, that I think that's opening up a lot of areas for all the grade levels. Because of what we did and how well it was received by the administration, other grade levels were told, "Yeah, this is the way we need to be moving." If it hadn't been for ASAP and our work as a grade level to make that change, it wouldn't have happened. Now, it is there for everybody to do it. But I'm not sure they're going to. . . . Why? Because the change is too difficult. It's too much work. They like the old way, some of them. And again they have to mesh as a team, as we do. They don't have that bond. And some people, as I said before, their style is this way; and others have their style; and they're not always willing to change. And face it, there's a lot of people, it's a lot easier to read from the script than it is to develop your own integrated materials. It's already there, you don't have to think about it.

These remarks show that Hamilton, a place where dissonant assumptions dominated, where the authority and accountability culture worked against change toward ASAP intents, nevertheless changed remarkably in the benchmark grade, through a series of actions by a committed group of teachers. The teachers scrounged for resources for capacity development, proselytized, and overcame resistance. They used ASAP in a conspiratorial, subversive way to advance an agenda they had apparently held in silence for some time. They used the principal's lack of knowledge, his inattention, and his belief that ASAP was a temporary phenomenon and that the state would soon come to its senses and retreat to a mandate that was consistent with a behaviorist model of teaching and testing.

Micropolitics at Franklin took a decidedly different turn, one that ended up retarding response to ASAP rather than advancing it. Some third-grade teachers

believed in and practiced literature-based reading and writing and integrated curriculum. Yet public discussion on these topics was discouraged. When the teachers tried to incorporate more constructed response items in the district tests, the principal intervened. Although they worked all summer in building the new tests to measure the Essential Skills and actually administered them one year, by the next year, "we're back to testing individual skills and silent consonants."

At Franklin, the dominant image of pupils held in the school and district was as deficient—because of their impoverished circumstances, teachers believed, their pupils were not able to benefit from holistic instruction and needed very fundamental drilling on separate skills. The isolated teachers who believed otherwise retreated to their separate classrooms and closed the door. One said:

> At one point we had people that were willing to risk and willing to try things. But without that support, without that backup, without feeling like you're floundering, a lot of them quit trying. Just, you know, I think it was really scary for them. It's hard to take risks if you're not convinced. If you don't have a backup from the administration saying, "It's right. It's right. Now you just have to go forth and try it." But without that, when nothing's said, you think, "Am I right or wrong? Am I left or right?" And we don't talk to each other about it. There's no help from the district. But, as long as my kids do okay on the CRTs, my principal lets me do what I want, so I just close my classroom door and do all the integrated units and reading with trade books that I want. I figure if my kids read and read and read, then they will do fine on whatever tests they have to take.

ASSERTION 5:   Inadequate Capacity and Capacity Building Impeded Coherent Response to ASAP Intentions

"It has taken me ten years." This teacher's statement represents the extraordinary complexity of learning how to access authentic literature and knowledge resources, combine different content areas into thematic units, acquire pedagogical content knowledge sufficiently broad and deep to teach students how to find and solve problems, and to assess achievement in authentic and nontraditional ways. To make this change, however, is not simply a matter of acquiring new teaching skills, according to this teacher, but requires a complete change in philosophical orientation. She noted, "You have to understand the rationale for what you are doing; it's more than just adding to your kit bag."

Moreover, this new brand of pedagogy represents a radical departure from instructional practice that has been dominated over the past decade by high-stakes, standardized testing and centralized curricula and textbook (e.g., basal series, workbooks, and worksheets) series prescribed by states and school districts. Research (Smith et al., 1989) has suggested that, under the pressure to raise standardized test scores, teachers and school administrators narrowed the curriculum to fit the contents of the tests, isolated the content areas from each other, and used teaching methods that simulated the forms that the tests took. As a result, teachers have been "de-skilled." To the extent that this pattern dominates, as it did in Arizona, one would expect that the change to performance assessment, integrated curriculum, and problem solving would require a major investment devoted to re-skilling or capacity building.

Capacity refers to knowledge construction of individual teachers; that is, to genuine expertise and deep understanding of principles and practices coherent with the intentions of the policy. Capacity building refers to the provision and procurement, by any actor or agency in the system, of time, resources, supportive climate, and technical support, and to the removal of incoherences and barriers. That is, if the intent of the policy is to promote literature-based reading instruction, a district- or school-mandated program of phonics-only instruction would be considered incoherent with the reform. Building capacity would include removal of that mandate.

In spite of the obvious demands of capacity building necessary for accomplishing ASAP reform intentions, the state invested little. The only form of capacity building on the part of the state consisted of meetings to promote ASAP, and training of volunteer teachers, school liaisons, and district test coordinators in the rationale for performance testing and in methods and procedures for administering and scoring the tests. The ADE staff believed that Forms A, B, and C could be used for multiple purposes: as test preparation for Form D, as means for accountability for Essential Skills under the DAP testing provision, for instruction, *and* for staff development. ADE assumed that teachers would build their capacity to teach toward ASAP reform intentions merely through familiarity with Forms A, B, and C. Beyond that, the state provided no professional development.

The virtual absence of state provision for capacity building set the conditions under which districts, schools, and teachers attempted to cope with the program.

Because of lack of financial resources, lack of local knowledge about constructivist education, and existing beliefs and practices that conflicted with ASAP (see also Assertion 5), local capacity building was highly variable. The extent of variability is evident in the survey data. Forty percent of the teachers surveyed reported having had 5 or fewer hours of training in curriculum and teaching in ways consistent with ASAP. In marked contrast, 12% reported having had more than 40 hours of training.

In those arenas where local capacity building was taken seriously, teachers spoke with satisfaction about their training. For example,

> The ASAP course I took cleared the cobwebs and misconceptions. I wish I had had the course before testing my first graders last year. It might not have seemed so much a burdensome requirement. (SOC)

At the low end of the local capacity-building scale, teacher frustration was obvious from the following excerpts and clearly explains why the effects of the ASAP program were so variable.

> I understand the concepts behind the ASAP and agree with them strongly. I especially like and understand the scoring rubrics and the concept of measuring the Arizona Essential Skills. However, teachers need more instruction and practice on integrated learning and problem solving. The students need more planned activities that include this. Actual teaching materials are needed, not just tests for evaluation.

> If the state wants us to use these types of test, they need to provide money to districts to print the assessments (Forms A, B, and C). They also need to make the A, B, and C Forms integrated like the Form D. New materials are required to teach in this manner, so the state needs to provide the funding for these materials. The state also needs to address the funding formula for districts and provide more equal funding.

> We need texts and materials that are compatible with ASAP. This is one of the great problems of not having adequate support materials to teach with for most of the subject areas. . . . Therefore we manage with what is given to us. (SOC)

> Teachers need more training in process style teaching and evaluation; and they need more time to communicate and share with each other. (SOC)

Especially for reforms with no capacity building or incentives, time is a crucial element—time to process the reform (come to a definition of it), to marshal resources toward its implementation, to learn what it takes to teach toward it, to experiment, fail, make advances, succeed, discuss, collaborate. But for the ASAP program, there was simply *inadequate time* to do these things.

> I think lack of understanding at the district level and the "rush" changes at the state level have caused a great deal of aggravation and have turned some against the idea, which I feel is great but faulty in execution. (SOC)

> I believe it was implemented too quickly. Before we had the confidence to attempt small parts of ASAP our report cards, essential skills, etc., were changed. Holistic teaching was not practiced at this school before the curriculum underwent a change. With many outside duties teachers have to cover, there is little time to absorb rationale. The whole idea of critical thinking is highly needed. However, the test assumes that teachers can manage it alone. (SOC)

> The major problem in education here is LACK OF TIME. We never have time to talk to others, let alone plan and work together. (SOC)

The failure of the state to fund local programs for capacity building (or to adequately staff itself for the technical requirements of administering the system) was joined to the existing *financial inequities among districts*. As one district administrator reported, "It's hard to fund professional development when you don't have enough money to unplug the toilets and remove the asbestos."

> There needs to be more money for schools to buy the resources they need in order to teach in such a way as to be compatible with ASAP. (SOC)

As shown in Assertion 5, lack of financial resources inhibited coherent response to ASAP reform intentions by limiting local capacity building. Valor, the rural school in our case study design, offered a telling example.

Lack of wealth was not the only reason that local capacity building varied. Many districts were able financially to fund adequate professional development and acquire consistent materials, but needed district personnel knowledgeable about what form this professional and curriculum development ought to take. In

those districts, financial resources were spent toward other aims besides those of ASAP (see Assertion 5).

On the survey, teachers were asked a series of questions about what they had been learning as a result of ASAP. In response, 24% reported having learned how to teach with thematic units and integrated subject matter, and 26% how to teach so that students will be able to solve complex, real-life problems. Only 18% agreed with the statement "Because of ASAP, our school has brought in consultants and experts in curriculum and instructional methods compatible with ASAP." Only 19% believed that teachers had ample time to experiment, collaborate, and reflect on possible changes in their teaching. About one quarter reported that their schools had adequate funds to purchase texts and materials consistent with ASAP. And finally, only 19% reported that adequate professional development had been provided to teachers to make changes consistent with ASAP.

The survey inquired about the amount of training the teachers had received. For the amount of training they had received about ASAP itself, that is, about the program and about how to administer and score the performance assessments, 60% reported having had 5 or fewer hours. Twenty percent reported having between 5 and 10 hours. Sixteen percent had had more than 10 hours of training in ASAP itself. The median was about 3 hours. Teachers in districts with large concentrations of poor and minority pupils reported substantially more training on ASAP, almost one third of a standard deviation unit. The effect size was .30 for the difference between the means of teachers in high and low minority schools. Teachers in rural schools also had a nearly one-quarter standard deviation advantage over suburban teachers in this kind of training.

The teachers were also asked to estimate the amount of training they had received in curriculum and teaching in ways consistent with ASAP; for example, integrated, thematic units, higher order thinking skills, literacy, problem solving, and the like. Forty percent reported having 5 or fewer hours over the past two years. Sixteen percent had had between 6 and 10 hours. At the other extreme, 12% reported having had more than 40 hours of training. The median number of hours was only about 8 hours of training in how to teach that which ASAP measures. Unlike the difference in pupil composition and type of community on ASAP training, there was no corresponding difference in the amount of professional development given to teachers in rural and high-minority districts.

Teachers were also asked to rate themselves in their knowledge of 18 items considered important to ASAP implementation. For each item (e.g., writing across the curriculum, higher order thinking skills), teachers marked whether they had no knowledge of the item, awareness only, good understanding, or expert knowledge.

About 60% of the sampled teachers reported they had "good understanding" of the Arizona Essential Skills, ASAP scoring rubrics, integrated units, problem solving, and cooperative learning. About half reported that they had "good understanding" of writing process, literature-based reading, hands-on science, student-generated research projects, and portfolio assessment. About 40% described themselves as having good understanding of conceptual math, probability and statistics, and alternative assessment. The rates of teachers describing themselves as "expert" in these topics were quite low, exceeding 10% only on ratings of knowledge of integrated units (16%), writing process (16%), literature-based reading (18%), problem solving (15%), and hands-on science (15%). On the variable "Self-Rated Knowledge" (Table 1, Appendix D), teachers in suburban districts had a one-quarter standard deviation advantage over teachers in rural districts.

District administrators were asked about the kinds of professional development provided to the teachers. Their responses were categorized as "related" (e.g., higher order thinking skills, thematic instruction, writing across the curriculum, authentic assessment) or "unrelated" (phonics, discipline, TQM) to ASAP. One quarter of the districts provided no ASAP-related professional development to teachers, and 17% provided three or more such activities. The median number of ASAP-related activities provided by the districts was one. About three quarters of the districts had provided some form of training in ASAP itself; that is, in how to administer and score it or some related form of performance or portfolio assessment. About 18% of the districts' professional development was judged to be close or very close to ASAP ideals. About 30% of the districts' professional development was judged to be opposite to or irrelevant to ASAP ideals. That is, many districts were providing some, perhaps even extensive, staff development, but it was on topics that built no local capacity to respond to the state reform. When asked to rate their teachers on their knowledge of curriculum and pedagogy related to ASAP, about two thirds of the district administrators rated teachers as "somewhat knowledgeable" about integrated

curriculum, holistic instruction, and higher order thinking. Only about 15% rated teachers as "very knowledgeable" in these areas.

Capacity building of students: Opportunity to learn. One of the most important issues in measurement-driven reform is whether students have courses and experiences relevant in kind and sufficient in amount to allow them to perform adequately on the assessments that the government agency mandates. This survey asked teachers to rate several conditions at their own schools that would indicate whether their pupils had adequate opportunity to learn that which ASAP entails. Results show that perhaps one quarter of the students lack opportunity to learn higher order thinking skills, to practice on alternative assessments, to write about their solutions to math problems, or to engage in lifelike learning experiences. Two thirds of teachers rated as very or somewhat true of their students that they have adequate opportunity to combine reading and writing, and to write about science and social studies topics.

The factor analysis revealed a consistent, underlying variable made up of opportunity to learn and resources supporting student-initiated investigations. Analyses of variance resulted in mean differences on Opportunity to Learn due to pupil composition, type of community, and district resources. Teachers in districts with low rates of poor and minority pupils rated opportunity to learn higher than did teachers with high rates of poor and minority pupils (effect size = .23). Teachers in suburban districts rated their students' opportunity to learn higher than did teachers in rural districts (effect size = .37). Teachers in districts with high taxation capacity/pupil rated their students' opportunity to learn higher than teachers in poorer districts (effect size = .35), and nearly one-half standard deviation difference separated the richest from the poorest districts on this variable.

ASSERTION 6: State Inattention to the Technical and Administrative Adequacy of the Assessment and Accountability System Impeded Coherent Responses to ASAP Intentions

Because ASAP was both an instrument to reform schools and a psychometric instrument, it must be subject to the standards and practices of the psychometric community. Yet the state devoted remarkably few resources to address the psychometric properties of ASAP Forms A-D. Nor was adequate time available to develop the tests, pilot, review, and revise them, to understand the

meaning of the scores, or even to work out the inevitable kinks in printing, distributing, administering, collecting, and scoring the assessments. The evidence from this study is convincing that the flaws of ASAP as a testing instrument worked against ASAP as a reform.

Performance assessment was then and still is an innovation whose technical aspects are unknown and controversial (e.g., Linn, 1994). Issues such as the number and representation of tasks to include in the assessment, the appropriateness of general versus specific scoring rubrics, the number of scorers necessary to generate stable scores, whether concepts and techniques of reliability and validity that are suitable for standardized tests are also suitable for performance tests, and methods of combining component scores are still far from solved. There is no consensus over whether performance assessment and standardized assessments are commensurable, or even whether process learning theory and outcome assessment of any variety are commensurable. Performance assessment generally, and ASAP Forms A-D in particular, are works-in-progress.

Essential to understanding the reactions of teachers on the surveys and interviews is a review of the history of ASAP as a psychometric instrument. The Arizona Department of Education had spent a number of years developing and refining the Essential Skills. Yet only one year was devoted to the development of ASAP performance assessment before it was introduced into the schools. Department officials were far more highly trained and experienced in the practice of standardized testing than in the practice of performance assessment. No experts in performance testing were hired, though some were consulted periodically. There was no functioning technical advisory committee (one was appointed but it met only once). Contracts for test development were let to Riverside Publishing, which had originally developed Form A and then conducted a technical analysis of its pilot test results. Based on the pilot test and analysis, Form D-1 was developed and administered without benefit of its own pilot testing (recall that Form D presented integrated subject matter whereas Form A tested separate subjects). Similarly, Form D-2 was administered prior to its correlation with Form D-1, even though they were meant to be parallel tests. No independent technical analysis was ever conducted or sought by ADE during this time. Meanwhile, the rubrics to be used for scoring the assessments were constructed along the way and changed several times, as were the systems for combining scores into scales and converting scales into normative metrics. Changes in

administration and scoring instructions were frequent. A district administrator tells the story this way:

> By the spring of 1993, the time the state Department of Education was at an implementation point for the first administration of Form Ds, we were positioned to implement the new measures with ease. But implementation turned into a nightmare. Deadlines kept changing, test booklets arrived late from the state Department, they came wrong, they came printed poorly. Some tests arrived only 24 hours in advance of the mandated date for administration. Teachers were required to use calculators and physical props or manipulatives that had not been announced or planned for. District support staff spent a large number of hours in overtime in order to provide the correct lengths of string, cardboard, etc., for third- and eighth-grade teachers poised to administer the tests. People at the state Department who had been assigned roles of coordination were viewed by district administrators as happy missionaries lacking in knowledge and skills about how to coordinate the thousands of logistical details associated with a major implementation like the ASAP. (District testing administrator in Parish 1996 study)

Teachers, even those who were supportive of ASAP reform ideals and performance assessment, became frustrated and confused.

> The directions handed down from one person to the next (beginning at the state level) have consistently changed from day-to-day and semester to semester. No wonder there is so much confusion and misunderstanding. It leaves most of us teachers wondering if the "bureaucrats" at the state office actually know what they're doing! I'd like to see *them* administer and score these tests! (SOC)

> I believe the ASAP is a very unreliable test. . . . The instructions are so complex and unclear that students become easily confused. Frustration level is very high!!! Students become burnt out after a short amount of time and basically rush through in order to finish it. (SOC)

> Many of the instructions, particularly on the math assessments are so complicated and unclear that even *I* don't understand what the children are expected to solve. Too many-step directions are not developmentally appropriate. (SOC)

> Our teachers are *still* arguing about what those directions were asking for. It's enormously frustrating for both us and the kids. I consider myself a good math teacher, and my kids learn to love math. But when I find kids crying because they don't understand what ASAP is asking them to do, I really feel bad. (Extension Study)

I feel the concept of ASAP is good but the math assessments are terrible! I feel most adults would fail the tests. The questions are too difficult to understand. After reading them you do not know what the problem is asking you. Students who have low reading and writing skills but very high math skills do not have a chance of passing the ASAP. (SOC)

It took so many hours to cut out and prepare the manipulatives for the Form A test in second grade that it left us resentful. Then it took us over 47 hours to score the tests in my class. (SOC)

How can you trust a test when the scoring is so subjective? (Extension Study)

I teach math up through pre-calc. and have always taught critical thinking to my students. I and many of my colleagues strongly support the *goals* of ASAP, but after two years of teaching practice assessments and training as a scorer, I have found that what the seniors experienced on test day was a far cry from the ASAP goals. We're not talking about minor bugs in the test, but well-intended questions that totally missed the mark. The questions are totally absurd. Is this incompetence or lack of management and common sense? State department ASAP folks have told various protesting math departments that there were *no errors on the test*. Bull feathers!! Please give us a chance before the ASAP program suffers any more damage. (SOC)

ASAP is not a valid testing tool to evaluate Arizona students. Students are frustrated not because they do not understand basic concepts, but because of the testing format. To not let students graduate because they do not score well on ASAP is horrible and unfair. (SOC)

If you can't read that well, you're pretty much out of luck [on both writing and math]. That's the bad part. (SOC)

When the results from the initial round of Form D assessment were reported, there was considerable confusion as well. A pupil's performance on the integrated assessment was disaggregated into separate scores for reading, writing, and math. The reading score was a simple sum of five components, each of which was comprised of rubric scores varying from one to four points. Thus reading scores could vary from 0 to 20. The sum was then converted to percentile equivalents and reported both ways. Since this reporting system was unfamiliar to both teachers and parents, many tried to impose their understanding of similar metrics

for standardized tests on the ASAP Form D results. As one teacher at Pine noted, "No one understands the score, really."

> I have not really used information gained from ASAP assessment at all. I tried to, but when we looked at the scores you didn't know if a twelve was good or bad. There were numbers given, but you didn't know how to interpret them, so I really couldn't tell if that was good or bad. You could tell the state average and how we did in comparison. I think the state's still looking at now what does it mean if you get a twelve or thirteen. So I guess it doesn't make a lot of sense to me. (Parish study, 1996)

The survey evidence on frustration over the inadequacies of the testing system corroborated information from the qualitative elements of the study. On the teacher survey, 68% agreed or strongly agreed with the statement "Confusing directions and procedures pose problems for the administration of ASAP." Only 12% of teachers disagreed with that statement. About half of the teachers in the survey distrusted the subjectivity in the scoring process. Nearly 70% believed that students with poor reading scores would not be able to perform well on the math score even when their math skills were adequate. Yet many believed that ASAP provided an accurate measure of students' ability to read with understanding (45%) and to write effectively (51%). Only 31% believed that ASAP accurately measured students' ability to solve real-world problems in math. Still fewer, 14%, believed that ASAP scores accurately reflect the quality of education a student has received.

Because the dominant definition of ASAP held by teachers was as the tests personally given, and because of the many technical and administrative weaknesses of the tests themselves, much energy of the reform was absorbed in reactions to these weaknesses, as if teachers were unable to get past them to search for legitimate ways to respond to the reform of instruction intent. They could not endorse the policy when its instrument seemed so flawed and caused them so much frustration.

What is clear from these comments is that the technical problems were much less related to the *form* of the test than they were to the problems of directions and procedures involved in administering the tests. One can make a good case that the latter problems might have been solved if sufficient time and trials (and reflection on them) could have been devoted to ASAP.

ASSERTION 7: The Reform Intention and the Accountability Intention of ASAP Conflicted With Each Other, and the Conflict Impeded Coherent Action

Analysis of the beliefs and actions of the policy makers who instituted ASAP revealed the existence of two sets of intentions (Noble, 1994).

The *first set of intentions* involved the use of ASAP for accountability, the definition of which was fixing responsibility for measured academic outcomes. In other words, the test was considered a tool to ensure that teachers taught what the Essential Skills specified. The policy makers with these intentions believed that the test would force teachers to change their level and direction of effort through a system of rewards and punishments based on the level of achievement scores their students attained.

An ADE official put it this way:

> This is all predicated on the following, a belief that every student can learn and will learn. . . . Every student who comes to our schools and leaves our schools must have skills that are going to enable them to be successful in whatever they want to do. What we are concerned about here is that students learn and what they learn. (Noble, 1994, p. 74)

Policy makers with this set of intentions equated learning with test scores and teaching with transmitting prescribed curriculum and for "making kids skilled" (Noble, 1994, p. 75). The competencies for which teachers would be held accountable were specified in the Arizona Essential Skills and embodied a mastery learning perspective, as these officials indicated:

> Those documents [Essential Skills] really define what students ought to know. Those are the documents the districts use for curriculum alignment. Those are the things that we are required to make sure are a part of our curriculum and included in what we teach.

> [A] good curriculum alignment plan would allow for decision making. "Look at the Essential Skills and decide who's going to teach what, when, how, and to what degree and who's going to pick it up from there the next year."

> The teachers need to know that the Essential Skills are the framework that they have to work around or within. I mean, they have to get the kids ready for the end of

3rd, 8th, and 12th grade to pass those examinations. That's what they're working for, nothing else. (Noble, 1994, pp. 81-83)

The second set of intentions was based on the notion that ASAP should be an instructional tool to guide teachers to a different kind of teaching. The assessment package could encourage teachers to teach holistically, but exclusive of staff development, the tests alone could not cause the changes that were needed. The following comments of ADE officials best capture this view:

I think the best we know about instruction [encourages] thinking about the purpose of what they're doing, being able to apply the knowledge that they've acquired to a real life situation. So that they can expand their own knowledge—not just information but use of information. (Noble, 1994, p. 81)

The ASAP as policy has been called the most profound incentive for change there has been in Arizona. Policy obviously cannot mandate what matters, but perhaps it can establish conditions for what matters." (Noble, 1994, p. 91)

The second set of intentions for ASAP is based on the notion of teaching as co-learning and coaching rather than transmitting a standard catalogue of skills to pupils. Teachers would have a certain degree of autonomy to pursue content according to pupil interests and sufficient space, time, and resources to experiment with teaching and testing forms, to reflect and collaborate with other teachers and curriculum experts. To change, teachers must "feel safe" without the personal risks of high-stakes tests attached to results (Noble, 1994, p. 77).

The dual set of intentions at the state policy-making level was embodied in the organization of the ADE. ASAP responsibility was divided between the "Pupil Achievement Testing Unit" (the members of which defined ASAP as an accountability program) and the "ASAP Unit" (the original members of which defined ASAP as a tool for changing teaching toward the tenets of constructivism and making assessment more authentic).

The following are comments of the Pupil Achievement Testing Unit:

I see ASAP as an example of a creative reform initiative that has at its core accountability for student achievement.

[ASAP] was the legislators' way of saying, "Educators, you're going to teach kids and we're going to know what they learn."

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.

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. . . . 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. (Noble, 1994, pp. 87-88)

The following are comments of the ASAP Unit:

[ASAP] is a very, very different approach to what we call testing, which we're now calling assessment. We're saying that it's not something that you stop teaching for, it's not something that you do power teaching for. This is something that's embedded in the instructional process day after day.

I would hope that assessment would become integrated into instruction. That we would get away from one day [that] we're testing all the students on a particular thing.

Our thought was that if we developed a system of assessment that was based on quality instructional methods, that teachers inherently would work themselves into better instruction by using and understanding the assessments. (Noble, 1994, pp. 87-88)

Why is this distinction important? Each set of intentions suggests a quite different assessment system. If a test is to be used for accountability, it must meet standards of comparability, objectivity, and fairness. Such a test would emphasize reliability of results, and consequently the content and task requirements would have to be standard across students. Scores must not vary according to who administers the test. Test content must be private and secured against the possibility that teachers would prepare their students inappropriately for it. Scoring must follow standard procedures and rubrics, so that scores would mean the same thing everywhere. Scores should not therefore vary according to who scored the test or where it was scored. Subjectivity, teacher judgment and

interpretation, and pupil choices would have to be reduced. The testing system would have to emphasize outcomes over learning processes, because processes are more difficult to measure objectively.

An assessment system following the second set of intentions must meet a different set of standards, de-emphasizing reliability and standardization, for example. Instead, the content of the test would have to be open to scrutiny and critique of all stakeholders. Over a period of years, therefore, the contents and tasks covered by the test and the assessment system (scoring, combining, reporting) would improve as a result of this scrutiny. There would be variations in content and tasks, because context and pupil choice would be important elements. Nor would the test be given on the same day to all students, as learning projects likely vary in length. Measuring process would be as important as measuring outcomes. Test results would be used to plan instruction. Test scores would be integrated with teacher judgment and other indicators of achievement. Stakes would be low on numeric scores.

In the history of ASAP, it is clear that the accountability intention (and the Pupil Achievement Testing Unit) won the micropolitical war over the definition of ASAP. Form D was designed to serve an "audit" function (making sure that the teacher-scored DAP testing did not distort the true picture of a district's achievement performance). It was administered in a standard way, on the same time schedule for all students. Results were scored centrally by the state generic rubric. The contents were secured, protected by publishers' copyright from widespread scrutiny, and given under such a tight timeline that stakeholder interests were ignored. In addition, the content and tasks of Form D were so standardized that an observing psychometric expert noted:

> . . . a lot of the items are very mechanistic. They're right-wrong. They're really the simplest form of constructivist response. . . . They could easily be transformed into multiple choice. (Noble, 1994, p. 115)

Yet, until the state published the results of the first year's administration of Form D, most teachers had only a vague understanding that ASAP was not only an innovative assessment tool but also had to serve an accountability function. The ADE's ASAP unit repeatedly disavowed the state's intent to use ASAP scores to compare districts and schools or to determine pupil promotion and graduation.

Newspaper headlines (e.g., "Tests say schools are failing"; "Math scores in state 'distress' officials") rudely awoke most educators to the accountability function of ASAP. ADE had reported out results by schools, and newspapers had transformed them into rankings. They also quoted the superintendent's distress that schools were not teaching adequately or in accordance with the Essential Skills and how students learn. An ADE official provided context:

> What's important to me now is that ASAP . . . is now being reported out by school, district and the state. Now, that's where the action is. That's where the stakes are. (Noble, 1994, p. 118)

In January of 1994, the State Board of Education added to the poorly foreseen high stakes to be attached to ASAP when they passed a resolution that high school graduation would be tied to ASAP performance beginning with the ninth-grade class of 1996.

The surprise that resulted from the triumph of accountability intentions over alternative ends emerged repeatedly in the data. For example:

> I feel the concept of ASAP is good but . . . we were told the tests were to be used for assessment not comparison, yet you see scores compared for districts and schools by the newspapers. What happened! (SOC)

The internally conflicting sets of intentions marred coherent response. Teachers at Valor, for example, moved gradually toward acceptance of ASAP reform ideals. Yet they objected to the stakes to be attached to ASAP test results. Since the state would eventually tie high school graduation to the test results, the district asked teachers to sign a contract to ensure that all students would be able to pass the ASAP test:

> I'm not feeling comfortable with that. It's not possible. I have no problems with accountability. We all need to be accountable. But I think you cannot set criteria like that and expect that it will be a reality, because it won't. (Extension Study, Valor)

> I believe in the philosophy of teaching, learning, and assessment behind ASAP. However, whenever such a complex program is mandated there are intrinsic problems: It gets tied to the legislature's goals, it gets watered down so that format is more important than philosophy, staff development is severely lacking (and to

bring about any *real* change, it must be done at the classroom level), early results get too much attention. The state needs to allow 5-7 years to see results. (SOC)

Form D and the other forms are dealt with by two entirely different agencies. . . . They don't even talk to each other, they don't even know each other's names. And I was really shocked, because Form Ds are nothing like the As, Bs, and Cs. And I was thinking, "You know, are these people really thinking that the As, Bs, and Cs are going to prepare us for the Ds?" (Extension Study, Hilldale)

They call Form D an audit. That cracks me up. Actually, it's kind of demeaning, because our district assessments, we give them and score them and turn them in. We say this is 75%, so the kids are mastering these essential skills. Then the state comes in with this audit, a dipstick they call it, to make sure that we're really doing what we say we're doing. It's not to improve instruction. It's an audit to see if districts are actually doing what they say they're doing. And when I saw the ASAP scores in the newspapers, after the state had promised us that ASAP would not be used for comparison purposes, to compete with other districts, I was infuriated. Plus, it's such a poor test; it has so many errors in it, that the scores mean nothing. But when people pick up the newspaper and see Hilldale math scores low, they don't think, "Well, it's a new test, or it's a bad test." They think that we're not teaching any math here. (Extension Study, Hilldale)

The contradictions at the state level produced conflicting messages to teachers. In addition, ASAP fell into an existing district culture that either emphasized or downplayed testing for accountability, a condition that either enhanced or impeded local change toward ASAP reforms ideals (also see Assertion 5). In districts that emphasized testing for accountability, the ASAP program was overlaid on their other assessments. The overlay increased "test burden," the amount of time teachers had to devote to test preparation and administration. In addition, teachers felt that they had to be responsible for multiple sets of achievement outcomes, some of which competed and conflicted with each other. Conflicting state and district testing requirements emerged at Pine (see Assertions 4 and 5).

I knew that ASAP was something that was coming up, but there were so many other things going on throughout the course of the year that I never thought about it. Then, all of a sudden, the district says, you've got to give this, kind of a last-minute thing. So, it was, like, we're giving it in two weeks, so we better start getting ready for it. And so we used a lot of the practice tests. The test itself, I thought was pretty neat. I liked the activities that went with it. I thought it was a lot more effective than like

a fill-in-the-bubble type of test. But as far as like integration into my classroom, you know, I really didn't integrate it into my classroom. It's hard, you know, because ASAP conflicts with the district CRTs, so it's hard to get the kids ready for both of them. We're pretty much overwhelmed with materials to practice specific skills that the CRTs cover. Now I hear that the state has told the district that the CRTs are no longer acceptable tool for assessment, so the district is replacing them with something that looks like the Form As. The subjects aren't really integrated, but at least the kids don't bubble in all of them, they write out their responses for some of them. But we still are required to teach all those specific skills. I really don't think there's enough time in the school year to teach all those things. (Extension study, Pine)

In districts with strong accountability culture, ASAP was used for accountability purposes along with existing tests, as teachers at Hamilton (also see Assertion 5) expressed:

ASAP Adds Accountability to Hamilton

What messages do we get from the administration or from the public in terms of the importance of ASAP scores? We didn't even get the scores until December. Based on last year's kids, we got the scores in December. That's a message of how important it is to the district. They're more interested in the district Basic Skills tests. But, you know, if they're published, you got to know the administrators want them to be high, they want to show off the school. But our district administration, even towards the end of last year, thought ASAP might go away and was not pushing it real heavily. We're not being come down on. I mean the district is not coming down on the principals, and the principals are not coming down on us yet on ASAP. Now that we know that ASAP is here to stay, if the state starts reporting scores on a teacher-basis, school by school or classroom by classroom, then that will probably change. Because then you will have teachers looking at them and thinking, "You know, those are *my* individual scores; they're under my name and this is what they look like." Then you're going to have a problem. That's the same thing that happens with all the other tests—ITBS, the basic skills tests, they're under your name. And if you've got a 64% mastery showing there, instead of 75% that's required, you better have a darn good reason why. Because the district administration pressures the school administration, and they pressure the teachers. When that happens with the ASAP, then that will bring even more pressure on third-grade teachers and third-grade students, and it shouldn't be. Then you'll have third-grade teachers bailing out, trying to get into other grade levels.

Even though we've got ASAP, the pressure on ITBS and CRTs was not changed a bit. Oh, yeah, heavy-duty pressure. And we're expected to have 80%. . . . They want us in the top quartile. And if they're not there, there'd better be a real good reason why. And then you're going to plan for how you're going to change that. If this year's scores come out not in the top quartile, then we will have to have a plan next year to make sure that doesn't happen again. If ASAP is going to continue. . . . I mean last year, it was anonymous basically. It was just the whole school; and the district's score was published as a districtwide score. If it's going to be broken down by school and then broken down by teacher, then that will also become an eventuality.

Pine teachers felt the same:

Right now our district tests align themselves really well with the Iowas. And when the newspaper reports come out and they want to publish Iowa scores, you're going to get better results by emphasizing those CRTs because then those skills will translate into better Iowa scores. I don't think ASAP skills translate into Iowa scores, which is what people look at. If they published ASAP scores in the newspaper, then I think you would see the districts emphasize it more because if they didn't then the parents would start calling and saying, "Rrrr, we're the lowest in ASAP scores, rrrr!" like they do now about the Iowas. (Extension Study, Pine)

In high-stakes districts particularly, the accountability function worked against ASAP's reform agenda and encouraged teachers to teach to the ASAP just as they taught to the standardized tests.

Administrators read a study that shows that the more you give them practice tests, the better they'll do on ASAP. So they think to themselves, "What can we do to get those scores up?" So that make us give more and more tests. And they are taking much too much time to give and to score, that we don't have time left over to teach thematically. (Extension Study)

Teachers inflate the scores because they are afraid of looking bad in the media. (SOC)

ASAP is not authentic. It is a performance-based test used to mirror holistic instruction. However, tests do not change practice without staff development, resources, and support for teachers. This is used primarily for teacher accountability. I don't see the difference this test makes to instruction. (SOC)

We have so many different types of tests that are being given at the end of the year, especially, and after awhile I don't know what student reaction to that is. I know, as a teacher, you can kind of sense the tedium of, "Oh, here's another test." (Extension Study)

To serve the accountability function, a test must yield valid inferences about the quality of education students have received. Teachers doubted that the ASAP was up to the task. The survey asked teachers the extent to which they agreed that ASAP accurately reflects the overall quality of education students have received. Only 14% agreed or strongly agreed with that statement. Surprisingly (given the wholesale rejection by teachers of the previous state test mandate), teachers doubted ASAP's validity even more than the validity of the standardized test. In the earlier study of teachers' reactions to the previous state testing mandate of ITBS and TAP, 39% of the sample agreed that those tests reflect the overall learning of a pupil throughout their schooling (Nolen et al., 1989).

Aside from their reservations about the validity of ASAP, many of the teachers believed that ASAP functioned more as an accountability device than as a way to assess learning authentically. Forty-eight percent reported that their districts use standardized tests to compare schools or evaluate teachers, and 34% reported that their districts use ASAP for these same functions. Sixty-two percent believed that teachers feel pressure to raise ASAP scores. Thirty percent reported that when teachers prepare pupils to take ASAP, they focus that preparation on those aspects of performance that will yield high scores, a practice tantamount to teaching to the test.

According to the survey of district administrators, in almost every district the ASAP was used for the same purposes as were standardized tests. That is, if traditional tests served high-stakes accountability functions, so did ASAP. Almost 44% of the district administrators judged that the emphasis placed on test scores was moderate (33% judged it to be high). When asked about how ASAP and other tests were used in their districts, only 15% reported that ASAP was used *neither* for accountability to the public, comparison of schools, nor for evaluation of teachers. Thirty-seven percent said that ASAP added to the existing test burden (the time devoted to preparing for and administering mandated tests). Only 5% reported that less time was spent on testing now. Asked about the validity they felt ASAP has for their students, 13% believed it had poor validity, 41% believed it has moderate validity, and 30% judged its validity as high.

As an indication of test burden, teachers estimated the number of hours devoted to test preparation and administration. Including the ASAP, ITBS, and state-mandated district assessments, the median number of hours spent in administering the mandated tests was about 18 hours, with 25% reporting 25 or more hours. The estimates for time spent in preparing pupils to take these mandated assessments had a median of about 28 hours, with 26% reporting over 50 hours. Seventy-two percent of the district administrators reported that more time was spent in testing as a result of the ASAP mandate than had been spent before.

Test burden was notably higher in urban, high minority, and resource disadvantaged districts. The mean for time spent in test administration of high minority districts exceeded that of low minority districts by one-third standard deviation. Urban districts had higher test burden than either rural or suburban districts (effect size = .22). The most financially advantaged districts had the lowest test burden (effect size = .23 compared to all others and effect size = .31 compared to the least financially advantaged districts). The effects on test preparation followed the same pattern.

ASSERTION 8:   The Lack of Attention at the State Level to Concerns for Equity and Fairness Inhibited Coherent Local Response to the Policy

Critics have long argued that standardized test scores are insensitive to cultural, socioeconomic, ethnic, and language differences. Poor children, children of color, and those with first languages other than English suffer when government agencies attach consequences such as track assignment and graduation to these biased scores.

Most Arizona educators we encountered in this study shared these criticisms of mandated standardized tests and hoped that performance assessment and holistic, student-centered curriculum might alleviate these sources of bias. As one teacher noted:

> Experiential learning is the only way to go with our kids. They have to see a connection with something they know. You have to give them a chance to speak and write in the language they feel comfortable in. They get those bubble-ins, and it means nothing to them. And they're not dumb, far from it. But what they know doesn't show up on the ITBS. (Extension Study)

On the national scene, as experience accumulated with alternative assessments, some experts raised doubts about whether the performance tests would be any less biased than traditional tests. For example, Darling-Hammond (1994), Shepard (1993), and Winfield and Woodard (1994) suggested that the disparity between rich and poor, minority and nonminority students would be as large as, or larger than, that produced by standardized tests. Among the reasons they offered for this prediction are that minority and economically disadvantaged pupils are most likely (a) to have experienced predominantly skills-based instruction and heavy preparation for standardized tests; and (b) to have attended schools with less capacity to develop in relation to the reform. As a result, poor children and children of color would lack opportunities to learn content and practice the tasks peculiar to performance assessments. Research by Madaus and his colleagues (Madaus et al., 1992) confirmed that test burden and time spent on intensive test preparation was more evident in schools with high concentrations of minority pupils.

Critics also pointed out that children with first languages other than English would also likely demonstrate lower scores than other children since they face more complicated tasks than English-speaking pupils. The task requirements embedded in performance tests are complex; for example, one must read or listen to the task instruction, read a passage of text that includes a math problem, comprehend both the text and the problem, solve the problem, compute solution, and then write about one's reasoning. For limited English-speaking children, the complexity is increased: that is, reading in translation, comprehending and solving the problem, then writing in translation. With this unseen difference in complexity, the construct being measured in the performance test actually differs from child to child, and the resulting score also has different meaning.

The Arizona Department of Education broke ground in the assessment field by including a Spanish language version of ASAP Form D for its largest group of limited English speakers. Unfortunately, its first draft used dialect inappropriate for the southwest. Even when this problem was corrected, there were inadequate pilot trials conducted to work out the inevitable problems of such new testing technology (B. Arias, personal communication, 1996)

Although there was remarkably little discussion and debate over these issues (for example, the state never appointed an equity advisory committee, and no community advocacy group came forward), issues of equity and bias did enter the

experience of many of the teachers we studied. The second-language issue seemed to evoke the greatest open-ended survey response.

> I am a strong advocate of the ASAP testing measure. I feel that it is a comprehensive test that really assess student achievement. The only problem is that the translated version (Spanish) was not well done. The English version was much easier to understand and administer. (SOC)

> I teach Mexican children, all of whom speak English as a second language. Many of the topics and written language are new to them or unknown by them. Also because the vocabulary is so limited, much of their writing is below grade level. (SOC)

> I find that Limited English speakers should be exempt. We need to provide them tests in their home language. They are totally frustrated and confused. They are very bright students but are being penalized for not speaking, writing, reading in English. (SOC)

> Cultural bias in the test is a reality, and bilingual students cannot find a way into the maze. (SOC)

On the survey, 65% of the teachers believed that many children who are acquiring English as a second language will score poorly on ASAP even though their educational achievement is adequate. On a related matter, 70% of the teachers believed that students with poor reading would suffer on the math test even when their math skills were adequate. The reading load on the math and writing tests would certainly affect limited English speakers more than most.

> If you can't read that well, you're pretty much out of luck [on both writing and math]. That's the bad part. (SOC)

Ethnic and socioeconomic bias of ASAP also came into the teachers' awareness. On the survey, 60% of the teachers agreed with the statement that the differences between minority and nonminority pupils on traditional achievement tests would also be produced on the ASAP. And 67% believed that ASAP scores reflect socioeconomic differences.

One of the arguments in favor of performance tests is that they activate pupils' prior knowledge and allow them to construct their own meaning. But some

teachers pointed out that this assumption is not plausible for poor children, as the following excerpts show:

> This is such a good program for the school districts and the state. However, thought needs to be given to the diverse cultural representation and socioeconomic levels in the state. Even after reading the article about hot air balloons in the eighth-grade test, many students had absolutely no concept of a hot air balloon which one must admit is an upper-middle-class toy. (SOC)

> Any ASAP topic is going to be artificial to a certain extent for some people. Everybody is not going to be equally excited about any one thing. But if you think about last year's ASAP [D-1] topic, about tarantulas, that was much more fair than this year's [D-2] one about the saving the burros in Grand Canyon. The spider topic was much easier for our kids to identify with in terms of being meaningful and related to their experiences. So here you have people from somewhere else selecting topics that they think the kids in Arizona will really identify with, because, when we think of Arizona, don't we all think of the Grand Canyon. But our kids don't think of the Grand Canyon. And in a class at this school, if you have one kid who has been to the Grand Canyon, you were very lucky. Whereas, my daughter's class [in a suburban, middle-class school], every kid but her had been to the Grand Canyon. So the whole notion of being on an equal footing as far as background experience goes, is off the mark. The kids here, they weren't interested in writing letters to save the burros or mules or whatever they were. They would have been more interested in helping the people in their own neighborhood. They'd write all day about that! (Extension Study)

In addition to asking teachers directly for their views, we used survey responses to look at the difference of response of teachers who teach in schools with high versus low rates of disadvantaged and minority students to key questionnaire items. For example, consistent with the findings of Madaus, we found that test burden—the number of hours spent in test preparation and administration—was substantially higher in schools with a high proportion of minority pupils (effect size = .33 for administration of tests and .23 for preparation for tests). Teachers in high-minority schools reported receiving more hours of training in ASAP itself (effect size = .37) but no advantage in amount of training in teaching and curriculum related to ASAP.

Schools with different pupil composition did not differ in the belief that they were pressured to raise test scores, so that testing pressure bore no relationship to the percentage of minority pupils in the school. However, the rate of agreement with the statement "Teachers feel pressure to raise ASAP scores" was quite high

(64%) overall. Teachers in high-minority districts were more likely than others to believe that the benefits of the testing program were worth the time, effort, and money required, although only 28% overall agreed with that statement.

The relationship of capacity building and equity produced a mixed picture. Teachers in high-minority schools were more likely than others to report that they were learning how to teach with thematic units and teach complex, real-world problem solving, although the rates were low in the sample overall. Nevertheless, teachers in high-minority schools reported that their students have less opportunity to learn in such areas as lifelike learning problems, long-term projects, higher order thinking skills, combining reading and writing, and writing about science topics. A difference of nearly one-quarter standard deviation separated teachers in schools with high versus low concentrations of poor and minority pupils.

CHAPTER 4
DISCUSSION

The Arizona Student Assessment Program is a part of the national movement to improve schools by means of various policy instruments, such as achievement standards, assessment systems designed to measure the attainment of these standards, and accountability mechanisms to reward and punish schools for (and inform the public about) their level of accomplishment.

Although the ASAP no longer exists as official state policy, its history may be seen as a microcosm for the movement as a whole. The consequences of ASAP offer insight into what might happen in similar sets of circumstances elsewhere. At the least, the evidence on the consequences of ASAP can shed light on the assumptions on which standards- and assessment-driven reform are based. As Cohen (1995, p. 11) noted, the assumption that these policy instruments will change instruction "remains a conjecture, for there is little evidence of direct and powerful relations between policy and practice."

The research we report here is part of a very early wave of empirical probes on the question about what happens to schools in the wake of reform by standards and assessments. This study was extensive, spanning the entire life of ASAP. Both policy and practice were examined. We looked intensively at what was happening in purposefully selected classrooms consequent to ASAP, inquired deeply into educators' beliefs and practices in context, and also tapped broadly into the responses of educators across the state. In striving for both generality and particularity of inference, we accomplished something uncommon in policy studies. In this chapter we organize the results into five discussion points and discuss them in light of the relevant theoretical and empirical literature.

Awareness of ASAP Was High But Acceptance Was Far From Universal

Cohen (1995, p. 13) argued that plausible criteria for success of a reform such as ASAP include "teachers' awareness of new policy directions," "attitudes toward the reform," and incorporation of the language of the reform into teachers' discourse. On these criteria, his research showed that similar reforms in California

were successful. He also noted that "most teachers have had great latitude to assign different meanings to new policies and to respond idiosyncratically" (p. 14).

To change what schools do, to make curriculum more challenging and coherent, to increase achievement, educators must at least be aware of the reform. Not every federal or state policy meets that test. The evidence from our study clearly shows that Arizona educators knew about ASAP, though what they knew about it varied dramatically among individuals and settings. Most thought of a particular test when they thought of "ASAP." Because they defined "ASAP" as a performance assessment, they associated it with the means to correct the injustices, invalidity, outmoded pedagogy, and dysfunctional accountability effects that they had come to associate with traditional, standardized tests. Other educators, when they thought about "ASAP," defined it as a state program to reform pedagogy toward constructivist practices and principles. Still others defined "ASAP" as an accountability system, as a way of potentially changing the relationships among the state, the districts, teachers, and students. This latter definition was a minority view, since few had experienced the accountability system directly, and the full extent of high-stakes uses of ASAP (e.g., as a graduation test) had not yet come into play.

How educators defined ASAP influenced whether or not they approved of it and probably their subsequent actions as well. Many teachers disagreed with the philosophy they thought that ASAP represented. Those who were steeped in a traditional, basic skills orientation to teaching and defined ASAP as a reform toward constructivism disapproved of the whole program. Reformers in Arizona very likely underestimated the depth of commitment of many districts and individual teachers to traditional ways of teaching. Reformers nationwide construe the divide between behaviorism and conservatism as merely a rational choice point. These data indicate instead that the divide represents a full-blown ideological conflict that likely stands in the way of reformers' ambitions.

Nevertheless, there was strong support for ASAP in many quarters. In addition, many educators believed that ASAP was a good idea that failed in execution, particularly with respect to the testing system itself, seen in confusing directions and procedures, changing messages from the state department, unclear expectations and the like. The majority of teachers felt that the benefits of ASAP were not worth the costs. Acceptance of ASAP testing was no greater than acceptance of the previous state mandate of standardized tests.

Do attitudes, awareness, and definitions translate into action? Cohen (1995) questioned whether the positive attitudes of California teachers and even the incorporation of reform language into their discourse actually seeped into actual teaching practices across the board. Elmore (1995) likewise found little change in teaching practice to match teachers' enthusiasm for reform ideas or even the structural changes that have been implemented in their schools to adapt to reform ideals.

ASAP Had Modest Effects Overall, Significant Effects in Some Places, With the Degree and Kind of Effect Influenced by Local Conditions

Did ASAP change instructional practice? This question seems simple, but could hardly be more complex. Those who believe that public education is a system assume that policies issued from the top work through the levels of organization in relatively straightforward and rational ways. Given enough time and alignment of companion policies, teachers will bring their practice into conformity (e.g., see Smith & O'Day, 1991). Other theorists (e.g., Hall, 1995) observe that policies issued from the government are redefined at each system level, so that the original intentions of a reform such as ASAP become distorted or completely ignored at the level of practice.

Whether one is an interactionist such as Hall or instead believes in the power of top-down, systemic reform, one must still question whether any real effects could logically be expected, given the short history of ASAP. Similar reforms in California, Vermont, and Kentucky have had more time to produce change (and state reformers more patience about the time required to bring change about). Perhaps the small effects that we observed in these data should be reason for optimism about the potential for educational change through mandated standards and assessment. The survey information that over half the teachers felt that ASAP had had little or no effect on their teaching is sobering nonetheless.

The remarkable changes made in some places such as Desert and Hamilton Elementary Schools show what can happen when the local circumstances are fortuitous. At Desert, teachers in the benchmark, third grade, started with beliefs and values predisposed to be favorable to ASAP but with little knowledge about what to do to change their practice in accord with it. With plenty of resources, a supportive principal, and a record of high test scores that gave them some room to maneuver, they built their capacity through an intensive professional

development effort. They brought in consultants, acquired curriculum and training materials, experimented and reflected together. By the end, they had reoriented their practice toward engaging, ambitious, and integrated subject matter, writing across the curriculum, reading authentic texts, and problem solving in math. The central feature of their classes was the use of product standards, with teachers and students using rubrics to interpret the quality of their work. According to Parish (1996), who observed as their practice evolved, these changes would not have come about without the state reform.

At Hamilton, teachers who had been predisposed to constructivist education, but silenced in a climate of pervasive behaviorism, used the ASAP mandate opportunistically to advance their own agenda. They conspired together to introduce practices that would not have been permissible otherwise. They took advantage of their district's intense interest in showing up well in the public display of test scores to justify introducing ASAP-related pedagogy, at least in the benchmark grade. This form of action in response to ASAP can best be understood as tactical, in comparison to Desert's wholehearted embrace.

Teachers at Hilldale and Jackson, already thoroughgoing constructivists and improving every year, initially approved of ASAP's reform aims and used ASAP opportunistically to advance their interests in the unfriendly micropolitics of their districts. Yet later they soured on ASAP over the burden ASAP testing put on instructional time. They came to know ASAP as just another test imposed on them for political reasons, motivated by absence of trust in professional educators. The changes toward ASAP aims would have been made even without the state mandate. Indeed, ASAP seemed to them to stand in the way of their progress.

Teachers at Valor never had a chance. Isolated geographically and without funds to purchase new materials or hire consultants and professional developers, they were the blind leading the blind, happy with their released time but without any means of combating their lack of knowledge.

The dramatic variation of effects of ASAP shows how central reforms interact with local conditions. When one understands the deeply rooted ideology that dominates places like Franklin, one can appreciate the power of district and school filters to either enhance, retard, or distort the intents of state policy. Franklin's practices were dominated by images of their pupils as too poor and too

deficient in ability and English language and social background ever to profit from constructivist pedagogy. For the teachers to have reacted in an effectual way to ASAP would have been remarkable.

Nor is Arizona unique in displaying these local variations to state reform (e.g., see Aschbacher, 1993; Firestone, 1989). Cohen (1995) described his research on systemic reform in Michigan this way:

> A few Michigan localities moved aggressively in the direction of reform, largely because local education leaders saw the state's proposals as a way to advance their own long-held ideas about instruction. Several other districts ignored the reforms, but many others moved cautiously and irregularly in new directions.
>
> Change was also fragmentary within districts, partly because many central offices sent missed signals. Even in districts with quite progressive central leaders who pressed for reform, other central office staffers argued for or persisted in a more traditional approach. Central office administrators often interpreted new state policies in different ways within districts because central subunits have quite different missions and make use of higher level policies accordingly. . . . The varied local responses to systemic reform also owe a good deal to differences among schools' responses to district-level guidance for instruction. Principals' response varied significantly within districts: some embraced the reforms and used them as an opportunity to try to change instruction while others maintained their attachment to traditional classroom methods, and still others adopted a neutral stance.
>
> As a result of these developments, variability in guidance for instruction has increased in consequence of efforts to reduce it. . . . One reason for increased variability in state and local guidance for instruction is that the growth of state instructional policy has not constrained local instructional policy-making. State guidance added messages, but so did local agencies. Nothing was subtracted. . . . local educational authorities acted as though they had undiminished authority to make instructional policy. (p. 12)

Lack of Attention to Capacity Development Undermined the Reform Agenda of ASAP

The essence of this reform movement is improving what students know and what they can do. But students cannot know what they have not been taught. And teachers cannot teach what they themselves do not know. It seems to follow logically that a state reform such as ASAP ought to include some provision for enhancing what teachers know. What actually happened consequent to ASAP,

however, was neither simple nor logical, and there is an important lesson in the Arizona experience for other states.

One of the most controversial aspects of the movement to reform schools by instituting standards and assessments is the weight to give outcomes (i.e., content and performance standards, test scores) in relation to inputs (i.e., delivery standards, opportunity to learn, opportunity to teach, capacity). Corcoran and Goertz (1995) attempted to clarify several alternative definitions of capacity and capacity development. For systemic (top-down) reformers, capacity development means the adoption of standards and assessments consistent with them, and the alignment of other policies that might otherwise conflict with them. But for school-based reformers (those who believe that schools can only be reformed from the bottom up), capacity development means the opportunity for teachers to develop and share their craft knowledge and the removal of governance barriers to teacher autonomy and growth. For some writers, capacity is the ability to do something. For others, capacity is the maximum or optimum amount of production. They expand on the latter definition and caution about its assumptions (Corcoran & Goertz, 1995):

> This definition focuses attention on the results of school reforms rather than on the means of implementing them, and it raises the issue of efficiency in the sense that capacity is viewed as the optimal amount of production that can be obtained from a given set of resources and organizational arrangements. . . . [which] raises the question, "what does the system produce?" The conventional wisdom is that the output of the system should be measured in terms of student achievement. Focus on results is one of the mantras of reform. But achievement is coproduced. It is simply not within the power of schools to ensure high performance by all students unless one assumes schools full of happy, hard-working youngsters with high aspirations or schools that function like total institutions, able to control the socialization of their charges and compel the necessary study. But we do not have students or schools like this. . . . We suggest that the defining "product" of the education system is high-quality instruction, which is central to the ability of the system to help all students reach high standards. . . . What are the key components of the instructional capacity of a school? . . . the intellectual ability, knowledge and skills of teachers and other staff; the quality and quantity of the resources available for teaching, including staffing levels, instructional time, and class sizes; and the social organization of instruction or instructional culture. (p. 27).

Concern for capacity in standards- and assessment-based reform has stumbled on issues of cost (professional development is vastly more expensive than assessment programs) and political expediency (it has proved easier to build a coalition to pass testing programs than staff development programs because of the issue of preserving local control of school operations). According to McDonnell (1994), however, it is necessary, both for reasons of efficacy and to couple standards and assessments with programs, to ensure that schools and teachers have the means to offer the requisite kinds of programs so that standards can be met. She explained her reasoning this way:

> Because many new forms of assessment require that teachers play a key role in their design, administration, scoring, and use, these assessments will not work as intended unless adequate training is provided. The need for major new investments in professional development is even greater for those assessment policies that are expected to change curriculum and instructional practices. . . . The second reason in favor of linking capacity building and assessment is a normative one that stems from the notion that political accountability is a reciprocal relationship. The public and their elected representatives are holding schools accountable through various forms of student assessment, but in return they are obligated to provide schools with the resources required to meet accountability standards. . . . [S]tudent assessment policies should be linked to capacity-building strategies that fulfill the public's side of the social contract between political communities and their schools. (McDonnell, 1994, p. 414)

Compared to the substantial efforts of states such as California (Goertz, Floden, & O'Day, 1995), Arizona's effort to build capacity to implement ASAP reform ideals was virtually nonexistent. ADE expected that the presence of Forms A, B, and C would make teachers aware of the kinds of performance the state desired and would focus their attention on the Arizona Essential Skills, and thus they would adjust their instruction accordingly. ADE offered an extensive series of meetings and workshops, but their emphasis was on the principles of performance assessment and the scoring rubrics. Beyond that, the state provided nothing for capacity development. Teachers were on their own, although many districts accepted the responsibility to align their formal curriculum to ASAP, and to acquire suitable texts and materials, so that they would be consistent with problem-solving math, higher order thinking skills, integrated subject matter, and writing across the curriculum. These districts sought out experts for consultation and mounted significant professional development programs.

But most did not. As the survey found, only 19% of teachers felt that adequate professional development had been provided so that they could adapt to ASAP. Eight hours over two years was the median amount of training they received in how to teach that which ASAP assessed.

The failure of the state to attend to capacity development is a clear explanation for the marginal effects of the reform. The effects that they envisioned are not likely to be realized short of the investment of substantial human and material capital. Research such as that of Borko, Flory, and Cumbo (1994) and Flexer and Gerstner (1994) shows how difficult is the task for teachers to make the change from traditional basic-skills orientation to an orientation toward constructivist pedagogy and assessment.

Failure to Solve Testing and Accountability Problems Undermined the Reform Ideals of ASAP

A remarkable and inescapable conclusion from the data of this study is the critical and perhaps fatal role played by measurement weaknesses in the ASAP program. First, most teachers defined ASAP as if it were simply the tests they gave. So, when they encountered a problem with the test, such as confusing directions or items too difficult for the students in their particular classes, they condemned and rejected the program as a whole. When they realized the state generic rubric was flawed, they defined "ASAP" as too subjective. When they were unable to figure out what the Form D scores meant, or when they experienced the vast amount of time and effort that preparing for and administering the tests was taking, even teachers who were predisposed toward ASAP reform ideals were disillusioned.

Later, when the state superintendent saw that Form D reliabilities were too low, and that Form D correlated too poorly with the other ASAP measures, she first suspended its scheduled administration and subsequently changed the nature of the program altogether (see Smith, Heinecke, & Noble, in preparation, for more complete details).

Although the testing problems embedded in ASAP were serious, we argue that nearly all of the problems of the ASAP performance tests could have been corrected or ameliorated, given sufficient time and psychometric expertise. 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 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 (Lorrie A. Shepard, personal communication, 1996). Given the primitive status of the technology of performance assessment, more time was needed. Faced with similar psychometric problems, Kentucky employed technical means to improve the reliabilities of its performance assessment, Vermont provided more development time and postponed attaching high-stakes consequences to the results of its portfolio assessments, and California appointed an professional review panel (Cronbach, Bradburn, & Horvitz, 1994; Koretz, Stecher, Klein, & McCaffrey, 1994). But Arizona simply canceled the tests.

It is doubtful whether the reliability coefficients of any performance assessment can match those of standardized tests. But what is high enough? Cronbach, et al. (1994) recommended that levels of reliability be commensurate with the stakes to be attached to the results. It should be a matter of broad-based dialogue among all stakeholding groups about how much emphasis there should be on reliability coefficients versus, for example, construct or consequential validity (Messick, 1994; Moss, 1994).

Although there cannot be any definitive answer to the measurement problems of ASAP, what was clearly needed was independent evaluation of the performance tests and of the program as a whole. We are mindful of the plaintive comment of one survey respondent about the fallacies in the testing: "Please give us a chance before the ASAP program suffers any more damage."

The form of test is one thing, but the function assigned to ASAP testing was something else. When ASAP was originally introduced, it was advertised as a testing program for which teachers would have substantial responsibility in designing, administering, scoring, and evaluating afterward. Teachers believed that ASAP would not be used for high-stakes purposes and were later disillusioned. Because of financial and time restraints the state relied more and more on test

publishers to develop, field test, score, and evaluate the results, and thus excluded professionals from the process.

Furthermore, the effects of ASAP as an accountability system were phased in gradually, and few teachers were aware of them all. A variety of research has shown that high-stakes accountability can have unintended and even pernicious effects on schooling, narrowing curriculum and focusing pedagogy only on tested tasks and content (Smith & Rottenberg, 1991). ASAP did not last long enough to probe this contention, although the majority of teachers in the survey reported that they felt pressured to raise ASAP scores in much the same way that they were pressured to produce high standardized test scores.

Cohen (1995, p. 13) noted the apparent anomaly in the movement's reform and accountability intentions. Motivated by perceptions that public schools are failing, "advocates of systemic reform propose to radically change instruction, and for that they must rely on teachers and administrators. But these agents of change are the very professionals whose work reformers find so inadequate."

In a similar vein, McDonnell (1994, pp. 408-409) distinguished between political and professional accountability in a way that echoes the ASAP dilemma of intentions:

[P]olicy-makers draw on a political model of accountability. This model assumes that the larger community and its elected representatives have a right not only to hold public institutions answerable, but also to circumscribe and control their behavior. Political accountability posits that a larger public interest transcends the interests and values of any single class or person, and that its pursuit is best ensured if individual behavior is held accountable to the larger body politic. . . . [W]hile the political model derives its legitimacy from its roots in the democratic electoral process, the professional model bases its claims on experts' mastery of a specialized body of knowledge . . . [and] assumes that, because their work poses complex and nonroutine problems, their application of that knowledge should be regulated by a code of ethics internal to the profession. . . . The application of professional knowledge to individual clients' needs requires judgment, so it cannot be reduced to rules or prescriptions for practice; thus professionals require autonomy from external political control in determining how the products of their expertise should be used.

The dilemma of intentions may not lend itself to resolution. Nevertheless, a public discussion of the standards-driven reform movement ought to consider

whether there is something about accountability that may work against the program that it appears to support.

How Can There Be Accountability and Reform Without Equity?

The movement to reform schools intends to raise the achievement of all students—rich and poor, minority and nonminority alike—by imposing standards common to all. Yet critics point out that what seems to be common and fair at the policy level falls out at the local level into quite different experiences for students of different groups, and different outcomes as well. We do well to heed the warnings of Darling-Hammond (1994), Madaus et al. (1992), and Winfield and Woodard (1994), that deficient opportunities to learn and dysfunctional accountability burdens experienced by poor students, students of color, and students learning English would contradict the positive intentions of the reform.

The evidence from this study shows that teachers were aware of the differential validity of ASAP for poor and minority pupils. If one assumes that tests can provide the basis for accountability, and if the tests are invalid for even a part of the student population, then the accountability intention of ASAP is thwarted.

The evidence of this study also shows that, according to the teachers' reports, the opportunities of these children to learn, for example, higher order thinking skills, are considerably less than those of other children. If so, then the intention of ASAP to change and improve education for all students is likewise thwarted.

Other evidence from this study raises serious issues of equity as well. Repeatedly, there were instances of teachers who questioned whether disadvantaged and minority pupils were capable of learning sophisticated curriculum, or anything more complex than basic skills. This suggests that both teacher knowledge and values must change for the ideals of the reform to be realized.

Differential validity and unequal opportunities to learn, demonstrated in this study, strongly suggest that, had decisions such as grade promotion, tracking, and high school graduation been tied to ASAP results, poor children, children of color, and children whose first language is not English would have been harmed by this reform. These possibilities point to a great need for public, professional, official, and interest groups' scrutiny of ASAP and its consequences, but little was evident.

APPENDIX A

CROSS-SITE DATA MATRIX

CROSS-SITE DATA MATRIX

	VALOR	FRANKLIN	PINE	HILLDALE
DISTRICT CHARAC-TERISTICS	Unified K-12 [K-5, 6-8, 9-12]. 1000 pupils. Dropout rate 12%. Transience rate fluctuates according to harvest seasons. Rural site, migrant influence. District and school are ethnically diverse.	Elementary, K-8 [K-2, 3-8]. 800-900 pupils. Dropout rate from high school district is 50%. Transience rate 90%. Urban core. District and school are ethnically diverse.	Elementary, K-8. 10K pupils in district. 14 schools are K-5 and 4 are 6-8. Dropout rate from high school district is 16%. Transience rate of 7%. Suburban. District is mostly middle class and Caucasian (17% ethnic minority).	Unified, K-12. Dropout rate is 10%. 24,000 pupils in district. Less than 10% minority pupils. Suburban.
DISTRICT RESOURCE BASE	One of the most resource-poor districts in AZ agricultural, commercial economy. Almost no discretionary funds are available for curriculum or professional development.	Nearby factories provide tax-base adequate for financing school operations. Moderate resource base.	Residential, commercial, light industry provide adequate tax base for financing school operations. Moderate to good resource base. District invests substantial amounts in professional development.	Residential, commercial. Adequate tax base for financing school construction and programs.

	VALOR	**FRANKLIN**	**PINE**	**HILLDALE**
DISTRICT TESTING MODEL	ITBS in Grades 1-8, due to Chapter 1 requirements. Purchased off-the-shelf criterion-referenced test package, similar to ITBS format, given twice/year. CRTs are objectives-referenced with 80% mastery level set. Administrator: "85% of students' composite scores (ITBS) fall below the 50th percentile." District pays for copying costs and costs of hiring teachers to score the ASAP Form A for Grades 9-11 to conform to state mandates for district testing. This is considered a significant expense. District feels state sends conflicting messages: that district tests too much AND that they should expand performance testing. "But, hell, give us a break."	ITBS in Grades 2-8, due to Chapter 1 requirements. District CRTs written by teachers, given 3/year with 70% passing rate specified. Teachers write CRT items, which resemble ITBS format plus writing test. District test results used for teacher evaluation and curriculum revision. Always one of the lowest scoring districts on ITBS. Administrator: "Two or three standard deviations from the state norm." Although district claims to "totally de-emphasize ITBS," it publishes scores in newsletter to parents.	ITBS in mandated grades plus Grades 5 & 6. District CRTs developed by district testing bureau, administered 2/year in math and communication arts. Format is standardized, objectives-referenced, multiple-choice with mastery levels specified; timed math tests. Frequent formative tests are expected, with results recorded by teachers and monitored by district. District test results used for teacher and principal evaluation and school comparisons. District claims disinterest in ITBS results, but publishes them in information packet, showing district is at or well above national norms in all tested areas. CRT results used in promotion/retention decisions.	ITBS in mandated grades. District testing went through revision post-ASAP from CRT to ASAP Preliminary Forms A, B, and C (performance assessment). No testing bureaucracy exists at district. Little commitment to a standardized testing model.

	VALOR	FRANKLIN	PINE	HILLDALE
TEST DEMAND	Although "there's a big emphasis on test scores," district does not expect high scores, based on its history and pupil composition. Although gain scores are perceived to be accurate indicators of how teachers are doing, administrators try not to focus teachers' attention on low scores or pressure them to increase scores. Teachers don't feel too much pressure, regardless of which test is used.	District mission statement's first goal is to raise pupil achievement as measured by CRT. Teachers are compared on CRT scores, are told to "fix" whatever skills their students are low on. Teacher: "Administrators will do anything to make those scores look better," even changing items in midyear. Total de-emphasis on ITBS score. High stakes on CRT only.	The high-stakes testing environment is an extension of its philosophy of learning, which emphasizes measured mastery of skills. District statement of philosophy: "Schools can maximize the learning conditions for all students through clearly stated outcomes, high expectations for all students, and continuous assessment of student learning." There is a strong perception that the schools exist on a distribution of their test scores. Principals of low-scoring schools are publicly admonished and subject to transfer or removal. District test coordinator to principal: "Pete, is there any life at Pines?" Teachers, in turn, feel the pressure exerted by principals. Anxiety is translated into rigid compliance to district testing policy. CRT results are used in part in the district career ladder program. Use of CRT results for promotion decisions indicates high stakes for pupils. Media attention to test scores is strong.	District administrators expect scores to be high because of socioeconomic status of residents and because standardized test scores have been historically high. District projects a positive image in part because of its high-ranking scores. Principals of low-scoring schools recognize this competitiveness, yet pressures are not particularly high to improve scores at any cost. District administrators are more interested in authentic improvement rather than in high scores for their own sake. Reaction to ASAP scores that were not as relatively high as ITBS/TAP scores, however, prompted district official to say to principals: "We should not be below the state or county average. I intend to be involved with you so the 1994 results show a different response." Principal in the school with highest ranking ITBS scores was concerned that ASAP school scores were not at the top, but did not encounter pressure from district officials to raise them. Strong media attention to scores of all kinds.

110

	VALOR	FRANKLIN	PINE	HILLDALE
DISTRICT ORGANI-ZATIONAL CULTURE	District is very loosely organized by tradition and rural character. Principal has relatively wide range of autonomy from district administrators, and in turn imposes few mechanisms that would withhold authority from teachers. No bureaucracy exists for either curriculum or assessment.	Mission statement: "Increased emphasis will be placed on instructional and management strategies which enhance improved academic performance." Top-down authority structure, with CRTs used as standardizing, centralizing mechanism.	District document: District is actively engaged in efforts to improve teaching, learning, and decision making through systematic efforts (Total Quality Management). Administrator: "This is a policy-driven district." While espousing site-based decision making, has developed elaborate procedures for monitoring the performance of schools, principals, and teachers and pushing a common, centrally-controlled set of goals and operations. District continues to experience rapid growth and frequent changes in administrative personnel. District testing plays a dominant role in the top-down nature of the district, with principals defined as responsible for producing good test scores and schools judged as good or bad based on scores. Administrator: District is driven by pattern of promoting and adopting whatever is perceived to be cutting-edge educational movement, which keeps teachers always in the role of novice, without time or resources to become experts.	There is a traditional value placed on school authority balancing central district authority. District is experimenting with site-based management. Schools differ from each other on the reading programs they offer, their graded organizational structure, etc. District officials lead by encouragement and capacity building, rather than by strict policy mandates or intimidation. Official: "It's up to the district to provide for staff development so that teachers can respond to innovations such as ASAP. We have the resources and intend to make them available to teachers."

	VALOR	FRANKLIN	PINE	HILLDALE
ASAP EXPERTISE/ COMMIT-MENT OF DISTRICT	District officials accept performance assessment and holistic instruction as an ideal, but lack in-depth understanding and expertise to pass on to teachers. Misunderstandings are evidenced, for example, by claiming that whole language is an emphasis in the district, but the adopted reading program follows a direct instruction model. Officials fail to recognize contradictions such as these.	Commitment of administrators to basic skills mastery precludes commitment to ASAP-like teaching. Administrator: "Local determination of ... what needs to be fixed with a kid academically always supersedes what the legislature says." No ASAP gatekeeper exists. Prior failure of holistic instruction experiment is additional barrier.	A significant and vocal minority of teachers profess commitment to ASAP principles of teaching and assessment. Most district administrators, however, are more interested in contrary principles, such as those embodied in TQM. Key administrator thinks that performance assessment is "the performance of objectives." No district official was expert in ASAP or advocated for it.	Expertise in ASAP and ASAP-consistent education was scattered across district, not centrally situated. ASAP coordinator was part-time only and was replaced after the year. Teacher (speaking about coordinator): "She's just a gifted ed teacher, so what good is she?"
BELIEF OF OFFICIALS IN ASAP PERMA-NENCE	Officials believe that ASAP has merit as a form of assessment, but are concerned about its costs, feasibility, and its possible use by the state as an accountability mechanism. They accept it as a permanent part of the state's accountability package and generally accommodate.	Administrators perceive ASAP to be temporary and of little value in shaping education for "these pupils." "It's going to collapse logistically after 2-3 years." Thought by some to be a "pilot."	Administrators believe that ASAP has problems but will probably be a permanent part of state test mandate, and will work out the kinks. But they see ASAP as much more a mastery model than a constructivist model.	District official, referring to performance assessment: "This is the wave of the future. This is reform." Official speaking to principals: "For those who believe this type of assessment is going away—it ain't going to happen." Teachers fear that ASAP will be like many movements in AZ education, here and gone, in relation to the career interests of state officials.

	VALOR	FRANKLIN	PINE	HILLDALE
REACTION TO ASAP RESULTS	District was above average on ASAP results (attributed by district officials to a fit between "what we teach and AZ Essential Skills"). This contrasts with their much lower average performance on ITBS. District test coordinator presented results at public school board hearing, expressing pride that Valor was second highest in the county. Results were not translated into any action, however.	District ASAP scores were only 1 standard deviation below state average (not as relatively low as ITBS scores). Little notice was taken. Some acknowledgment that state emphasis on ASAP might require some district adjustments. Third-grade teachers' rewrite of test items to reflect ASAP principles was reversed in favor of more basic skills.	ASAP results were not given to schools in any systematic or purposeful way. Most teachers did not even know about them. Because the school scored higher in the district ranking of ASAP than it typically scores on ITBS or CRT, the principal felt some pride and received a few congratulations. However, because ASAP was given no importance by the district, this accomplishment was greatly overshadowed by the school's performance on the more valued district tests.	Official: "We should not be below the state or county average. I intend to be involved with you so the 1994 results show a different response. It is time to get on board. Make sure those in your building understand that alternative assessment is an important tool for teaching and assessment. If your students have not been exposed to higher order thinking then they will not be prepared to do these things." Given the small variance of ASAP scores across schools, this statement was an overinterpretation and undue cause for concern. Principal was perplexed at Hilldale not being the highest ranked school, since ASAP was closest to "what we teach."

	VALOR	FRANKLIN	PINE	HILLDALE
PROSPECTS FOR SECOND YEAR CHANGE	Given the laissez-faire district culture, some incremental changes are possible among teachers who are already predisposed to holistic instruction and performance assessment, but little change should be expected among teachers with other predispositions. No strong organizational values act as barriers to change. Neither are there supportive values or mechanisms. The lack of resources for curriculum and professional development also mitigate against change in the immediate future.	The inherent conflicts between ASAP principles and local beliefs about the nature of pupils, etc. make change in year 2 unlikely. Some accommodation may occur with great cognitive dissonance. After ASAP scores came out, administrators conceded that they might have to pay more attention to ASAP scores.	Support for ASAP-consistent instruction and assessment are overwhelmed by the top-down organizational culture, tightly structured connection between curriculum, testing, and management philosophy, strong testing bureaucracy, absence of time and support for alternatives. Incremental changes may occur, such as incorporating Forms A-C in district assessment of writing, yet wholesale paradigm shift will have to occur to make significant changes.	District goals, organizational culture, testing program are consistent with the ASAP reform, though not as a consequence of that reform. Resources are available to continue to develop teachers and curriculum. Scope and sequence and text book adoption have been aligned with ASAP. A critical mass exists of teachers and administrators that have training in whole language and conceptual math. These factors suggest that prospects are positive for second year movements toward ASAP goals. Negative foreshadowing concerning the concentration by teachers on those aspects of ASAP that are scored, and possible narrowing of focus toward them.

	VALOR	FRANKLIN	PINE	HILLDALE
DISTRICT IMAGE OF PUPIL	The most salient image is that pupils with economic disadvantages have DEFICITS that schools must address and that hold back progress toward higher order thinking instruction. Students are perceived as having levels of ability and motivation that exist independent of teaching efforts. "I don't understand the standard for higher order thinking skill. The student down here is thinking at the highest level he can possibly think."	Pupils who are highly disadvantaged are unlikely to benefit from reform efforts. Disadvantage is the driving force that shapes goals and activities, including the emphasis on test scores on basic skills. Low scores are attributed to poverty, an enduring quality irrespective of teacher efforts. Pupils are empty receptacles when they come to school. Pupils here can't transfer or generalize.	Pupils have particular levels of learning needs that must be diagnosed, then must be given sufficient opportunities to practice the skill until it is mastered. Focus is accurate mastery rather than understanding. Learning is extrinsically motivated. Convergent learning is more important than divergent. If skills are not repeated, they will be forgotten.	Pupils bring with them a variety of knowledge, interests, and capabilities that must be considered in designing curriculum and school organization, by those closest to them. Pupils are active meaning-makers. Pupils need thinking and problem-solving skills to help them succeed in the world. Pupils' social and psychological needs are as important as their academic needs.
DISTRICT IMAGE OF TEACHER	Officials believe that teachers should all be following common objectives and materials but impose few mechanisms to ensure that. They believe that low test scores are sometimes good for teachers in that they provide chances for self-evaluation and revision. Teachers are generally on their own, for better or for worse. As a result, teachers vary, and their own beliefs and images about pupils, teaching, and curriculum are more salient than district images.	Each teacher's scores are public, and teachers with low scores are told to just "fix it." Teachers are evaluated by scores. Teachers with highest CRT scores are designated as "master teachers." They are given release time to develop skill-building activities for other teachers to use.	Teachers are the transmitters of curriculum. Their role is to diagnose "the correct level of difficulty" each pupil needs in relation to a required curriculum component and devise strategies to make pupils successful at that level. Teachers provide subskills practice in anticipation of district testing. Teachers regarded as emotionally overreacting to test emphasis in district; may not have enough competence to teach higher order thinking curriculum.	Teachers ultimately govern what goes on in classrooms, and teachers will change toward ASAP style instruction only if their philosophies are consistent with it. Change comes through building teachers' knowledge and demonstrating what can work. Teachers can be leaders in supporting the changes that other teachers should make. ASAP mandate can serve as a stimulus or guide to provoke teachers' experimentation and reflection. Ideally, teachers provide instructional opportunities and materials for pupils to use in constructing meaning.

DISTRICT IMAGE OF CURRICU-LUM	VALOR	FRANKLIN	PINE	HILLDALE
	District officials would prefer a curriculum that is standardized and one that results in an accumulation of specific skills. "Without a district curriculum, teachers get married to a text" and don't think beyond it to a coherent curriculum. Curriculum should be aligned to common state and district objectives and tests. Five years earlier, preferences for whole language teaching were explored through a series of staff development activities. But momentum was lost and no funds were available to change texts and district tests. Most teachers ignore or are unaware of district curriculum. No textbook adoption procedures. Principal feels curriculum is district's weakest area.	There is no curriculum other than what is on the CRTs. Principal: "If you're teaching something that isn't there, either quit teaching it, or else" write an item. Basic skills must be mastered before thinking skills. Constant repetition and drill are considered essential. Skills are taught separately, not integrated. Reading instruction is highly tracked. Progress through grades influenced by scores.	Curriculum guide is centerpiece for controlling what goes on in schools. A pie chart specifies how much instructional time should be spent in each of 12 curriculum components. 59% is to be spent in reading and communication arts (the components covered by CRTs) at the third grade. The curriculum is packed, leaving little room for divergence. Skills are hierarchically arranged. Even a reading series with "Integrated" in its title is used by fragmenting literacy into separate subskills and practicing those to mastery levels.	AZ Essential Skills plus district objectives provide a framework within which teachers select appropriate materials and instructional opportunities. Writing process is pervasive across curricular areas. Schools choose texts and curriculum packages. Curriculum should be integrated because that is what is meaningful to students (although secondary teachers resist the idea). Some key district officials subscribe to constructivist ideals. The message to principals that, although district has adopted Open Court (direct instruction), schools should do whatever makes sense and succeeds for their particular population.

	VALOR	FRANKLIN	PINE	HILLDALE
DISTRICT IMAGE OF ASSESSMENT	Officials believe that tests should be aligned with local curriculum to be valid. ASAP fits this criterion, but is problematic because of its costs and feasibility and because the state does not mandate (or pay for) administration of ASAP at each grade level. "The concept is great but the feasibility is humongous." They believe that assessment should provide comparative data, but also should reflect "what we teach" at every grade level; should also show mastery of basic skills but also diagnose pupil needs for teachers' use; yet should be inexpensive and feasible.	Externally mandated tests are culturally and linguistically biased. Only locally constructed tests are valid for "our population." Testing and curriculum must be matched because pupils can't transfer or generalize. Teachers are told to do whatever it takes to raise CRT scores.	Assessment is for continuous monitoring of progress and is inextricably linked to curriculum. Document: "The foremost purpose of district-wide testing of ... objectives is to evaluate and monitor student growth, document that growth ... help the District evaluate instructional program effectiveness, and ... partially evaluate teacher effectiveness." A district administrator has sole responsibility for the testing program and actively and effectively promotes that agenda. Administrator: "Assessment must be tied to objectives, systematic and objective." Validity is content validity and demonstrated pre-post gain (sensitivity to instruction). Teacher accounts aren't credible for making systemwide conclusions and decisions. ASAP also does not likely to meet that standard.	Officials believe that instruction and assessment should be integrated and ASAP supports this. Test results reflect in part the socioeconomic composition of the community. Still, test results should be high to show the district or school is doing well. Test scores play only a small part in teacher or principal evaluation, but a pattern of low scores might indicate need to improve instruction. High ASAP scores would show that districts have changed in the right direction; i.e., toward holistic, real-world, problem-solving education (consequential validity). But, officials reserve judgment on ASAP, waiting to see how the state will report or use scores. Ranking schools by ASAP scores or using scores to indict teachers will alter reactions and encourage teaching toward the test in possibly inappropriate ways.
COMMUNITY CHARACTERISTICS (SIZE AND TYPE)	School is located in an unincorporated town of 600 residents, within 100 miles of Phoenix, near an Indian reservation. Economy is predominantly agricultural.	School is located in urban core. A few small, well-kept homes are near government projects, shacks, and homeless shelters.	School is located in a middle-class suburb, amidst well-kept tract homes. The school draws from an attendance area that is much more economically and ethnically diverse than the district average.	School is located in a prosperous, middle- to upper-middle-class urban/suburban community.

	VALOR	FRANKLIN	PINE	HILLDALE
COMMUNITY SES	Community is predominantly lower-middle to lower class. Unemployment rates are high due to fluctuation in agricultural industry.	One of the most impoverished communities in AZ.	School serves a mixture of middle-class and upper-middle-class (professional and management), as well as a poor enclave of poorer, minority "apartment" dwellers.	School serves middle- to upper-middle-class families, most of whom believe that "their children are all gifted."
PARENT INTEREST AND PARTICIPA-TION	Parents want to know how students compare with students in the rest of state and nation. "There's a big interest in test scores." Parents rarely question school's authority. Many are undereducated and unfamiliar with the workings of school. Parents want to know how students perform because they accept the authority of the test scores.	About 25% of district families take an active interest in the school and participate in school activities. Majority are only minimally involved. Teachers: "Parents have little interest in test scores here."	The upper-middle-class parents are actively involved in school life and intensely interested in test scores. Principal and experimental reading program designed to raise low scores were the focal point of conflict during year of study. Principal: "Parents came breaking down doors, with newspaper articles in their hands, saying what's the deal? Why are we always lowest? And we had to respond to that and make efforts to make the scores look better."	Parents participate in school activities at high rates and are extremely interested in test scores. Principal, however, plays down standardized test scores and actively tries to educate parents on alternative ways of indicating achievement.
PUPIL SES	75% of pupils at school qualify for free/reduced lunch. Many migrant families from Mexico send children to school here during harvest.	97% of pupils qualify for free/reduced lunch. Administrator: "The most disadvantaged, at-risk students in the state are at this school."	22% of pupils at school qualify for free/reduced lunch program.	School does not participate in reduced price lunch program.
PUPIL ETHNIC COMPOSI-TION	Anglo 41%, Hispanic 35%, Native-American 24%.	Anglo 8%, Hispanic 84%, Native American 2%, Black 6%.	Anglo 72%, Hispanic 20%, Native American 2%, Black 4%, (Unknown 2%).	Less than 5% non-white.

118

	VALOR	FRANKLIN	PINE	HILLDALE
PUPIL LANGUAGE DOMINANCE	Most Native American and Hispanic students come to school with English proficiency because of reservation preschool program or Valor preschool. About 35% of pupils are classified LEP.	Over two thirds of pupils have first language other than English. One half enter kindergarten with little or no English proficiency.	5% of school's pupils are limited English proficient.	Less than 5% have first languages other than English.
SCHOOL ORGANIZATION AND SIZE	K-5 plus federally-funded preschool. 500 students.	Grades 3-8 plus Head Start, migrant education, transition English programs. 550 pupils.	Grades prekindergarten to 5. 550 students.	K-5. 900 pupils.
SCHOOL STRUCTURE	Traditional graded structure. Community distribution center provides food and clothing to community. Pull-out programs for Chapter 1 and computer literacy. Class sizes 25-32. Classes are heterogeneously grouped. School policies discourage retention and few teachers practice it except in cases of extreme absenteeism.	Traditional graded structure. Class size averages 22-28. Chapter 1-eligible pupils pulled out for remediation. Retention based on CRT mastery (but attendance and teacher judgment enter in decision); little retention occurs. Instructional aides in all classes Transition classes and diagnostic/counseling services provided to transient students. Community Education classes provided in English literacy, computer ed, family math, and parenting.	Graded structure, but with several two-grade, combined classes. In each class, pupils are grouped for instruction and regrouped periodically based on CRT results. Promotion/retention decisions partially based on mastery test results. Chapter 1 school.	School experiments with various multiage and combined grade (3/4, 1-2) combinations. Enhanced fine arts, physical education and computer lab. Reading Recovery program available, but few are referred.

	VALOR	FRANKLIN	PINE	HILLDALE
GRADE-LEVEL CONFLICT	The disparities among classrooms in accommodation to ASAP are due to differences in teacher beliefs rather than grade-level differences. Third-grade teachers experience a greater burden of testing on instructional time than other grades. First- and second-grade teachers are interested in ASAP but express no sense of responsibility for it.	Third-grade teachers feel victimized by the need to satisfy conflicting demands and instructional principles (accountable for ASAP and CRT mastery). They feel that rest of school culture is discordant and nonsupportive. When third-grade teachers revised CRTs to be more compatible with ASAP, fourth-grade teachers reversed the changes, making them even more emphasizing basic skills.	Third-grade teachers tend to dismiss the importance of ASAP because it is not administered at every grade. Teacher: "We tried to get the second-grade teachers to try it, but they rejected it because it wasn't required." Third-grade teachers feel the test demands on them are doubled because of ASAP mandate, or would be if district cared about its results.	Each grade level above K does some ASAP administration. Third-grade teacher: "It took a lot of pressure off us knowing that it wasn't just our responsibility."
PRINCIPAL ROLE	Principal plays a laissez-faire role with respect to ASAP, curriculum, etc. Sees her role as one of support and encouragement but, in rare instances, feels she has to coerce some teachers to perform. Believes in Hunter's Elements of Effective Instruction model.	Role of principal is to centralize authority over curriculum and teaching methods by evaluating teachers based on CRT scores, aligning instruction and assessment, and driving out untested content. A behaviorist with respect to both pupils and teachers—reinforces correct behaviors; exposes and shames incorrect behaviors; ignores process.	Principal is beleaguered because of the relatively low position of his school in the district distribution of scores. He serves as a conduit in a top-down hierarchy from district to teachers, conveying exactly the demands and expectations of the district. He has so little room to maneuver that he ignores ASAP altogether, along with any other goals and activities not specifically endorsed by district. He frequently reminds teachers of the importance of CRTs. Defines ASAP as alien to Pines teachers: "Teacher is teaching one method or style but testing method is another way."	Principal is a catalyst for change in directions ASAP seems to support. She can "play the testing game," but prefers to de-emphasize inflating scores in favor of authentically changing instruction and curriculum and developing teachers as professionals. Because she believes ASAP is valid as an indicator of constructivist teaching, she was bemused that Hilldale did not score at the top of the district. She is extremely knowledgeable about constructivist teaching and performance assessment and presses the teachers to get on board.

	VALOR	FRANKLIN	PINE	HILLDALE
ACCOMMO- DATION/ RESISTANCE OF PRINCIPAL	Principal accommodates to ASAP in the same way as to any other external mandate, with passivity.	Ignores and resists ASAP reform as irrelevant to local population and inconsistent with beliefs that basic skills must be mastered before thinking skills can be taught.	Principal recognizes that, in the current district culture, ignoring ASAP is safe.	Principal not only accommodates but acts as resource to the district in promoting activities consistent with ASAP.
TEST BURDEN	Chapter 1 requirements mean that ITBS is given to every pupil. CRTs given early and late in year. Sixteen days per year are spent in administering tests in third grade. Preparation time varies among teachers.	CRTs given 3 times/year with preparation time beforehand. ASAP adds to third-grade burden. ITBS in all grades. Six weeks of instructional time taken up in testing in non-ASAP grades, eight weeks for Grade 3.	Teachers sometimes feel that all they are doing is testing, in one form or another.	ASAP is the district test at all levels but K, therefore test burden is light. ITBS only at mandated grades. Some teachers give CRTs at their discretion. District requires (and provides scoring for) an analogies test for determining academic ability.
TEST PREPARA- TION	Amount of preparation varies by teacher, based on perceptions of stakes. All but one teacher perceived ASAP as low-stakes and therefore didn't prepare pupils for it. One teacher had been burned by accountability function of tests in the past and therefore prepped for it.	Since curriculum and teaching are synonymous with CRTs, test preparation is constant. As testing schedule nears, teachers become more anxious and prepare more. Practice consists of timed tests and worksheets that use formats similar to CRT format. This is a daily event in some classes. Practice for ITBS is rare. Little practice ASAP. Form A stayed in box until week before D was to be given.	Since teaching and CRTs are part of the same package, test preparation for CRTs is on-going, becoming nearly frenzied as the spring post-testing schedule approached. 50% of instructional time is spent in test preparation. Preparation for ASAP was minimal. Some teachers didn't even "open the box" containing ASAP practice materials until the week before testing schedule. Focal teacher taught test-wiseness using Form A for 2 weeks prior to test schedule and administered ASAP-like exercises in form of homework. Emphasized ways of getting high rubric scores.	Principal has discouraged staff from using Scoring High. ASAP preparation consists of doing "activities" related to Forms A, B, and C.

	VALOR	**FRANKLIN**	**PINE**	**HILLDALE**
CURRICU-LUM/TEXTS	SRA text series (direct instruction model) is adopted reading text, although that is used only by some teachers. Those with constructivist tendencies embed more literature and de-emphasize the text. Most staff are unaware that a district curriculum even exists.	Curriculum IS the CRTs, said to also reflect AZ Essential Skills. Teachers follow Madeline Hunter's Critical Teaching Skills and have developed structured lesson plans for all skills. Reading text selected for match with CRTs, phonics-oriented basal series. Math text is available, but less important than CRT.	District specifies curriculum in 12 areas, separated, not integrated. Reading program, Success for All, has been substituted at Pines to raise low scores. Phonics-based, oriented to skills and some low-level comprehension. Math text is Addison Wesley (skills-oriented). Everyday Math is a less preferred alternative text that emphasizes problem-solving and writing about it. Text for social studies takes backseat to curriculum "bible." Science has taken a "backseat."	Reading program is literature-based. No basals are used, except by three teachers. Writing process unifies curricular areas. *Math Their Way* and *Math, A Way of Thinking*. Although science and social studies texts are available, teachers tend to incorporate material in thematic units and projects. District provides materials for "hands-on" math and science activities.
TEACHER EXPERIENCE	Two focal teachers with 5 and 3 years of experience. Another had more than 20, with two master's degrees and intentions to become administrator. District has tendency to hire new teachers who can't get into more desirable assignments. After a couple years' experience, they leave. Teachers express pride in themselves as professionals and participants in the community.	Focal teachers had 16 and 30 years' experience. All teachers certified ESL as well as elementary.	Fourth-grade focal teacher has 15 years of experience and a penchant for technological innovations, particularly computers. Third-grade teacher has 6 years' experience and interest in "total quality learning."	Focal teacher had 4 years' experience, all with same principal and constructivist training and continual professional development.

122

	VALOR	FRANKLIN	PINE	HILLDALE
TEACHER COMMITMENT TO HOLISTIC THINKING INSTRUCTION	Because of decentralized structure, teachers varied on all teacher dimensions. Teachers fell roughly into two clusters: concrete-sequential (i.e., behaviorist, skills-oriented) and constructivist (e.g., whole language). Constructivist teachers were moderately committed, but even these have had limited exposure. Commitment a function of limited exposure and knowledge.	Those few who lean toward holistic instruction are viewed as outcasts. Emphasis by school on bit-by-bit instruction of basic skills and CRT content precludes integrated activities or thinking education.	Teachers see themselves as supporting the concept of holistic teaching, yet experience dissonance between the two paradigms. They redefine holistic instruction so as to make it compatible with a skills approach. Teachers engage in "dis-integrating." Teachers called ASAP a "round-about kind of learning." Teachers who favor constructivism rarely get the time or support to put these ideas in practice.	Teacher regards herself as a whole language teacher. Experts in whole language refer to the school as "in transition," i.e., having the rhetoric and moving toward thorough knowledge of holistic principles, but not yet deeply experienced or expert. Her reading program is literature-based and uses literature logs and literature studies. However, the interaction in groups is teacher-directed and focuses on convergent responses (e.g., "How does [author] use adjectives?" Has participated in several years' training in writing process and thematic teaching and is expert in these areas. Constructivist math expertise seems limited to thorough training and several years' experience in *Math Their Way*.

	VALOR	FRANKLIN	PINE	HILLDALE
PRIOR KNOWLEDGE OF ASAP-LIKE INSTRUCTION	The behaviorist teachers have none. Constructivist teachers had been practicing some form of holistic instruction in language and literature for about five years, since a series of workshops and training activities were held. But even constructivist teachers fell back to a basic skills approach to teaching math, because of inadequate scaffolding in the conceptual principles of mathematics.	Some Grades 1 and 2 teachers express some awareness, but only one is expert. "It took me ten years to [get to] the point of being comfortable with whole language, and I'm not as good as some." Prior knowledge is actually negative, as an earlier experiment with whole language instruction failed to raise test scores so was abandoned. Teacher: "We tried it and it didn't work here."	Teachers who claim knowledge actually misunderstand the principles of constructivist teaching; e.g., thinking that process-teaching and assessment cannot be grasped by low-ability pupils. Focal teacher shows she understands integrated, real-world teaching when she teaches a science lesson on owls, but doesn't translate that into teaching reading and writing, because latter is governed by CRTs.	Teachers have had workshops and college courses "for years" on writing process and literature-based reading. Some professional development has taken place in math (use of manipulatives to teach number concepts), but not nearly so extensively. Focal teacher admits less expertise in math. Teacher: "ASAP tests probability, which isn't in district scope and sequence, and we never took a workshop in it."
OPPORTUNITIES FOR RELEVANT PROFESSIONAL DEVELOPMENT	Little course taking, except for teacher pursuing administrative certification. Lack of resources has severely constrained ANY development activities for teachers. Five years earlier ADE staff presented a series of training sessions for teachers in holistic teaching and alternative assessments. But teachers hired since then have not had similar activities, nor has district provided any follow through on earlier training.	Though resources are available to develop teachers and curriculum, they are directed toward activities more consistent with school's basic skills orientation. Teachers say they have not been provided with materials and skills that would allow them to change toward ASAP.	Resources for professional development are ample (1 day /month/teacher), but are directed toward activities consistent with district philosophy (e.g., TQM) rather than ASAP. No professional development in ASAP-consistent instruction was made available to teachers. Staff development is monologues of experts directed toward teachers rather than clinical models. No time for teachers to experiment or collaborate over activities outside the district demands.	Resources for staff development are ample and marshaled toward goals consistent with ASAP. More so for writing and literacy than for math. Absence of competing demands that characterize other sites means that more time is available to experiment, collaborate, and reflect. Teachers serve as resources for each other. Focal teacher has taken a course every semester to enhance her capacity in literacy, literature, integrated curriculum, cooperative learning, etc.

	VALOR	FRANKLIN	PINE	HILLDALE
PRESENCE OF ASAP GATE-KEEPER	ASAP coordinator resigned before end of year. Coordination and communication lacking between anyone knowledgeable and the teachers.	Teachers were unaware there was even a coordinator present in the district. Coordination and communication were thus lacking. No one person took responsibility for informing teachers about ASAP or constructivist education.	ASAP coordinator was school psychologist who reported to the director of testing, a strong CRT and skills-teaching advocate. She provided information about ASAP requirements to teachers, but was not expert in curriculum or alternative assessments. No one teacher or principal stepped forward to be a spokesperson for ASAP. Constructivists were not in leadership positions.	Focal teacher serves as effective gatekeeper. Since the outset of ASAP planning, she has been on state-level and district committees. She regularly attended the state training sessions on ASAP administration and scoring and kept fellow teachers apprised of what the ASAP program meant. She is also trained to be an official scorer.
TEACHER IMAGE OF PUPIL	Concrete-sequential teachers' image of pupils was consistent with district's image: Students are empty containers into which curricular skills are poured. Pupils have innate ability and motivation, irrespective of teachers' efforts. "There is a maturational rate in each individual that they have no control over. And we'll never change that individual who can't grasp the concepts." Learning is a solitary process that occurs within the learner. Constructivist teachers believe students come to school with knowledge, experience, interests that interact with materials and teaching acts.	Teachers' view is consistent with district's: Degree of disadvantage is central. "These pupils" arrive at school with nothing. "Every point they get on those tests, they get from here." Pupils can't transfer or generalize.	Teachers would prefer to teach in constructivist ways, because that is the way pupils learn best (intentional, process- and real-world-oriented, interactive, higher order problem solving, etc.), but these views are suppressed in the dominant district culture, which views pupils in the opposite way. Because of the dissonance, teachers come to accept district image of pupils, that pupil ability is an enduring trait that must be diagnosed and learning activities geared toward. Pupils come to school empty, learn by practicing hierarchi-cally-arranged subskills. Pupil failure is attributed to personal or family deficiencies.	Pupils' interests and prior knowledge vary; therefore starting instruction from "where kids are" means that appropriate instruction must be constructed by teachers within the context, not standardized or determined from "higher up." Pupils' skills develop unevenly and as a consequence of understanding, not as precursor. Pupils' social nature means that learning should be collaborative. Pupils make meaning together. Learning is active. Prior knowledge and experiences and interests are vital elements in arranging learning opportunities for them.

	VALOR	FRANKLIN	PINE	HILLDALE
TEACHER IMAGE OF TEACHING	Concrete-sequential teachers believe that teachers transmit curriculum as received from external authority. They must maintain the proper conditions of order so that private learning can occur. They must provide opportunities for pupils to practice skills repetitively until mastery is achieved. The teacher "needs to be really focusing on the deficits." Constructivist teachers believe that teachers should design instructional activities that interact with pupils' interests and prior knowledge, and social nature, at least in regard to language, literature, and reading (but not in regard to math).	Irrespective of CRTs, teachers feel they can "still shut the door" and determine what happens in class (yet they still comply with district requirements). Teacher: "If there were no guidelines for teachers, it would be chaos."	Although teachers maintain some semblance of control over what happens in the classroom, that diminishes under pressure to raise scores by practicing subskills. Teachers would like to teach in ways consistent with ASAP or constructivist models. Some even come into the district with such competencies, but lose them in the shuffle of packed curriculum and test demands. Teacher: "If I'm thorough teaching the basics, I don't have time for this or that." To do more integration, teachers must teach and test faster and more efficiently, otherwise pupils get cheated and basics will get neglected.	Teachers should start from where kids are and the problems or themes pupils find compelling. Teachers should serve as resources in organizing and accessing information sources, encouraging exploration and problem solving, showing the connections between subject matter, and assessing pupils' growth. Teachers need to collaborate with each other and model that process for pupils. Teachers need to look after pupils' social and psychological development and model effective citizenship. Teachers coach, encourage, make good materials available, and provide quality experiences. Teachers don't see themselves as filling deficits, but neither do they ignore problem areas. Teachers need to be self-directed. "It's actually research and data that has changed my teaching techniques" rather than external pressure or presence of ASAP.

TEACHER IMAGE OF CURRICULUM	VALOR	FRANKLIN	PINE	HILLDALE
	Concrete-sequential teachers believe that curriculum is a hierarchically-arranged collection of skills to be mastered. Even supposedly integrated curriculum is "dis-integrated." The goal of learning should be "getting it right." Integrated, holistic science and social studies programs were systematically ignored or "dis-integrated." Constructivist teachers' goals for pupils are to "get it," that is, to understand conceptually. The nature of learning is social and interactive. Students are meaning-makers. When teachers react to pupils' struggles (under conditions of inadequate prior knowledge) with activities that "add sense," learning is meaningful, integrated, and whole.	Curriculum presents each skill sequentially. Repetition is key to mastery. Emphasis on accuracy, not understanding. Anything not on the CRT is regarded as frill, and there is no time for frills when you have to repeat drill on skills. The few teachers who favor whole language feel frustrated and disenfranchised and conflicted. "I don't have to get a reading book to do reading. I can get a science or social studies lesson and that to me is reading for the day. But that's the way I do it, and I know we're not supposed to do it that way."	Teachers feel that that there ought to be more to curriculum and teaching than what the district requires or that complying with test mandates demands of them. Teacher: "There's much more to teaching … how to write and explore your creativity." But diverging from standard curriculum exacts a price. When the formative test results show their students are falling behind the sequence, their anxiety rises and the pace of activity quickens, until they give up and go back to "playing the game," even though that produces "robots." Curriculum is packed, it is something that comes from above and must be covered. "They're giving me more stuff to do, more shit, but they're not giving me any more time to prepare." Nominally integrated units are dis-integrated. Though using a conceptual math program, teacher invented worksheets to drill pupils to mastery on math facts.	With encouragement of principal, teachers endorse a constructivist image of curriculum: whole language and conceptual math. Subject matter should be integrated, address real-world concerns and issues and pupil interests. Much value placed on interactional and cooperative learning opportunities. Teachers use literature-based reading program and *Math Their Way* and *Math, A Way of Thinking* rather than basals and workbooks. Students form mathematical hypotheses and understanding based on "experiments." Principal tells parents who are "shopping" for schools not to expect seeing many worksheets, as teachers are more interested in authentic problem-solving than drilling to ensure getting the right answer. Writing process permeates and unites all curricular areas. Skills come as a consequence of understanding. Extended project work and thematic units cross disciplinary boundaries and respond to pupils' real-world concerns and interests. No textbooks evident.

127

	VALOR	FRANKLIN	PINE	HILLDALE
TEACHER IMAGE OF ASSESSMENT	Both concrete-sequential and constructivist teachers distinguish between assessment and testing. For concrete-sequential teachers, assessment amounts to pupils' "displaying" correct answers so that instruction can proceed along the hierarchy of skills; or if incorrect, teachers can regroup pupils and repeat instruction. For constructivist teachers, assessment is teacher judgment of pupil understanding and integration into knowledge structure (particular to each pupil). Both groups define testing as that which is done for external audiences. Regardless of the form of the measure (CRT, ITBS, or performance test), teachers orient toward the accountability function the test serves. Testing is a silent, individualistic "display" of correct responses. ASAP, which is supposed to be integral with instruction, is a separate activity, stressful for both teachers and pupils, an add-on to what they would be doing otherwise, noninteractional and nonteacher mediated.	"Our principal has not supported whole language and doesn't care about ASAP results." ITBS is linguistically and culturally biased. Valid assessment is demonstrated growth, but lack of growth or low scores on tests are due to lack of vocabulary, motivation to do well, not necessarily bad teaching. Teachers know pupils best and therefore should determine what gets assessed. But some think that the CRTs "don't mean a thing."	CRT testing is done for accountability purposes, to satisfy external audiences. Teacher: "Parents said, 'we want better schools, dang it.' [So message from principal] is we need to appease this community," whether that is the right thing to do or not. To teachers, CRT is just a reality of this district, something they have to live with. Principal provides frequent reminders of its importance through comparative graphs, updates, and suggestions for instructional strategies for pulling up scores. Teacher on effects of testing: "We zero in on those ... that's what our district wants us to teach." Teacher: "The first year I didn't know a lot about [CRTs] and I didn't really care. I was teaching very whole language and math manipulatives-based. When the kids' test scores came back, they didn't do very well at all, but I knew we had learned but the test didn't show it. So I added more subskills attention and went with that."	Teachers view testing as fundamentally different from assessment. Testing should be objective to be valid; assessment is valid when it is local, teacher-determined, and shows growth. Role of testing is high-stakes accountability to satisfy external audiences; role of assessment is to advance instruction. Testing is getting the right answer. Assessment is more oriented to process, open-ended, not standard responses across pupils and socially constructed. ASAP is something closer to the way teachers teach, so it is more successful and less stressful. ASAP can be a lever to change behaviorist teachers to holistic teachers. There is discontinuity between Forms A, B, C and monitored Form D. A, B, and C are interactive, relaxed, process-oriented, teacher- and peer-mediated, like instruction. D is silent, individual, independent, teacher-centered, disconnected from instruction.

	VALOR	FRANKLIN	PINE	HILLDALE
TEACHER IMAGE OF ASSESSMENT (CONTINUED)	Teachers would prefer to have tests play a diagnostic function rather than an accountability function, to help them better understand pupils' learning problems.		ASAP is a step in the right direction, and more in keeping with better instruction, but is still perceived negatively because (a) it is not aligned with district objectives and CRTs, (b) it is administered in spring and reported too late to provide useful information to teachers, (c) it comes from ADE as just another externally mandated accountability device, and (d) administrative and scoring problems. Instead, testing needs to be teacher-centered, objective, standardized in administration and scoring. "ASAP is just another thing we have to do."	In Form D, teacher worries about what is to be scored and points pupils' attention to that. Warns them not to spend so much time on drawing (unscored) that they won't have enough time for their story (scored). Teachers believed that ASAP was the state's preferable alternative to ITBS. If ASAP fails, they fear a return to ITBS testing. They feel vindicated that what they perceive to be ASAP's philosophy matches what they perceive to be their own philosophy.

129

	VALOR	FRANKLIN	PINE	HILLDALE
TEACHER ACCOMMODATION / RESISTANCE TO ASAP	Teachers tend to comply with whatever authority dictates.	Teachers tend to comply with whatever authority dictates.	Even teachers who favor holistic instruction come to comply with the district's concrete-sequential model. Teacher: "I've done it. I've done my tap dance." Teacher with personal philosophy most consistent with ASAP resigned in frustration and transferred to an alternative school.	The form of accommodation taken is deliberate and tactical. Teachers and principal see the ASAP program as consistent with the direction they were going independent of the reform. They see it as an inducement to concrete-sequential teachers and top-down administrators to think in different (constructivist) ways about pupils and teaching. On the other hand they are wary of the accountability possibilities inherent (later realized) behind ASAP and attend to that which is to be scored as a possible high-stakes instrument on which they will be expected to score high.
TEACHER FAMILIARITY WITH ASAP, RUBRICS, ESSENTIAL SKILLS	Only one teacher, who resigned before year's end, was a trained scorer. Others had one on-site ADE workshop. Most had heard of AZ Essential Skills but had not seen or used them except for those lesson plans they had had to code according to subskills tested on CRTs. Limited use of Forms A, B, and C.	Teachers speak knowledgeably about Essential Skills, because the district CRTs are in part built around them. Only one third-grade teacher seemed to have a strong grasp of ASAP and scoring rubrics.	Teachers received some training during previous year in conjunction with ASAP pilot, but level of awareness is still low. They express understanding of Essential Skills in that district objectives are perceived to have been aligned with them. No one was expert in rubric scoring during this first year. They first got exposure to rubrics during the training they received two weeks before ASAP was to be administered.	Key teachers are well versed in ASAP and scoring rubrics. They thus are savvy about focusing pupils' attention on what is to be scored and have adapted a test-wiseness strategy to maximize scores. Teacher reminded pupils to use complete sentences, for example, because she knew that would make a difference between getting a score of 2 and 3. "Giving the test is much easier if you start with the scoring rubric and work backwards" because that means you are giving the scorers exactly what they want.

130

APPENDIX B

ANALYTIC MATRIX

	VALOR	FRANKLIN	PINE	HILLDALE
RESOUR-CES FOR CHANGE	MATERIAL: Impoverished. KNOWLEDGE: No gatekeepers. Incompatible curriculum. No professional development.	MATERIAL: Adequate, but dissonant with ASAP. KNOWLEDGE: Scattered, marginalized. No gatekeepers. Incompatible curriculum and professional development.	MATERIAL: Rich, but dissonant with ASAP. KNOWLEDGE: No gatekeepers. Incompatible curriculum. Incompatible professional development.	MATERIAL: Rich, consonant with ASAP goals. KNOWLEDGE: Several knowledgeable gatekeepers. Professional development in sync with ASAP.
POWER TO CHANGE	Laissez-faire organization. School and teacher discretion.	Top-down power, centralized district, no teacher power.	Policy-driven, top-down, centralized, standardized. No teacher or principal power.	Site-based power, principals lead and persuade teachers. District encourages, supports. Teachers have power.
ASSUMP-TIVE WORLDS	DISTRICT: Concrete-sequential. TEACHERS: Vary. Dominant image of pupil as disadvantaged/deficit.	DISTRICT: Concrete-sequential. TEACHERS: Concrete-sequential or marginalized. Dominant image of pupil as disadvantaged/deficit.	DISTRICT: Concrete-sequential image drives everything. TEACHERS: Concrete-sequential or marginalized.	DISTRICT: Constructivist. TEACHERS: Constructivist or marginalized.
ROLE OF TESTING	Low stakes, low expectations, low burden. ASAP is just add-on accountability requirement.	High stakes on district CRT. Low expectations. Rejection of ASAP as inappropriate for disadvantaged. ASAP adds to high test burden.	Low expectations. Very high stakes on district CRT. Very high test burden. ASAP adds to test burden.	High expectations, but moderate stakes from district. Low test burden. ASAP fits instruction, but still fills accountability function.
STATUS AT END OF YEAR 1: REACTION	Teachers accommodate by "dis-integrating" integrated tests and lessons. No capacity building. Resistance: constructivist teacher resigns.	Passive acquiescence. Disinte-grating. Added test burden creates frustration. Constructivist teachers isolated. No capacity building.	Minimal, symbolic accommoda-tion. Contrary philosophy still dominant. No capacity building. Constructivist teacher resigns. Test-wiseness training.	Alignment of district scope and sequence to ASAP. Teachers vindicated by perceived match between their philosophy and ASAP. Test-wiseness training. Constructivism NOT CAUSED BY ASAP. Capacity building.

APPENDIX C

TEACHER QUESTIONNAIRE

South West Educational Policy Studies

April 15, 1994

Dear Colleague:

As an Arizona teacher, you know that this is a critical time for educational voices to be heard. Decisions are being made by legislators and others that affect all of us. The Arizona Student Assessment Program (ASAP) is one key element in Arizona's efforts to improve schools. We are requesting your participation in a research study we are conducting about the role of ASAP at your school. Through the study, we hope to gain a better understanding of what mandated tests such as ASAP mean to school professionals and the factors that influence the implementation of educational reforms.

Your school has been selected in a scientifically designed sample so that your views will represent teachers throughout the state. Of course, your participation is voluntary. Because of the nature of the design, however, the results will be much more meaningful and the message to policy makers much more effective if you take part. All responses will be held in strict confidence, and no individual or school identities will be revealed. A copy of the report will be made available to you at the end of the study. In addition, we are offering a small monetary honorarium for your school's participation.

The organization conducting the study is Southwest Educational Policy Studies, a nonprofit educational research organization based in Arizona. We represent over 30 years combined research experience and 20 years of professional practice in public schools. We are an independent group, not affiliated with any government agency or professional association. It is our intention to consider all sides of the issue of state assessment, use the best research procedures, and represent your viewpoints as forcefully as possible. The results of this study will be disseminated widely to key groups of policy makers and others, including Arizona legislators, the State Board of Education, candidates for State Superintendent of Public Instruction, professional associations, teacher education programs, and policy researchers. The study is funded by the Center for Research on Evaluation, Standards, and Student Testing at the University of California, Los Angeles, which studies the impact of testing on public schools nationally.

The enclosed questionnaire takes between 15 and 20 minutes to complete. It consists of statements about teaching, learning, testing, and school practices. These statements were derived from data of an earlier study, in which we spent a year in several classrooms to understand the role of mandated assessment. For example, the statement "Using ASAP means that there are just more things that teachers have to cover" comes directly from teachers in schools that we observed. The statements represent a range of viewpoints that we discovered.

Please respond to the statements according to the scales provided in each section. The final section asks for your comments on ASAP-related issues that you feel have not been adequately addressed in the statements and questions. Be assured that we will closely analyze and document each comment. When you are finished, enclose the questionnaire in the envelope provided and seal it. In this way we can insure that no one at your school can associate your identity with the information you provided. Your principal will collect the envelopes and return them to us.

We are aware how busy you are, particularly at this time of year. However, we hope you agree to participate in the study, to lend your voice to the debate about state testing. If you have any questions or comments about the study, call and leave a message. Your reactions are important to us, and we are grateful for your help.

Sincerely,

Mary Lee Smith, Ph.D.
Project Director

Audrey J. Noble, M.Ed.
Research Director

LaVern Tarkington, M.Ed.
Research Associate

3646 E. Ray Road, Ste. B16-125   Phoenix, AZ 965-1684

Arizona Student Assessment Program Survey

SECTION A. BELIEFS ABOUT TEACHING, LEARNING, AND ASSESSMENT

This section contains statements of beliefs about teaching, learning, and assessment. Please indicate the extent to which you agree or disagree with each of the following statements.

> *Circle the number that corresponds to your agreement using the following scale.*
> *(5=Strongly Agree...3=Don't Know/No Opinion...1=Strongly Disagree.)*

		Strongly Agree	Agree	Don't Know/No Opinion	Disagree	Strongly Disagree
1.	If the district curriculum is aligned with the Arizona Essential Skills, then every student will have equal opportunity to learn challenging and important material.	5	4	3	2	1
2.	ASAP represents the best we know about how students learn.	5	4	3	2	1
3.	Pupils at this school need to master skills before they can progress to higher-order thinking and problem solving.	5	4	3	2	1
4.	When ASAP is used as it was designed to be used, it can be integrated into daily instruction.	5	4	3	2	1
5.	Using ASAP means that there are just more things that teachers at this school have to cover.	5	4	3	2	1
6.	Teachers at this school are now competent to teach the type of instruction compatible with the ASAP performance assessment.	5	4	3	2	1
7.	Some pupils will never be able to profit from integrated, thematic instruction.	5	4	3	2	1
8.	The primary role of teachers should be to teach the curriculum that the district defines.	5	4	3	2	1
9.	When I get an integrated unit or test such as ASAP Form A, I have to break it down into its separate parts and teach the skills in each part so that students will be able to get the right answers.	5	4	3	2	1
10.	Many students at this school have a very difficult time transferring and generalizing what they have been taught.	5	4	3	2	1
11.	The Arizona Essential Skills represents high standards of achievement.	5	4	3	2	1
12.	The Arizona Essential Skills is the curriculum framework that all Arizona teachers should follow.	5	4	3	2	1
13.	A student should achieve a minimum mastery score on the ASAP to be graduated from high school.	5	4	3	2	1

SECTION B. POSSIBLE EFFECTS OF THE ASAP PROGRAM

This section contains statements which represent possible effects of the ASAP program. Please indicate the extent to which you agree or disagree with each of the following statements.

> *Circle the number that corresponds to your agreement using the following scale.*
> *(5=Strongly Agree...3=Don't Know/No Opinion...1=Strongly Disagree.)*

	Strongly Agree	Agree	Don't Know/No Opinion	Disagree	Strongly Disagree
1. I don't understand ASAP well enough to adapt my teaching to it.	5	4	3	2	1
2. ASAP has had little or no effect on my teaching.	5	4	3	2	1
3. As a result of ASAP, major changes in curriculum have been made at school.	5	4	3	2	1
4. ASAP has given me moral support to do the kind of holistic teaching that I have been doing already.	5	4	3	2	1
5. ASAP takes away from instructional time teachers should be spending on something more important.	5	4	3	2	1
6. As a result of ASAP, many teachers at this school have changed the way they think about teaching and learning.	5	4	3	2	1
7. I feel like I am struggling to do BOTH a skills-oriented type of instruction AND a holistic-type of instruction.	5	4	3	2	1
8. The texts and materials that the district requires are compatible with ASAP.	5	4	3	2	1
9. Most or all of the tests this district requires have standardized formats (e.g., multiple choice, norm- or criterion-referenced).	5	4	3	2	1
10. Because of ASAP, I am learning how to teach with thematic units and integrated subject matter.	5	4	3	2	1
11. Because of ASAP, I am learning how to teach so that students will be able to solve complex, real-life problems.	5	4	3	2	1
12. Because of ASAP, our school has brought in consultants and experts in curriculum and instructional methods compatible with ASAP.	5	4	3	2	1
13. Many teachers are fed up with the outside demands and just hope ASAP will go away.	5	4	3	2	1
14. Our district has revised its Scope and Sequence (curriculum) to be aligned with ASAP.	5	4	3	2	1
15. Our district still uses the scores of standardized tests (ITBS, TAP, CRT) to compare schools and/or evaluate teachers.	5	4	3	2	1
16. Our district uses ASAP scores to compare schools and/or evaluate teachers.	5	4	3	2	1
17. Teachers feel pressure to raise students' ASAP scores.	5	4	3	2	1
18. Unless the state equalizes resources among schools, the ASAP program will not be able to reduce the differences in achievement between disadvantaged and nondisadvantaged students.	5	4	3	2	1

	Strongly Agree	Agree	Don't Know/No Opinion	Disagree	Strongly Disagree
19. Overall, the benefits of ASAP testing are worth the investment of effort, time, and money it takes to administer the program.	5	4	3	2	1
20. Overall, the benefits of the <u>entire</u> state testing program (including ASAP, ITBS/TAP, and state-mandated district tests) are worth the costs.	5	4	3	2	1
21. ASAP is just another fad that will disappear.	5	4	3	2	1

SECTION C. VALIDITY AND MEASUREMENT

This section contains statements about the validity of ASAP performance test and what it measures. Please indicate the extent to which you agree or disagree with each of the following statements.

> *Circle the number that corresponds to your agreement using the following scale.*
> *(5=Strongly Agree...3=Don't Know/No Opinion...1=Strongly Disagree.)*

	Strongly Agree	Agree	Don't Know/No Opinion	Disagree	Strongly Disagree
1. What ASAP measures is about the same as what ITBS/TAP measures.	5	4	3	2	1
2. ASAP scores accurately reflect a student's mastery of basic skills.	5	4	3	2	1
3. ASAP scores accurately reflect a student's ability to solve authentic, real-world math problems.	5	4	3	2	1
4. ASAP scores accurately reflect a student's ability to read for understanding.	5	4	3	2	1
5. ASAP scores accurately reflect a student's ability to write effectively.	5	4	3	2	1
6. ASAP scores accurately reflect the quality of education a student has received.	5	4	3	2	1
7. ASAP scores reflect things outside the school's control, such as intelligence, parental interest in education, or emotional adjustment.	5	4	3	2	1
8. Students have adequate motivation to perform on ASAP at a level that accurately reflects achievement.	5	4	3	2	1
9. No test, not even the ASAP, is as valid as a teacher's judgment based on the student's performance on class assignments.	5	4	3	2	1
10. Students with poor reading skills will not be able to perform on the ASAP math portion at a level that accurately reflects their math achievement.	5	4	3	2	1
11. The differences between minority and nonminority pupils that exist on traditional standardized achievement tests also exist on the ASAP performance assessment.	5	4	3	2	1
12. ASAP scores, like other test scores, reflect in part the socioeconomic background of pupils and schools.	5	4	3	2	1
13. The ASAP is free of gender bias.	5	4	3	2	1
14. The ASAP is fair to all cultural groups.	5	4	3	2	1

138

	Strongly Agree	Agree	Don't Know/No Opinion	Disagree	Strongly Disagree
15. The time limits on ASAP Form D unfairly penalize many students at this school.	5	4	3	2	1
16. Many children who are acquiring English as a second language will score poorly on the ASAP even if their educational achievement is adequate.	5	4	3	2	1
17. Confusing directions and procedures pose problems for the administration of ASAP.	5	4	3	2	1
18. Since ASAP is scored by a person rather than by a machine, its results cannot objectively or accurately measure achievement.	5	4	3	2	1
19. When most teachers at this school prepare their students to take ASAP, they encourage them to spend most of their time and energy on those activities that will earn high scores.	5	4	3	2	1

SECTION D. CONDITIONS AT PUBLIC SCHOOLS

This section contains statements that describe conditions at public schools. Please rate each statement according to how accurately it describes YOUR OWN SCHOOL.

> *Circle the number that corresponds to the degree of accuracy using the following scale.*
> *(5=Very Accurate...3=Somewhat Accurate...1=Not at all Accurate.)*

	Very Accurate		Somewhat Accurate		Not at all Accurate
1. Most changes of curriculum, instruction, and classroom organization are made by teachers on their own.	5	4	3	2	1
2. Most changes of curriculum, instruction, and classroom organization are made by the district officials and handed down to the teachers.	5	4	3	2	1
3. Most changes in curriculum, instruction, and classroom organization are decided at the school by the principal, department head, or by teachers acting together.	5	4	3	2	1
4. A departmental organization (e.g., English, Math) makes curriculum integration difficult.	5	4	3	2	1
5. Teachers of different grade levels seldom plan together.	5	4	3	2	1
6. Teachers have ample time to experiment, collaborate and reflect on possible changes in their teaching.	5	4	3	2	1
7. There are so many classroom requirements that teachers have little freedom to teach what they believe is important.	5	4	3	2	1
8. There is so much pressure for high test scores that teachers have little time and energy to teach anything that is outside the scope of required tests.	5	4	3	2	1
9. The district and school have inadequate funds to purchase texts and curricular materials consistent with ASAP.	5	4	3	2	1

		Very Accurate		Somewhat Accurate		Not at all Accurate
10.	Some or all of the teachers are knowledgeable and experienced about holistic, integrated, problem-solving instruction.	5	4	3	2	1
11.	None of the teachers has had much training or experience with performance (alternative) assessments.	5	4	3	2	1
12.	There is at least one person at this school that teachers can turn to for accurate information about the ASAP program.	5	4	3	2	1
13.	Most teachers see themselves as agents of change rather than objects of change.	5	4	3	2	1
14.	Adequate professional development has been provided for teachers to make changes necessary to implement ASAP.	5	4	3	2	1
15.	The principal is supportive of integrated, holistic, problem-solving education.	5	4	3	2	1
16.	Many teachers disagree with the philosophy of curriculum, instruction, and assessment that ASAP represents.	5	4	3	2	1
17.	There is an adequate level of resources to provide lifelike learning experiences for students.	5	4	3	2	1
18.	Students at this school have adequate experience with alternative or performance assessments, so that the ASAP format is familiar to them.	5	4	3	2	1
19.	Students have adequate opportunity to learn higher order thinking skills.	5	4	3	2	1
20.	Students have adequate opportunity to combine reading and writing tasks.	5	4	3	2	1
21.	Students have adequate opportunity to write about science and/or social studies topics.	5	4	3	2	1
22.	Students have adequate opportunity to write about their solutions to math problems.	5	4	3	2	1
23.	Students have adequate opportunity to engage in long-term projects and/or investigations.	5	4	3	2	1
24.	Students have adequate opportunity to pursue topics of their own choice.	5	4	3	2	1
25.	Students have adequate opportunity to engage in lifelike learning experiences.	5	4	3	2	1

SECTION E. TIME FOR REQUIRED TESTING

1. Considering ALL the testing you are required to do (including ITBS/TAP, ASAP, District CRT's), please estimate the number of hours you spend during a school year on ADMINISTERING these tests. Circle your most accurate estimate.

    5 hours or fewer   6-10   11-15   16-20   21-25   More than 25 hours

2. Estimate the number of hours you spend preparing students for these required tests during a school year. Circle your most accurate estimate.

    10 hours or fewer   11-15   16-20   21-25   26-30   31-35   36-40   41-45   46-50   More than 50 hours

SECTION F. PROFESSIONAL DEVELOPMENT AND KNOWLEDGE RELATED TO ASAF

1. Indicate the amount of training (workshops, staff development) you have received about ASAP itself (for example, how to administer and score it).

    5 hours or fewer   6-10 hours   More than 10 hours

2. ASIDE FROM THE TRAINING YOU LISTED ABOVE, indicate the amount of professional training, staff development and credit courses you have completed in teaching consistent with ASAP. (Consider such things as courses, in-service sessions or workshops on thematic units, conceptual math, integrating writing and reading, etc.) Please circle the approximate TOTAL number of hours of training you have received over the past two years.

    5 hours or fewer   6-10   11-15   16-20   21-25   26-30   31-35   36-40   More than 40 hours

3. Please describe the most recent professional development of any type in which you participated.

    • Topic addressed: _____

    • Date of activity: _____   Duration (hours or days): _____

    • What type of person led the activity (e.g., fellow teacher, district curriculum coordinator, consultant, professor, other)
    _____

    • Your judgment of the value or relevance of the activity to your teaching. _____
    _____

4. Circle the number that corresponds to your level of knowledge of each of the following. Use the scale provided below.

> *(1=None...2=Awareness Only...3=Good Understanding   4=Expert.)*

		None	Awareness Only	Good Understanding	Expert
a.	Arizona Essential Skills	1	2	3	4
b.	Forms A, B, and C of the ASAP performance test	1	2	3	4
c.	The rubrics used to score the ASAP performance test	1	2	3	4
d.	The District Assessment Plan filed by your school district	1	2	3	4
e.	The Writing Process or Writing Across the Curriculum	1	2	3	4
f.	Literature-Based Reading Instruction	1	2	3	4
g.	Conceptual or Everyday Math (e.g., *Math Their Way*)	1	2	3	4
h.	Integrated, thematic teaching units	1	2	3	4
i.	Cooperative Learning	1	2	3	4
j.	Direct Instruction of Basic Skills	1	2	3	4
k.	Psycholinguistics	1	2	3	4
l.	Constructivist learning theory	1	2	3	4
m.	Student generated research projects	1	2	3	4
n.	Hands-on science activities	1	2	3	4
o.	Probability and statistics	1	2	3	4
p.	Problem-solving or critical thinking	1	2	3	4
q.	Portfolio assessment	1	2	3	4
r.	Alternative or classroom performance testing	1	2	3	4

SECTION G. DESCRIPTION OF PARTICIPANTS

1. How many years of teaching experience have you had (including the current year)? _____ Years

2. Please describe your current teaching assignment.

    Grade level(s): _____

    Subjects taught: _____

3. Have you...   (Please circle the most appropriate response for each of the following.)

    ...participated in the administration of ASAP?   No   Yes

    ...TRAINED as a scorer for ASAP?   No   Yes

    ...SERVED as a scorer for ASAP?   No   Yes

    ...had any other role in the development
       or implementation of ASAP (e.g., such as
       ASAP coordinator)?   No   Yes

4. Gender: _____

5. Age: _____

6. Ethnic Group: _____

SECTION H. PARTICIPANT REACTIONS

A program such as ASAP is complex. So are responses to it. If issues important to you have not been covered in the questionnaire so far, you may use the space below to describe them.

Thank you for your participation. If you have experienced any problems in connection with the survey or wish to comment further, please call and leave a message at 965-1684.

APPENDIX D

FACTOR ANALYSIS TABLES

Table 1

Variable Characteristics of "Self-Rated Knowledge"

Number of items	18
Cronbach's alpha	.88
Items	Arizona Essential Skills Form A, B, C of the ASAP performance test Rubrics used to score ASAP performance test Your district's District Assessment Plan Writing process or writing across the curriculum Literature-based reading instruction Conceptual or everyday math Integrated, thematic teaching units Cooperative learning Direct instruction of basic skills Psycholinguistics Constructivist learning theory Student generated research projects Hands-on science activities Probability and statistics Problem-solving or critical thinking Portfolio assessment Alternative or classroom performance assessment
Mean = 45.84 Standard deviation = 7.9	Skewness = -.23 n = 1229

Table 2

Variable Characteristics of "Belief in the Validity of ASAP and Essential Skills"

Number of items	11
Cronbach's alpha	.87
Items (paraphrased)	ASAP represents the best we know about how students learn Arizona Essential Skills represent high standards of achievement Arizona Essential Skills is the curriculum framework all teachers should follow A student should score a minimum mastery on ASAP to graduate from high school ASAP reflects basic skills achievement ASAP reflects ability to solve real-world math problems ASAP reflects ability to read with understanding ASAP reflects ability to write effectively ASAP reflects the quality of education a student has received ASAP reflects things such as emotional adjustment of student Students have adequate motivation to perform well on ASAP
Mean = 30.4 Standard deviation = 8.19	Skewness = -.20 n = 1355

Table 3

Variable Characteristics of "Opportunity to Learn"

Number of items	9
Cronbach's alpha	.89
Items (paraphrased)	Resources at this school are adequate to provide lifelike learning experiences for students Students at this school have adequate opportunity to learn higher order thinking skills and complex problem solving Students have adequate opportunity to combine reading and writing Students have adequate opportunity to write about math problems Students have adequate opportunity to write about science topics Students have adequate opportunity to engage in long-term projects Students have adequate opportunity to pursue topics of their own choice Students have adequate opportunity to engage in lifelike learning experiences
Mean = 27.48 Standard deviation = 6.81	Skewness = -.01 n = 1339

Table 4

Variable Characteristics of "Accountability/Authority Culture"

Number of items	6
Cronbach's alpha	.61
Items (paraphrased)	Most changes in curriculum are made by the district and handed down to teachers There are so many classroom requirements that teachers have little discretion District uses standardized tests to compare schools and evaluate teachers District uses ASAP to compare schools and evaluate teachers Teachers feel pressure to raise ASAP scores
Mean = 19.13 Standard deviation = 3.80	Skewness = -.28 n = 1339

Table 5

Variable Characteristics of "Belief in the Beneficial Effects of ASAP"

Number of items	9
Cronbach's alpha	.78
Items (paraphrased)	ASAP has had little or no effect on my teaching (-) I am learning how to teach with integrated units I am learning how to teach complex problem solving in math Our school has brought in expert consultants Our district has revised Scope and Sequence to be aligned with ASAP Benefits of ASAP are worth the investment Benefits of entire testing program are worth investment Major changes in curriculum have been made ASAP has given me moral support to do the kind of holistic teaching I was already doing
Mean = 24.19 Standard deviation = 6.07	Skewness = -.08 n = 1349

Table 6

Variable Characteristics of "Belief in the Dis-Equities of ASAP"

Number of items	8
Cronbach's alpha	.73
Items (paraphrased)	Unless state equalized district sources, achievement differences will persist Students with poor reading will perform poorly on ASAP math Minority/nonminority differences on standardized tests will also show up on ASAP ASAP reflects in part SES differences ASAP is fair to all cultural groups (-) Time limits on ASAP penalize students here LEP students will score poorly on ASAP Confusing directions and procedures plague ASAP
Mean = 29.07 Standard deviation = 4.69	Skewness = -.43 n = 1351

Table 7

Variable Characteristics of "Opportunities for Teachers to Learn"

Number of items	10
Cronbach's alpha	.74
Items (paraphrased)	Teachers at this school are competent to teach consistent with ASAP Teachers here are knowledgeable about holistic education No teacher is expert with performance assessment (-) At least one teacher has information about ASAP Adequate professional development has been provided for teachers to adapt toward ASAP Principal is supportive of holistic education Students here have experience with performance assessment District has revised its Scope and Sequence District texts are compatible with ASAP
Mean = 30.19 Standard deviation = 4.82	Skewness = -.36 n = 1353

Table 8

Variable Characteristics of "Rejection of ASAP Ideals"

Number of items	6
Cronbach's alpha	.68
Items (paraphrased)	Pupils need to master basic skills before they progress to problem solving Using ASAP just means there is more to cover Some pupils will never be able to profit from integrated, thematic instruction Many students have a difficult time generalizing and transferring ASAP takes away from instructional time that could be better spent ASAP is just another fad that will fade away
Mean = 22.30 Standard deviation = 4.82	Skewness = -.06 n = 1355

APPENDIX E

SAMPLING PLAN TABLES

Table 9

Sample Size Allocated to Stratification Cells of District Type, District Size and Percent Minority Pupil Composition (based on $N = 654, 506$ and $n = 3000$)

| Size | Type | Percent minority | | | | | |
| | | High | | Mid-range | | Low | |
		n	$\%$	n	$\%$	n	$\%$
Large	Unified	56	1.9	372	12	762	25
	Non-unified	80	2.7	342	11	262	8.7
Medium	Unified	167	6	148	4.9	279	9.3
	Non-unified	68	2.3	182	6.1	114	3.8
Small	Unified	38	1.3	19	0.6	19	0.6
	Non-unified	23	0.8	23	0.8	23	1.5

Table 10

Numbers and Percentages ($n = 2696$) of Teachers in the Sample Ultimately Distributed Arrayed by District Type and Size and Minority Pupil Composition

| Size | Type | Percent minority | | | | | |
| | | High | | Mid-range | | Low | |
		n	$\%$	n	$\%$	n	$\%$
Large	Unified	73	2.7	140	5.2	408	15
	Non-unified	244	9	327	12	268	10
Medium	Unified	176	6.5	34	1.3	307	11
	Non-unified	109	4	170	6.3	149	5.5
Small	Unified	63	2	74	2.7	16	0.5
	Non-unified	34	1	58	2.2	46	1.7

Note. Excludes teachers in schools with which no contact was ever made. Includes replacement samples.

Table 11

Response Rates of Teachers and Schools Arrayed by District Type and Size and Minority Pupil Composition (n of teachers = 1360/2696; n of schools = 97/111)

Size	Type	Percent minority		
		High	Mid-range	Low
Large	Schools	73%	88%	91%
	Teachers	40%	61%	49%
Medium	Schools	93%	100%	89%
	Teachers	37%	57%	51%
Small	Schools	88%	71%	86%
	Teachers	69%	43%	56%

Note. Excludes teachers in schools with which no contact was ever made. Includes replacement samples.

ITEM-LEVEL DATA FROM TEACHER QUESTIONNAIRE

Effects Due to Pupil Composition: Group Means and Common Standard Deviations (n =1355)

Variable	Mean for high minority	Mean for mid-minority	Mean for low minority	Common standard deviation
Self-rated knowledge	45.38	46.39	45.63	7.90
Validity of ASAP and essential skills	31.77	29.17	30.76	8.20
Opportunity to learn	26.8	26.53	28.39	6.81
Accountability/Authority culture	19.26	19.30	18.96	3.80
Beneficial effects of ASAP	25.52	23.30	24.20	6.06
Dis-equities of ASAP	29.23	30.54	28.02	4.69
Opportunities for teachers to learn	29.61	29.58	30.83	6.41
Rejection of ASAP ideals	22.18	22.84	21.99	4.82
Time for test administration	4.19	4.07	3.64	1.67
Time for test preparation	5.97	5.99	5.14	3.59
Training about ASAP	1.69	1.58	1.46	.76
Professional development	3.37	3.25	3.10	2.81

Effects for Type of Community

Variable	Mean for urban districts	Mean for suburban districts	Mean for rural districts	Common standard deviation
Self-rated knowledge	46.12	46.87	44.79	7.90
Validity of ASAP and essential skills	31.39	30.38	29.72	8.19
Opportunity to learn	27.18	28.99	26.46	6.80
Accountability/Authority culture	19.96	19.09	18.49	3.80
Beneficial effects of ASAP	24.33	24.29	23.97	6.07
Dis-equities of ASAP	29.33	28.42	29.42	4.69
Opportunities for teachers to learn	30.05	31.30	29.36	6.41
Rejection of ASAP ideals	22.54	21.24	23.01	4.82
Time for test administration	4.16	3.80	3.77	1.67
Time for test preparation	6.06	5.43	5.35	3.59
Training about ASAP	1.58	1.42	1.63	0.77
Professional development	3.15	3.28	3.18	2.81

Effects for Four Categories of Tax Capacity (per pupil)

Variable	Mean for tax capacity >$2500	Mean for tax capacity $1000-2500	Mean for tax capacity $500-1000	Mean for tax capacity <$500	Common standard deviation
Self-rated knowledge	46.71	46.02	45.43	45.20	7.90
Validity of ASAP and essential skills	29.14	30.70	30.31	31.13	8.19
Opportunity to learn	29.63	27.23	27.49	26.27	6.80
Accountability/Authority culture	18.03	19.43	19.01	19.48	3.80
Beneficial effects of ASAP	22.55	24.41	24.18	25.02	6.07
Dis-equities of ASAP	28.25	29.20	29.33	29.07	4.69
Opportunities for teachers to learn	30.82	30.17	30.18	29.69	6.41
Rejection of ASAP ideals	21.75	22.23	22.17	23.14	4.81
Time for test administration	3.72	3.84	3.87	4.24	1.67
Time for test preparation	4.29	5.72	5.88	6.00	3.59
Training about ASAP	1.49	1.50	1.67	1.55	0.77
Professional development	3.21	3.21	3.35	3.00	2.82

APPENDIX G

EFFECT SIZE CALCULATIONS

Percentage of Sample Responding to Each Item Option. Item Means and Standard Deviations ($n = 1355$)

PART ONE: THIS SECTION CONTAINS STATEMENTS OF BELIEFS ABOUT TEACHING, LEARNING, AND ASSESSMENT. PLEASE INDICATE THE EXTENT OF YOUR AGREEMENT WITH EACH STATEMENT

	Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree	Item mean	Standard deviation	Missing data
B1. If the district curriculum is aligned with the Arizona Essential Skills, then every student will have equal opportunity to learn challenging and important material.	10%	44.8%	15.6%	22.5%	4.6%	3.34	1.08	34
B2. ASAP represents the best we know about how students learn.	2.3%	15.4%	24.2%	39.5%	17.9%	2.44	1.03	10
B3. Pupils at this school need to master basic skills before they can progress to higher order thinking and problem solving.	20.7%	42.5%	7.5%	22.4%	6.3%	3.49	1.22	9
B4. When ASAP is used as it was designed to be used, it can be integrated into daily instruction.	12.3%	50.3%	18.1%	15.4%	3.5%	3.53	1.01	7
B5. Using ASAP means that there are just more things that teachers at this school have to cover.	13.6%	28.2%	12.5%	38.7%	6.4%	3.04	1.22	
B6. Teachers at this school are now competent to teach the type of instruction compatible with the ASAP performance assessment.	15.1%	37.2%	24.1%	19.0%	4.2%	3.40	1.09	4
B7. Some pupils will never be able to profit from integrated, thematic instruction.	8.2%	23.9%	12.0%	37.4%	18.2%	2.66	1.25	7
B8. The primary role of teachers should be to teach the curriculum that the district defines.	9.8%	46.1%	10.4%	27.7%	5.1%	3.28	1.13	12

	Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree	Item mean	Standard deviation	Missing data
B9. When I get an integrated unit or test such as ASAP Form A, I have to break it down into its separate parts and teach the skills in each part so that students will be able to get the right answers.	10.0%	35.4%	21.5%	27.3%	4.4%	3.20	1.09	17
B10. Many students at this school have a very difficult time transferring and generalizing what they have been taught.	18.5%	45.8%	11.7%	21.5%	2.3%	3.57	1.09	3
B11. The Arizona Essential Skills represents high standards of achievement.	9.7%	40.4%	23.0%	20.1%	6.1%	3.28	1.08	9
B12. The Arizona Essential Skills is the curriculum framework that all Arizona teachers should follow.	6.1%	32.8%	26.5%	25.3%	8.8%	3.02	1.09	6
B13. A student should achieve a minimum mastery score on the ASAP to be graduated from high school.	10.5%	31.1%	22.1%	22.9%	13.1%	3.03	1.22	6

THE FOLLOWING STATEMENTS REPRESENT POSSIBLE EFFECTS OF THE ASAP PROGRAM. INDICATE YOUR AGREEMENT OR DISAGREEMENT WITH EACH STATEMENT (Strongly Agree = 5 . . . Strongly Disagree = 1)

	Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree	Item mean	Standard deviation	Missing data
E1. I don't understand ASAP well enough to adapt my teaching to it.	3.0%	19.6%	11.9%	50.2%	14.5%	2.46	1.06	12
E2. ASAP has had little or no effect on my teaching.	8.9%	42.4%	11.2%	30.0%	6.8%	3.17	1.16	10
E3. As a result of ASAP, major changes in curriculum have been made at school.	4.2%	26.6%	25.0%	35.7%	7.3%	2.85	1.04	16
E4. ASAP has given me moral support to do the kind of holistic teaching that I have been doing already.	4.5%	23.8%	23.6%	32.6%	14.4%	2.71	1.12	14

153

	Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree	Item mean	Standard deviation	Missing data
E5. ASAP takes away from instructional time we should be spending on something more important.	16.8%	26.0%	17.6%	32.0%	6.6%	3.15	1.23	16
E6. As a result of ASAP, many teachers at this school have changed the way they think about teaching and learning.	2.1%	18.1%	39.7%	29.4%	10.0%	2.73	0.94	11
E7. I feel like I am struggling to do BOTH a skills-oriented type of instruction AND a more holistic type of instruction.	16.6%	36.4%	12.8%	28.4%	4.9%	3.32	1.19	11
E8. The texts and materials that the district requires are compatible with ASAP.	3.6%	33.2%	21.8%	29.7%	10.3%	2.90	1.09	18
E9. Most or all of the tests this district requires have standardized formats (e.g., multiple choice, norm- or criterion-referenced)	7.7%	33.5%	24.6%	26.9%	5.9%	3.10	1.07	19
E10. Because of ASAP, I am learning how to teach with thematic units and integrated subject matter.	1.8%	21.9%	17.0%	43.1%	15.1%	2.52	1.05	14
E11. Because of ASAP, I am learning how to teach so that students will be able to solve complex, real-life problems.	2.8%	23.5%	16.9%	39.5%	16.1%	2.57	1.10	16
E12. Because of ASAP, our school has brought in consultants and experts in curriculum and instructional methods compatible with ASAP.	1.4%	16.9%	21.4%	36.5%	22.8%	2.37	1.06	13
E13. Many teachers are fed up with the outside demands and just hope ASAP will go away.	33.1%	30.9%	22.8%	10.1%	2.1%	3.84	1.07	13
E14. Our district has revised its Scope and Sequence (curriculum) to be aligned with ASAP.	6.2%	34.6%	31.7%	18.9%	7.3%	3.14	1.04	17

	Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree	Item mean	Standard deviation	Missing data
E15. Our district still uses the scores of standardized tests (ITBS, TAP, CRT) to compare schools and/or evaluate teachers.	10.0%	38.1%	29.2%	17.0%	4.4%	3.33	1.02	17
E16. Our district uses ASAP scores to compare schools and/or evaluate teachers.	6.4%	28.9%	40.7%	17.3%	5.6%	3.13	0.97	14
E17. Teachers feel pressure to raise ASAP scores.	22.1%	40.4%	21.3%	12.9%	2.4%	3.68	1.03	12
E18. Unless the state equalizes resources among schools, the ASAP program will not be able to reduce the differences in achievement between disadvantaged and nondisadvantaged students.	33.1%	35.0%	22.8%	6.1%	1.8%	3.93	0.99	16
E19. Overall, the benefits of ASAP testing are worth the investment of effort, time, and money it takes to administer the program.	4.5%	22.2%	26.3%	23.0%	22.5%	2.63	1.19	19
E20. Overall, the benefits of the entire state testing program (including ASAP, ITBS/TAP, and state-mandated district tests) are worth the costs.	2.1%	17.6%	28.3%	31.0%	19.1%	2.52	1.06	26
E21. ASAP is just another fad that will disappear.	17.0%	24.0%	36.1%	15.9%	5.0%	3.33	1.10	28

THIS SECTION CONTAINS STATEMENTS ABOUT THE VALIDITY OF ASAP PERFORMANCE TEST AND WHAT IT MEASURES. INDICATE THE EXTENT OF YOUR AGREEMENT OR DISAGREEMENT WITH EACH STATEMENT (Strongly Agree = 5; Strongly Disagree = 1)

	Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree	Item mean	Standard deviation	Missing data
V1. What ASAP measures is about the same as what ITBS/TAP measures.	0.9%	6.9%	23.0%	50.4%	18.1%	2.22	0.85	10
V2. ASAP scores accurately reflect a student's mastery of basic skills.	1.0%	18.5%	17.9%	41.7%	19.9%	2.38	1.04	14

155

	Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree	Item mean	Standard deviation	Missing data
V3. ASAP scores accurately reflect a student's ability to solve authentic, real-world math problems.	2.7%	28%	22.5%	30.3%	15.6%	2.71	1.12	12
V4. ASAP scores accurately reflect a student's ability to read for understanding.	3.5%	41.6%	16.6%	26.7%	10.7%	3.01	1.12	11
V5. ASAP scores accurately reflect a student's ability to write effectively.	5.3%	45.8%	15.0%	24.1%	8.9%	3.15	1.12	11
V6. ASAP scores accurately reflect the quality of education a student has received.	1.3%	12.8%	21.4%	40.4%	23.1%	2.28	1.00	12
V7. ASAP scores reflect things outside the school's control, such as intelligence, parental interest in education, or emotional adjustment.	10.8%	33.1%	23.2%	20.1%	12.2%	3.10	1.20	10
V8. Students have adequate motivation to perform on ASAP at a level that accurately reflects achievement.	1.0%	13.4%	21.6%	41.3%	21.9%	2.30	0.99	11
V9. No test, not even ASAP, is as valid as a teacher's judgment based on the student's performance on class assignments.	29.2%	39.3%	14.5%	14.8%	1.2%	3.81	1.06	12
V10. Students with poor reading skills will not be able to perform on the ASAP math portion at a level that accurately reflects their math achievement.	26.7%	44.2%	19.3%	7.6%	1.3%	3.88	0.93	12
V11. The differences between minority and nonminority pupils that exist on traditional standardized achievement tests also exist on the ASAP performance assessment.	20.7%	39.0%	29.8%	8.8%	1.0%	3.70	0.93	10
V12. ASAP scores, like other test scores, reflect in part the socioeconomic background of pupils and schools.	20.7%	46.6%	22.6%	8.0%	1.4%	3.78	0.91	10

	Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree	Item mean	Standard deviation	Missing data
V13. The ASAP is free of gender bias.	3.6%	24.6%	52.1%	13.0%	5.8%	3.07	0.87	12
V14. The ASAP is fair to all cultural groups.	9.8%	24.4%	47.8%	14.6%	1.8%	3.26	0.90	21
V15. The time limits on ASAP Form D unfairly penalize many students at this school.	7.2%	20.5%	50.4%	18.5%	1.4%	3.14	0.86	27
V16. Many children who are acquiring English as a second language will score poorly on the ASAP even if their educational achievement is adequate.	17.7%	47.2%	28.1%	5.5%	0.4%	3.77	0.82	15
V17. Confusing directions and procedures pose problems for the administration of ASAP.	32.1%	36.7%	18.7%	10.9%	0.4%	3.90	0.99	16
V18. Since ASAP is scored by a person rather than by a machine, its results cannot objectively or accurately measure achievement.	19.6%	30.2%	20.8%	25.6%	2.7%	3.39	1.15	16
V19. When most teachers at this school prepare their students to take ASAP, they encourage them to spend most of their time and energy on those activities that will earn high scores.	7.7%	22.4%	50.6%	16.0%	2.3%	3.17	0.87	15

THIS SECTION CONTAINS STATEMENTS THAT DESCRIBE CONDITIONS AT PUBLIC SCHOOLS. RATE EACH STATEMENT ACCORDING TO HOW ACCURATELY IT DESCRIBES YOUR OWN SCHOOL (Very accurate = 5 . . . Not at all accurate = 1)

	Very Accurate	Some-what Accurate		Not at all Accurate	Item mean	Standard deviation	Missing data
C1. Most changes of curriculum, instruction, and classroom organization are made by teachers on their own.	13.8%	29.4%	34.7%	13.7%	6.9%	3.30	1.09	21
C2. Most changes of curriculum, instruction, and classroom organization are made by the district officials and handed down to teachers.	8.2%	20.7%	34.2%	25.9%	9.4%	2.92	1.09	22
C3. Most changes in curriculum, instruction, and classroom organization are decided at the school by principal, department head, or teachers acting together.	9.6%	28.4%	33.6%	19.0%	7.5%	3.14	1.08	27
C4. A departmental organization (e.g., English, Math) makes curriculum integration difficult.	9.4%	15.3%	30.8%	23.3%	17.4%	2.75	1.21	51
C5. Teachers of different grade levels seldom plan together.	34.6%	24.8%	19.5%	12.5%	6.9%	3.69	1.26	24
C6. Teachers have ample time to experiment, collaborate and reflect on possible changes in their teaching.	3.8%	6.7%	14.9%	26.3%	46.3%	1.93	1.12	27
C7. There are so many classroom requirements that teachers have little room for discretionary activities.	19.4%	21.3%	33.3%	19.0%	5.2%	3.31	1.15	23
C8. There is so much pressure for high test scores that teachers have little time and energy to do other things.	14.6%	17.9%	31.4%	26.0%	8.3%	3.05	1.17	24
C9. The district and school have inadequate funds to purchase texts, curricular materials consistent with ASAP.	23.1%	18.9%	27.4%	18.3%	10.0%	3.28	1.29	32

Item	Very Accurate		Somewhat Accurate		Not at all Accurate	Item mean	Standard deviation	Missing data
C10. Some or all of the teachers are knowledgeable and experienced about holistic, integrated, problem-solving instruction.	9.8%	26.6%	39.3%	16.7%	6.3%	3.17	1.03	16
C11. None of the teachers has had much training or experience with performance (alternative) assessments.	13.6%	28.0%	34.5%	16.5%	5.1%	3.29	1.07	32
C12. There is at least one person at this school that teachers can turn to for accurate information about the ASAP program.	27.7%	29.6%	21.4%	10.8%	9.0%	3.57	1.26	21
C13. Most teachers see themselves as agents of change rather than as objects of change.	7.9%	25.8%	39.3%	17.3%	6.9%	3.11	1.02	38
C14. Adequate professional development has been provided for teachers to make changes necessary to implement ASAP.	4.1%	15.0%	29.9%	28.9%	20.7%	2.52	1.11	20
C15. The principal is supportive of integrated, holistic, problem-solving education.	26.7%	35.1%	24.8%	6.9%	4.3%	3.75	1.07	31
C16. Many teachers disagree with the philosophy of curriculum, instruction and assessment that ASAP represents.	17.9%	22.7%	35.8%	17.4%	4.4%	3.33	1.10	23
C17. There is an adequate level of resources to provide lifelike learning experiences for students.	3.2%	13.3%	30.7%	31.7%	19.6%	2.48	1.06	20
C18. Students at this school have adequate experience with alternative or performance assessment, so that the ASAP format is familiar to them.	3.2%	14.1%	32.8%	30.3%	17.8%	2.54	1.05	24
C19. Students have adequate opportunity to learn higher order thinking skills.	8.6%	32.4%	34.5%	18.9%	4.4%	3.22	1.00	17

159

	Very Accurate		Some-what Accurate	Not at all Accurate	Item mean	Standard deviation	Missing data
C20. Students have adequate opportunity to combine reading and writing tasks.	21.1%	46.0%	25.3%	1.0%	3.82	0.86	19
C21. Students have adequate opportunity to write about science and/or social studies topics.	13.8%	36.8%	32.2%	2.8%	3.48	0.98	31
C22. Students have adequate opportunity to write about their solutions to math problems.	7.8%	23.5%	37.1%	6.7%	3.04	1.03	38
C23. Students have adequate opportunity to engage in long-term projects and/or investigations.	10.3%	27.8%	34.8%	5.8%	3.18	1.05	24
C24. Students have adequate opportunity to pursue topics of their own choice.	8.5%	21.5%	34.9%	9.4%	2.96	1.09	22
C25. Students have adequate opportunity to engage in lifelike learning experiences.	7.3%	23.9%	38.1%	7.4%	3.02	1.03	21

TIME FOR TESTING

Consider ALL the testing you are required to do (including ITBS/TAP, ASAP, District CRTs). Estimate the number of hours you spend during a school year on administering these tests.

Option	Percentage
5 hours or fewer	8.9%
6-10	14.5%
11-15	16.1%
16-20	18.6%
21-25	11.3%
More than 25 hours	25.3%
Item mean	3.90
Standard deviation	1.68
n	1282

Estimate the number of hours you spend preparing students for these required tests during a school year.

Option	Percentage
10 hours or fewer	20.4%
11-15	7.3%
16-20	5.6%
21-25	5.5%
26-30	7 9%
31-35	4.8%
36-40	5.6%
41-45	3.2%
46-50	4.2%
More than 50 hours	26.3%
Item mean	5.59
Standard deviation	3.59
n	1231

Amount of training received on how to administer and score ASAP itself.

Option	Percentage
5 hours or fewer	59.6%
6-10 hours	20.1%
More than 10 hours	16.2%
Item mean	1.55
Standard deviation	0.77
n	1301

Amount of training and staff development and credit courses completed in teaching consistent with ASAP (over past two years).

Option	Percentage
5 hours or fewer	40.4%
6-10	15.6%
11-15	8.6%
16-20	5.8%
21-25	4.8%
26-30	3.0%
31-35	1.6%
36-40	3 0%
More than 40 hours	11.7%
Item mean	3.21
Standard deviation	2.82
n	1280

Self-Rated Knowledge: Percentages of Sample Responding to Each Level of Knowledge. Item-Level Descriptive Data ($n = 1355$)

Item	None	Awareness only	Good under-standing	Expert	Item mean	Standard deviation	Number missing
Arizona Essential Skills	3.8%	30.9%	57.2%	4.2%	2.64	0.63	53
ASAP Forms A, B, C	8.9%	39.6%	44.4%	3.5%	2.44	0.71	49
ASAP scoring rubrics	8.4%	27.1%	54.5%	6.8%	2.62	0.74	44
District DAP	15.2%	41.5%	34.6%	4.5%	2.30	0.79	56
Writing process	5.6%	22.0%	53.5%	15.8%	2.82	0.78	42
Literature-based reading	7.8%	20.8%	50.6%	17.2%	2.80	0.82	49
Conceptual math	16.7%	32.5%	39.3%	8.0%	2.40	0.87	48
Integrated, thematic units	3.7%	20.3%	57.0%	15.9%	2.88	0.72	41
Cooperative learning	1.1%	14.4%	65.8%	15.9%	3.00	0.59	43
Direct instruction of basic skills	5.5%	14.9%	58.1%	18.4%	2.93	0.75	43
Psycholinguistics	42.8%	36.3%	14.5%	2.1%	1.75	0.79	58
Constructivist learning theory	46.7%	33.7%	13.4%	1.8%	1.69	0.78	61
Student-generated research projects	7.8%	31.7%	47.1%	9.7%	2.61	0.78	49
Hands-on science	8.9%	24.0%	49.5%	14.2%	2.72	0.83	49
Probability and statistics	15.4%	39.0%	36.1%	6.1%	2.34	0.64	42
Problem-solving/critical thinking	2.4%	16.0%	63.6%	14.9%	2.94	0.64	42
Portfolio assessment	8.2%	32.9%	46.9%	8.6%	2.58	0.77	47
Alternative assessment	10.6%	35.6%	43.2%	7.2%	2.49	0.79	48

REFERENCES

Arizona Department of Education. (1992). *Arizona district profile*. Phoenix, AZ: Author.

Aschbacher, P. R. (1993). *Issues in innovative assessment for classroom practice: Barriers and facilitator* (CSE Tech. Rep. No. 359). Los Angeles: University of California, Center for Research on Evaluation, Standards, and Student Testing.

Ball, S. J. (1987). *The micropolitics of the school*. London: Methuen.

Behrens, J., & Smith, M. L. (1996). Data and data analysis. In D. Berliner & R. Calfee (Eds.), *Handbook of educational psychology* (pp. 945-989). New York: Simon & Schuster Macmillan

Blumer, H. (1969). *Symbolic interactionism: Perspective and method*. Berkeley: University of California Press.

Borko, H., Flory, M., & Cumbo, K. (1994*). Teachers' ideas and practices about assessment and instruction* (CSE Tech. Rep. 365). Los Angeles: University of California, Center for Research on Evaluation, Standards, and Student Testing.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56,* 81-108.

Cohen, D. K. (1995). What is the system in systemic reform? *Educational Researcher, 24*(9), 11-17.

Combs, A. W. (1991). *The schools we need: New assumptions for educational reform*. Landam, MD: University Press of America.

Corbett, H. D., & Wilson, B. L. (1991). *Testing, reform, and rebellion*. Norwood, NJ: Ablex Publishing Corporation.

Corcoran, T., & Goertz, M. (1995). Instructional capacity and high performance schools. *Educational Researcher, 24*(9), 27-31.

Cronbach, L. J., Bradburn, N. M., & Horvitz, D. G. (1994). *Sampling and statistical procedures used in the California Learning Assessment System*. Report of Select Committee to the Acting State Superintendent of Public Instruction, California State Department of Education, July 25.

Darling-Hammond, L. (1994). Performance-based assessment and educational equity. *Harvard Educational Review*, *64*, 5-30.

Elmore, R. F. (1995) Structural reform and educational practice. *Educational Researcher, 24*(9), 23-26.

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.

Firestone, W. A. (1989). Using reform: Conceptualizing district initiative. *Educational Evaluation and Policy Analysis, 11*, 151-164.

Flexer, R. J., & Gerstner, E. A. (1994). *Dilemmas and issue for teachers developing performance assessments in mathematics* (CSE Tech. Rep. No. 364). Los Angeles: University of California, Center for Research on Evaluation, Standards, and Student Testing.

Glaser, B. G. (1978). *Theoretical sensitivity*. Mill Valley, CA: The Sociological Press.

Goertz, M. E., Floden, R. E., & O'Day, J. (1995). *Studies of education reform: Systemic reform. Volume I*. Philadelphia, PA: Consortium for Policy Research in Education.

Hall, P. M. (1995). The consequences of qualitative analysis for sociological theory: Beyond the microlevel. *The Sociological Quarterly, 36*, 397-423.

Koretz, D., Stecher, B., Klein, S., & McCaffrey, D. (1994). The Vermont portfolio assessment program: Findings and implications. *Educational Measurement: Issues and Practice, 13*(3) 13-16.

Linn, R. L. (1994). Performance assessment: Policy promises and technical measurement standards. *Educational Researcher, 23*(9), 4-14.

Lipsky, M. (1980). *Street-level bureaucracy: Dilemmas of the individual in public services*. New York: Russell Sage Foundation.

Madaus, G. W., West, M. M., Harmon, M. C., Lomax, R. G., & Viator, K. A. (1992). *The influence of testing on teaching math and science in grades 4-12: Executive summary*. Chestnut Hill, MA: Boston College, Center for the Study of Testing, Evaluation, and Educational Policy.

McDonnell, L. M. (1994). Assessment policy as persuasion and regulation. *American Journal of Education, 102,* 394-420.

McDonnell, L. M., & Elmore, R. (1987). Getting the job done: Alternative policy instruments. *Educational Evaluation and Policy Analysis, 9*, 133-152.

Messick, S. (1994). The interplay of evidence and consequences in the validation of performance assessments. *Educational Researcher, 23*(2), 13-23.

Miles, M., & Huberman, A. M. (1994). *Qualitative data analysis*. Newbury Park, CA: Sage.

Moss, P. A. (1994). Can there be validity without reliability? *Educational Researcher, 23*(2), 5-12.

Noble, A. J. (1994). *Measurement-driven reform: The interplay of educational policy and practice*. Unpublished doctoral dissertation, Arizona State University, Tempe.

Noble, A. J., & Smith, M. L. (1994). Old and new beliefs about measurement-driven reform: "Build it and they will come." *Educational Policy, 8*, 111-136.

Nolen, 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* (Tech. Rep. 89-2). Phoenix: Arizona State University-West Campus, Education and Human Services.

Norusis, M. J. (1993). Chicago, IL: *SPSS for Windows*. SPSS, Inc.

Olivares-Seck, M. (1994). *Districts' capacity to respond to state-mandated reform*. Unpublished doctoral dissertation, Arizona State University, Tempe.

Parish, C. A. (1996). *Changes in practice accompanying the implementation of the Arizona Student Assessment Plan*. Unpublished doctoral dissertation, Arizona State University, Tempe.

Shepard, L. A. (1993). Evaluating test validity. In L. Darling-Hammond (Ed.), *Review of Research in Education* (Vol. 19, pp. 405-450). Washington, DC: American Educational Research Association.

Smith, M. L., Edelsky, C., Draper, K., Rottenberg, C., & Cherland, M. (1989). *The role of testing in elementary schools* (CSE Tech. Rep. No. 321). Los Angeles: University of California, Center for Research on Evaluation, Standards, and Student Testing.

Smith, M. L., Heinecke, W., & Noble, A. J. (in preparation). *Report on the Arizona Student Assessment Program (ASAP)*. Tempe: Arizona State University.

Smith, M. L., Noble, A. J., Cabay, M. Heinecke, W., Junker, S., & 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 Evaluation, Standards, and Student Testing.

Smith, M. L., Noble, A. J., Haag, S., Olivares-Seck, M., & Taylor, K. (1994). *Consequences of measurement-driven reform: Survey of Arizona school practitioners*. Los Angeles: University of California, Center for Research on Evaluation, Standards, and Student Testing.

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

Smith, M. S., & O'Day, J. (1991). Systemic school reform. In S. Fuhrman & B. Malen (Eds.), *The politics of curriculum and testing* (pp. 233-268). Philadelphia, PA: Falmer Press.

Stiggins, R. J. (1994). *Student-centered classroom assessment*. New York: Macmillan College Publishing Company.

Winfield, L. F., & Woodard, M. D. (1994). Assessment, equity and diversity in reforming America's schools. *Educational Policy, 8*(1), 3-27.