DOCUMENT RESUME

ED 342 799 TM 017 966

AUTHOR Aschbacher, Pamela; Herman, Joan
TITLE Alternative Assessments in Schools: Report on Status and Results of Local Projects.
INSTITUTION Center for Research on Evaluation, Standards, and Student Testing, Los Angeles, CA.
SPONS AGENCY Office of Educational Research and Improvement (ED), Washington, DC.
PUB DATE Nov 91
CONTRACT RI17G10027
NOTE 25p.
PUB TYPE Reports - Descriptive (141)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Academic Achievement; Case Studies; *Educational Assessment; *Educational Environment; Educational Policy; Elementary Secondary Education; Longitudinal Studies; Policy Formation; Program Evaluation; Program Implementation; School Districts; State Programs; *Student Evaluation; Test Use; *Theory Practice Relationship
IDENTIFIERS Alternative Assessment; *Alternatives to Standardized Testing; California; Performance Based Evaluation

ABSTRACT

To understand how assessment, particularly alternative assessment, and practice interactions may differ among schools with different implementation contexts and different school cultures, this project is studying over time a set of schools representing different state and local policy contexts and including different student populations. Research questions are: (1) whether the implementation strategies facilitate or inhibit teacher use of information from alternative assessments; (2) how alternative assessments interact with teachers' perceptions of achievement and student ability; (3) how alternative assessments interact with teachers' instructional decisions; (4) how alternative assessments interact with students' perceptions of performance and abilities; and (5) how each of the above differs in different schools. Longitudinal case-study methods are being used in two phases. In phase 1, four sites in California are being studied; phase 2 may expand to include other schools and states. The project will result in a series of products that should enhance the understanding of policymakers, researchers, and practitioners about ways to implement alternative assessments to improve learning and teaching. One table summarizes the study framework. There is a 35-item list of references. (SLD)

National Center for Research on
Evaluation, Standards, and Student Testing

Final Deliverable – November 1991

Project 3.1: Alternative Assessments in Schools

Report on Status and Results of Local Projects

Project Directors:
Pamela Aschbacher and Joan Herman

U.S. Department of Education
Office of Educational Research and Improvement
Grant No. R117G10027 CFDA Catalog No. 84.117G

Center for the Study of Evaluation
Graduate School of Education
University of California, Los Angeles
Los Angeles, CA 90024-1522
(310) 206-1532

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

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

ALTERNATIVE ASSESSMENTS IN SCHOOLS
REPORT ON STATUS AND RESULTS OF LOCAL PROJECTS*

Pamela Aschbacher and Joan Herman

We are currently witnessing an explosion of interest in new forms of assessment as tools for improving the quality of teaching and learning. This enthusiasm is fueled not only by widely heralded advances in assessment technology, but also by increased concern about the shortcomings of both traditional, teacher-based classroom assessments and large-scale testing programs. Researchers claim that the bulk of traditional classroom and large-scale assessments fail to measure significant, higher-order learning outcomes and thereby narrow the focus of schooling and undermine our educational system.

Performance-based assessment strategies have been highly touted, although perhaps prematurely, as a way to obtain a broader understanding of students' grasp of fundamental concepts of the disciplines and a deeper insight into students' thinking processes. Increasingly, researchers, practitioners, and policymakers believe these new techniques will not only monitor student learning better and more equitably, but also, and more important, improve the quality and fairness of teachers' own student assessments and of classroom practice generally.

Despite high hopes and lofty promises, we have little solid information about what is likely to happen when new assessment strategies actually enter the classroom. However, we do know that for improved assessment technology to have even a chance of improving practice, the development and implementation of new strategies must be sensitive to the realities of the schooling process and the complexities of educational change. Particularly important will be efforts to understand and accommodate the complex

* This report describes the progress made this year in Project 3.1, Alternative Assessments in Schools, and outlines our plans to continue the project next year as Project 3.3, Alternative Assessments in Classroom Practice.

interactions among new forms of assessment, the local and state assessment policy contexts in which schools are situated, the demographic and cultural characteristics of schools, and the ways teachers think about assessment and instruction. Without a sensitive tailoring of the introduction of new technologies to school conditions, the prospects for improving the quality of education with new assessment strategies are dim. Once again, after all our technical work, we are likely to find that the more things change, the more they stay the same.

Project Objectives

Given current assessment problems, growing enthusiasm for new strategies, and the complexity of school cultures and educational change, the objective of this project is to investigate two fundamental issues:

- Under what conditions are alternative assessment strategies most likely to become a valued part of classroom practice, that is, to be seen by teachers as providing useful information for evaluating student progress and shaping classroom instruction? and
- What are the consequences of new assessment strategies for central features of classroom teaching and learning? Under what circumstances do new forms of assessment alter teachers' and students' learning-related perceptions and behaviors?

Why These Objectives?

Current educational practice includes a troublesome mix of assessment practices. Classroom-based assessments of questionable quality serve as teachers' primary source of evaluative information. At the same time, externally-mandated standardized tests with technical sophistication but debatable validity and instructional consequences serve as policymakers' primary tool for monitoring student progress and school quality. While new assessment technologies promise both better information about student learning and an impetus for improved practice, the complexity of the culture of schools makes clear that a genuine integration of new assessment technologies into classroom practice will not be easy.

Current Problems with Classroom-Based Assessment

As a part of their daily practice, teachers observe and question children, taking note of their class participation, social competence, work habits, classroom behavior, abilities, and achievements (Shavelson & Stern, 1981). These informal assessments are spontaneous and unplanned, occurring quite naturally as a part of everyday classroom interactions between teachers and learners (Jackson, 1968). Teachers also judge the products that students produce during class assignments and homework, and they give quizzes and tests (Crooks, 1988).

Teachers generally view their assessment practices as a source of important and useful evaluation information (Dorr-Bremme & Herman, 1986). Researchers are less convinced. While few studies have examined the nature and effects of teachers' assessment practices (Crooks, 1988), some scholars have questioned the quality of teacher-based assessments and the validity and fairness of the decisions based on them (e.g., Natriello & Dornbush, 1984; Stiggins, Frisbie, & Griswold, 1989). Certainly, most teachers have had scant training in assessment concepts and practices (Stiggins, 1988).

Despite the dubious quality of classroom assessments, they have profound consequences in the lives of children and for the quality of schooling. Classroom-based assessments inform teachers' instructional decisions (Dorr-Bremme & Herman, 1986). In the worst cases, such assessments can mislead teachers into believing that students have learned when, in fact, they have not, or that remedial materials are needed when, in fact, more advanced materials would help to advance learning. Teachers use information from their own assessments as the basis for assigning grades and providing other feedback to children and their parents about how well children are doing in school. Classroom-based assessments are often the basis for decisions about students' ability group, classroom, and track placements that have consequences both for children's educational opportunities and their life chances (Oakes, 1985; Rosenbaum, 1976). The potential for harm is great if assessments support overly narrow, biased, or simply incorrect conclusions about what students have learned and what future learning opportunities are most appropriate for them.

What we don't know, however, is whether highly-touted new forms of assessment can actually be integrated into classroom practice and what their effects on practice will be. While hopes are high, we have little evidence to support expectations that teachers will find information from new methods of formal assessment more useful for judging student learning and guiding instruction, that students' will be encouraged to stretch toward further learning, and that parents will have more satisfactory answers to pressing questions about their children and their future prospects.

Current Problems with Externally-Based Assessments

Externally-developed standardized tests of achievement are also commonplace in classrooms. Until the 1970s, however, such tests were rather disconnected from ongoing classroom practice. And, despite continued technical refinement of these measures, most teachers have not seen standardized tests as useful in either shaping instruction or evaluating students (Dorr-Bremme & Herman, 1986) or as consequential (Goslin, 1967). Rather, teachers have viewed such tests as a necessary nuisance imposed by local administrators for their own purposes.

However, with the advent of minimum competency testing to certify students for graduation in the 1970s (Jaeger, 1989) and the current reform movement's emphasis on state and national testing programs as the primary tools for school accountability and improvement (Finn, 1991; Office of Educational Research and Improvement, 1988), students' scores have been increasingly employed to trigger other policy actions with serious consequences for both schools (ranging from public exposure to financial rewards or sanctions to state takeovers) and students (for example, promotion and graduation status). Thus, the types of tests formerly seen by many teachers and students as tests that don't "count" now count a great deal. At the same time, control over important assessment-related decisions—for example, what knowledge and skills to test, how often to test, which instruments to select or construct, how to use and report results—has been assumed by state legislatures and state education agencies, with recent, increasing momentum towards federal control of a national testing program.

These shifts have generated considerable concern and debate, not only about the validity and value of standardized tests, but also about the impact of these "high stakes" tests on school and classroom practice. While some have argued that such tests can "drive" practice toward reform and improvement (e.g., Popham, 1987), others counter with evidence that more negative consequences are likely. Rather than exerting a positive influence on students' learning, pressure to "do well" on externally-mandated tests can lead teachers to trivialize the learning and instruction process, distort the curriculum, mimic standardized test formats in their own evaluations, divert valuable instructional time to test preparation, and cheat (Bracey, 1989; Cannell, 1987; Darling-Hammond & Wise, 1985; Dorr-Bremme & Herman, 1986; Romberg, Zarinnia, & Williams, 1989; Smith, Edelsky, Draper, Rottenberg, & Cherland, 1989). Such changes in classroom practice appear to be motivated, for the most part, by political rather than pedagogical reasons (Madaus, 1987).

Moreover, more subtle classroom effects may also occur. Research that has compared teachers' judgments about students' abilities and achievements and standardized test results suggests a rather close match (see review by Hoge and Coladarci, 1989). One interpretation of these results is an optimistic one—that teachers are quite good judges of students' abilities. Another, more sobering conclusion is that teachers' conceptions of students' abilities and performance simply reflect the assessment measures with which they are familiar. That is, teachers may learn to define ability and achievement as the narrow range of knowledge and skills demanded by standardized tests, measures whose construct validity is open to question (Linn, 1986; Snow, 1980).

We don't know yet whether alternative assessments that are externally mandated by local or state policymakers will produce any different effects than traditional measures. For example, once mandated, will alternative assessments be trivialized and accommodated by teachers as traditional multiple-choice measures currently are? Or, will teachers find them more useful and more consistent with their own assessments? How likely is it that alternative assessments will encourage alterations of such fundamental dimensions of practice such as teachers' conceptions of academic ability?

Concerns About Alternative Assessment

Most of the rhetoric surrounding alternative assessment extols the anticipated virtues and consequences of new strategies that are "authentic," reflect higher-order thinking skills and metacognition, and are consistent with good instructional theory and practice. However, there are those who realize that implementing good alternative assessments on either a large or small scale may be quite complicated, and furthermore, that poorly designed, poorly implemented, or poorly utilized new assessments may well result in significant negative consequences (Hoover & Hill, 1991; Linn, Baker, & Dunbar, in press; Linn, Graue, & Sanders, 1990).

In a survey of state assessment officers, for example, Aschbacher (1990) found that although the push for alternative assessments at the state level has been strong and about half of the states are involved at some level in alternative assessments, many of the testing directors expressed serious concerns and reservations about the feasibility of mounting successful new assessment programs. Their concerns were focused in four areas:

- the high costs of development, training, administration, and scoring that are associated with the new measures;
- the need for new technical methods to assure reliability of scoring, and validity of tasks, scoring and utilization of results;
- difficult and complicated logistical arrangements; and
- a lack of solid understanding of and support for new assessment methods (and the accompanying curricular and instructional innovations they require), often despite apparent "acceptance" by the various constituencies involved in their implementation.

Clearly there is a broadly-felt need for research on the implementation and effects of alternative assessment techniques.

The Complexity of Practice and Change

The typical American school is subject to a multitude of expectations and caught between the often conflicting forces of idealism, diversity, and efficiency. How have schools responded to these expectations and conflicts? With an almost paradoxical culture—one that tries to respond to all demands while, at the same time, resisting fundamental change.

Sarason's (1971; 1982) seminal work on the largely unsuccessful attempts in the 1960s to improve mathematics curriculum points clearly to the conservative nature of the school culture. His analysis shows that traditional ways of organizing classes, curriculum, instruction, even assessment—what Sarason called cultural "regularities"—block change. When confronted with reforms, schools most often simply reconfigure them to fit their cultural status quo. Making matters more complicated, Sarason found that educators rarely question school and classroom "regularities" and often have a difficult time explaining and justifying them. Those who propose changes risk an uproar of protest and, perhaps even more powerful, silent and persistent resistance.

This cultural perspective has pushed researchers interested in innovation and change (e.g., Fullen, 1982; Sirotnik & Oakes, 1986) to consider the impact of both the formal regularities of schools and the less formal ones. What is clear is that efforts to bring innovations into schools nearly always fail unless implementation efforts consider such matters as how principals, teachers, and students think about their work, how they relate to one another, and what practices make sense and are important to them. While schools may respond superficially to externally-developed reforms and mandates, genuine alterations in school practice and outcomes are unlikely to take place.

For CRESST, research on school cultures and change suggests that if we hope to bring new forms of assessment into schools (either as a part of teachers' ongoing practice or as useful, externally-mandated programs), we need a much better understanding of the assessment "culture" within schools. We need to investigate what forms the schools' responses (and resistance) to new assessments are likely to take. And we need to identify the development and implementation strategies most likely to help teachers alter their conceptions and beliefs about assessment strategies and the information provided.

Project Activities

To understand how assessment and practice interactions may differ among schools with different implementation contexts and different school cultures, this project is studying over time a set of schools that represent different state and local policy contexts and include different student

populations. It is only by assisting in local and state efforts to make their general goals concrete and workable, and by evaluating the diverse effects of new assessment efforts, that we can effectively address these issues of national importance. This project will attempt to involve practitioners in the development of assessment tools and conditions that satisfy their information needs and further their instructional goals.

This project attempts to better understand the types of school conditions and support that are required to help schools and teachers improve their school- and classroom-based assessment, including incorporating high-quality assessment technology (traditional and alternative) into their own assessments. In the project, CRESST researchers and school practitioners will undertake collaborative efforts to understand the transfer of assessment technology to the classroom and the conditions under which innovative arrangements for school-based assessments can be developed.

Criteria beyond traditional psychometrics must also be used to evaluate the quality and impact of performance assessment. The diverse intended and unintended *consequences* of an alternative assessment program are a primary focus. Other criteria include *transparency, fairness, content quality, cognitive complexity, generalizability, meaningfulness,* and *cost* (Linn, Baker, & Dunbar, in press).

Conceptual Framework and Research Questions

The study's efforts to provide information about the implementation and effects of alternative assessments will be guided by five research questions:

1. What implementation strategies facilitate or inhibit teachers' use of information from alternative assessments in making their classroom-based judgments of students' progress and in shaping their curriculum and instructional decisions?

2. How do alternative assessments interact with teacher perceptions such as:

   - conceptions of achievement in a subject area;
   - views of students' capabilities.

3. How do alternative assessments interact with teacners' instructional decisions in such areas as:

- how they group students for instruction;
- what curricular content they select for particular students;
- what instructional tasks they have students tackle;
- how they distribute class time among various activities;
- what information they provide to parents about student progress.

4. How do .ternative assessments interact with students' perceptions of their own performance and abilities and with their learning-related dispositions such as confidence, willingness to work hard and to persist at difficult learning tasks?

5. How does each of the above differ in schools with different cultural characteristics (e.g., student race/ethnicity, socioeconomic status, and prior achievement levels)?

The conceptual framework in Table 1 suggests that implementation of new assessment strategies will be a function of what is tried and what already exists. Furthermore, we expect that successful implementation of new assessment strategies will be neither straightforward nor everywhere the same. Implementation will be influenced by who is involved in the process and their perceptions of the consequences. And, we do not expect that successful implementation strategies will necessarily be the same in schools serving different student groups. In fact we expect to find some interaction between schools' history of "success," or lack of it, with traditional educational practices including achievement measures (often strongly related to student-body composition) and the implementation of new measures.

Second, the framework in Table 1 suggests that, when successfully implemented, alternative assessments may influence teachers' most fundamental conceptions of what achievement means. For example, teachers may shift from a current emphasis on "right" answers and errors as the standard for achievement (mirroring the emphasis of paper-and-pencil tests) to a broader view—such as the extent to which students' work matches some exemplary or expert model of high levels of competence. New assessments may also change teachers' views of students' ability to learn various subjects and how ability is distributed among their students. For example, instead of seeing a wide spread of ability in the class that parallels the spread of correct answers on a multiple-choice test, teachers may come to view more students

as capable of performing complex tasks when assessed in less restricted circumstances. Such altered perceptions, if they occur, could be expected to affect teachers' decisions about classroom practice. For example, teachers may alter curriculum and instruction so that complex problems and student thinking dominate classroom tasks. Teachers may also change their standard conceptions of the need for ability grouping in reading or mathematics. Of particular concern here is how teachers' decisions shape the learning opportunities offered to students and whether teachers provide different opportunities to different students. Finally, we expect that alternative assessments, themselves and in combination with teachers' perceptions and teaching strategies, may affect how students relate to their classroom learning.

Table 1

Framework for Studying the Implementation and Effects of Alternative Assessments

Implementation – Under what conditions do alternative assessment strategies become a valued part of classroom practice? (What are the impediments and facilitators?)

Description of assessment strategies and context:

- student outcome goals: cognitive, conative, and/or affective skill domains assessed and for which students; breadth and depth of coverage, student response mode; content areas; grade levels; topics of focus; evidence of content validity (match of goals to measures)
- administrative support required (department, school, district, state), staff and time required
- costs: public relations, training, transportation of materials, substitutes, administration, reproduction, scoring, processing and interpretation, reporting
- administration conditions: physical setting, time and staff requirements, number of students involved, materials, storage and set-up, group or individual administration, sampling plan, other logistics
- scoring characteristics: nature of rating/scoring scheme, records needed, staff and training required, procedures to ensure reliability, turnaround time
- activities: student and teacher schedules and activities
- level of implementation

Table 1 (continued)

Conditions or features of implementation:

- student characteristics: demographics; components of appropriate instructional background for alternative assessment and grouping experiences; dispositions (e.g., confidence, willingness to work hard)
- teacher characteristics: cognitions of students, learning, teaching, content area expertise; dispositions, background; involvement in development of alternative assessment; perception of importance of their using new assessments and of the results
- administrator characteristics (including counselors): same as for teachers
- related staff development: key elements in training or preparation for teachers and administrators
- assessment purpose, relevant local and state policy or mandates, the "stakes" involved; involvement of state and/or local education agencies

Impact – What are the consequences of alternative assessment strategies for the central features of teaching and learning?

- students: cognitive, metacognitive, affective, and social outcomes, including attitudes about learning and academic self-confidence
- teachers: for instance, professional knowledge, attitudes and expectations about learning and instruction; professional role and power within the school structure; instructional practices and decision making (e.g., shift in focus from class activities to student outcomes, grouping of students for instruction, selection of content and tasks for particular students, distribution of class time among various activities, information provided to parents about student progress); and morale and self-esteem
- administration: for instance, support for teachers' use of alternative assessment and concomitant instructional decision making; information provided to parents, teachers and students
- parents: for instance, involvement in school, conceptions of learning and achievement, support for student learning
- curriculum, instruction, learning environment: equitable access to learning and assessment opportunities; efficacy of decisions made about students especially in comparison to those made on the basis of traditional assessments; "authenticity" and substance of outcomes; breadth and depth of curriculum; public scoring criteria; generative holistic approach to learning (including cognitive, metacognitive, affective skills)
- staff development: content (e.g., integration of instruction and assessment, new assessment methods), frequency, quality, audience
- policy: changes in department, school, district, or state policies related to assessment, curriculum, and instruction,(e.g., conceptions of desired student outcomes, access to learning opportunities)

Third, just as teachers may alter their conceptions of students' achievement and abilities, students may change the way they view themselves vis-à-vis classroom learning. For example, it's possible that students may develop more confidence and willingness to work hard at learning as a consequence of assessments that involve complex cognitive tasks that allow for multiple "right" solutions to problems and multiple ways of arriving at and expressing those solutions. Here, too, however, we expect that the interaction of new assessments and domains of practice may differ for different groups of students and in different schooling contexts.

We have incorporated in our framework a number of criteria for judging alternative assessment (Baker, 1990): consequences, fairness, transfer and generalizability, cognitive complexity, content quality, meaningfulness, and cost and efficiency. We expect that this framework could serve as a model for later research by CRESST and others.

Design and Procedures

Given the complexity of the questions listed above and the framework that underlies them, the study is using longitudinal, comparative case study methods (Yin, 1984) that will enable us to describe the implementation and effects of assessment in several contexts, explore in detail the specific dimensions of interest noted above, and develop explanations for the effects we observe.

The project consists of two phases—Phase 1: an initial planning and implementation phase during Years 1-2, in which we are working closely with teachers and administrators in the selected sites; and Phase 2: the impact phase during Years 3-5, in which we will study the impact of assessments in a number of different school settings. We expect that the development and implementation process may overlap with the impact phase in some sites.

Four sites (or cases) are included in the first phase of work. They include:

1. a large urban, racially and socioeconomically diverse school district where the district office sought our assistance in working with a small group of teachers to pilot alternative assessment strategies for classroom use;

2. a middle-sized suburban, less diverse school district where the district office sought our assistance in working with a team of teachers and administrators to develop, pilot, and implement alternative assessment strategies for both classroom and district use;

3. a culturally diverse elementary school not bound by state education policies and whose mandate includes participation in educational research projects, where the administration and faculty jointly decided to work with us on an assessment model for an eventual school-wide alternative assessment program; and

4. an innovative elementary classroom with children from primarily minority and lower SES backgrounds, in which the teacher has been developing and implementing alternative assessments for the past several years.

In the second phase of the study, during years 3-5, the number of cases may increase to include schools in other state and local policy contexts where successful implementation of alternative assessments has been achieved. Here, we may draw from schools in states as such Vermont or Kentucky with statewide efforts to develop alternative assessment programs, from schools involved in programmatic reform efforts such as the CHART Project in the humanities, and from schools involved in teacher professionalism reform efforts in the content areas, such as the Math + program. A varied set of cases will enable us to understand better how implementation, assessment, and practice interactions may differ among schools with different policy and implementation contexts and school cultures.

Data will be collected primarily through interviews, observations, and questionnaires. Among the instruments to be used will be measures of student achievement, student dispositions, and school and classroom context and processes. Analytical strategies will include the identification and verification of patterns that characterize each site and comparisons across sites to identify those interactions that appear to be generic and those that vary in different contexts.

Planning and Implementation Phase

The general approach of this project is to integrate research and practice in alternative assessment by working together with selected teachers, schools, and districts to help them develop and pilot alternative assessments, which we

can then study. Our assistance includes providing them with training in alternative assessment methods and development, providing some alternative assessment materials developed at CRESST, facilitating their development of new materials and methods for assessing performance, monitoring their implementation process, and documenting the impact of their innovative assessments.

During Year 1, two major tasks were completed and one initiated:

1. We developed preliminary plans for assisting the four case study sites with development and implementation of alternative assessments and worked collaboratively with school administrators and practitioners to develop plans unique to each site;

2. We trained groups of teachers and administrators from a number of schools in alternative assessment methods and concomitant instructional theory; and

3. We developed an initial framework for studying implementation and impact.

During spring of 1991, we contacted several districts, schools, and teachers about the possibility of collaborating with us to develop and improve alternative assessments in various subject matter areas and grade levels. We met with the most promising, including San Diego School District (SDSD) and La Cañada School District (LCSD) regarding high school social studies, UCLA's University Elementary School (UES) regarding elementary math or science, Conejo Valley School District and Fallbrook School District regarding math, and Charlotte Higuchi, an elementary teacher in the Los Angeles Unified School District (LAUSD) regarding elementary math. During the meetings we explored ways in which we might collaborate, such as our providing teacher training in alternative assessment and compatible instructional methods, studies of their implementation of new instruction and assessment methods, technical support for teachers who wish to develop new assessments, and some evaluation of the impact of such innovations. We selected two content areas in which to focus our efforts: math and social studies.

Alternative assessment in secondary social studies. As the foundation for future work in both this project and Project 2.2 (Alternative Approaches to Measuring School Subjects), CRESST staff conducted a week-long workshop

during the summer of 1991 for 11 secondary social studies teachers from three districts: SDSD, LCSD, and LAUSD. The workshop was held at UCLA under the auspices of CRESST's Project 2.2. The training consisted of an introduction to the value and use of alternative assessment and the CRESST model for assessing content area understanding (in social studies and science) by examining student writing (Baker, Aschbacher, Niemi, Yamaguchi, & Ni, 1991). In addition, training was provided in instructional methods based on a generative model of student learning (Wittrock, 1990) as an appropriate instructional precursor to performance assessment. This approach is based on cognitive research showing that students must elaborate and question new information and link it to prior knowledge in order to be able to use it generatively to interpret new situations and solve problems.

Teachers in the CRESST summer training institute worked in teams to develop instructional plans based on cognitive learning theory to accompany CRESST's model social studies assessments. Their instructional plans together with the CRESST assessment materials are being used by teachers this fall and winter. Additional workshops this winter will be held to debrief teachers and to provide further input and reinforcement of what they learned last summer.

The teacher training workshop laid the foundation for teachers in the field to adapt and expand the CRESST assessment model in their secondary social studies classes during FY 92. We anticipate that, for example, they may expand the model to include assessment of extended research and/or extended writing aspects of the task, or they may create new assessment tasks in new history topic areas or in geography. As they develop and use alternative assessments, we will study their implementation and impact. The materials they develop will become part of a sampler of alternative assessments to share with other practitioners.

Alternative assessment in math. There are three sites of study for math assessment: a school, an inner-city classroom, and small district. Each is described below.

School-wide reform. Several meetings were held with UES faculty and administrators to discuss plans to collaborate with CRESST on the development of new alternative assessments. The school faculty is eager to

participate in educational research and to create models for educational reform. Our early plan was to help them develop portfolios and journals in two content areas, probably science and math, as a model alternative assessment strategy that could be adapted to other content areas in the future. After much discussion, however, we narrowed the focus to math alone and decided to work solely on journals during FY 92 for several reasons. A number of the teachers already use journals as part of their class activities but do not use them yet for assessment purposes. Working on only one new method in one subject area would minimize logistical and other implementation barriers and allow us to focus more squarely on assessment. Also, we all agreed that it is best to start small and expand later. The entire school is emphasizing math this year, so that is a natural choice of subject area for our project, and the faculty is intrigued by the idea of using journals in math. Furthermore, journals may be viewed as a sort of subset or special case of the portfolio concept. Focusing on journals this year will allow us to deal with many portfolio-related issues on a small scale now, thus segueing nicely into a proposed math portfolio project at UES the following year to be funded by the National Science Foundation. At additional meetings this fall, CRESST staff joined UES faculty and administrators in examining current practice with regard to journals, defining "journal" for this project, specifying which student outcomes or characteristics they want to be able to assess in this way, and outlining appropriate journal content to ensure such measurements. They are quite interested in metacognitive and affective outcomes as well as cognitive ones.

Inner-city classroom. Charlotte Higuchi, a teacher-researcher at a culturally diverse elementary school in the LAUSD, met with us several times to discuss her philosophy and methods of assessment and instruction and to share research ideas with us for the coming year. She uses a variety of personally-created performance assessments throughout the year with her students, including math journals. This winter CRESST researchers will visit her classroom again and meet with her to share our different approaches to content specifications and scoring rubrics for math journals. Higuchi's class is a particularly rich and unusual setting in which to study the implementation and impact of alternative assessment. She is a very dedicated and innovative teacher who has been developing and refining alternative methods on her own over the past several years. She has had most of the same

students in a multi-age class for three years, thus eliminating some of the difficulties of assessment that arise from lack of background knowledge about individual children. In addition, her students are primarily from lower SES, minority backgrounds, and a number of them need special education.

District-wide reform. To provide the basis for future work in this project, CRESST staff conducted a three-day workshop during the summer of 1991 for representatives from seven districts. The training covered an introduction to the value and appropriate use of alternative assessment and a systematic model for developing performance assessments. Participants worked in groups to develop an alternative measure to use in their own district classrooms. Two follow-up workshops in December of this year and next spring will help participants refine and expand their measures.

Several districts that attended the June workshop on developing performance assessments are very interested in developing district-wide alternative assessments in math. Conejo Valley, for example, has a less diverse student population than Los Angeles, but has significant experience in direct writing assessment and is already familiar with the development of rubrics, training and scoring procedures, and the integration of assessment with good instruction. In fact, they already have formed a 10-person committee of administrators and teachers to explore and develop new assessments.

Other alternative assessment initiatives. In addition to working directly with teachers, schools, and districts on alternative assessments, we are participating in at least two other initiatives for developing alternative assessments that could be studied by this project in subsequent years. One of these is a group of over 20 states convened this fall by the Council of Chief State School Officers (CCSSO), with support from CRESST and OERI, to collaborate in the development of alternative assessments in a variety of subject areas including math and social studies. CRESST staff are facilitating the networking of the Social Studies Consortium, which may provide a site for study in future. Currently involved states include: Alabama, California, Connecticut, Hawaii, Iowa, Nebraska, New Jersey, North Dakota, Texas, and Washington. At the fall meeting, the consortium developed a method of summarizing state curriculum frameworks and assessment methods, which

members will apply to their state's program and share among the group. At a meeting this winter or spring, participating states will share what exists, identify overlapping content, and map out potential areas for future joint development projects in social studies assessment.

The second initiative in which we are participating is a group of practitioners, content area experts, and assessment experts who will be brought together for the first time in January by the Rockefeller Foundation's CHART Project to explore new ways of assessing student outcomes in the humanities. This programmatic approach may provide new venues or new assessment models for future research in this project.

Year 2 Plans

During Year 2, two major tasks are to be completed and one initiated:

1. Complete the conceptual framework for understanding the likely facilitators and barriers to the implementation of alternative assessments in different school contexts. The result of this activity, the November 1992 deliverable, will be a paper synthesizing barriers and facilitators to the implementation of alternative assessments.

2. Plan and collect baseline data at several sites on selected areas of impact, such as dimensions of student achievement, teachers' perceptions and practices, and students' dispositions.

3. Collect a sampler of alternative assessment strategies tried at each site.

We will continue to work with teachers, schools, and districts in Year 2 to document the tryout and revision of new assessment methods in brief case studies at the sites described above. In so doing, we will refine and apply our framework for evaluating the implementation and impact of alternative assessments. We expect to monitor the development of new measures during winter and spring 1992, and, with spring 1992 field tests, we plan to collect implementation and impact data. We also will assemble a collection of the alternative assessments created in math and social studies during this project to share with other practitioners and researchers. We anticipate that this document will include the actual assessment materials used, together with the relevant specifications for learner goals, assessment materials, administration procedures, training required, and scoring criteria and standards. The document will likely contain information about journals and

open-ended items in math and about essays and short-answer questions in history and geography. Our work with the CCSSO state consortia and the CHART social studies assessment group may allow us to enrich the document with additional assessment pieces.

The project will result in a series of products that should enhance policymakers', researchers', and practitioners' better understanding of the complex process of implementing alternative assessments in ways that can improve teaching and learning. It will also improve our knowledge of how various forms of assessments interact with fundamental classroom perceptions and practices that shape students' opportunities to learn. Additionally, the project will provide the opportunity to pilot new measures of student learning developed in other CRESST projects and to share site-developed measures with other practitioners and researchers.

References

Aschbacher, P.R. (1990). *Alternative assessment: State activity, interest, and concerns* (CSE Tech. Rep. No. 322). Los Angeles: UCLA, Center for Research on Evaluation, Standards, and Student Testing.

Baker, E.L. (1990). *Responses to Negotiation Questions, Institutional Grant Proposal for OERI Center on Assessment, Evaluation and Testing.* Los Angeles: UCLA, Center for the Study of Evaluation.

Baker, E.L., Aschbacher, P.R., Niemi, D., Yamaguchi, E., & Ni, Y. (1991). *Cognitively sensitive assessments of student writing in content areas.* Los Angeles: UCLA, Center for Research on Evaluation, Standards, and Student Testing.

Bracey, G.W. (1989). *The $150 million redundancy.* Phi Delta Kappan, 70(9), 698-702.

Cannell, J.J. (1987). *Nationally normed elementary achievement testing in America's public schools: How all 50 states are above the national average* (2nd ed.). Daniels, WV: Friends of Education.

Crooks, T.J. (1988). The impact of classroom evaluation practices on students. *Review of Educational Research, 58,* 438-481.

Darling-Hammond, L., & Wise, A.E. (1985). Beyond standardization: State standards and school improvement. *The Elementary School Journal, 85*(3), 315-336.

Dorr-Bremme, D., & Herman, J.L. (1986). *Assessing student achievement: A profile of classroom practices* (CSE Monograph 11). Los Angeles: UCLA, Center for the Study of Evaluation.

Finn, C. (1991). The case for national testing. *GAO Journal,* 13, 30-42.

Fullen, M. (1982). *The meaning of educational change.* New York: Teachers College Press.

Goslin, D.A. (1967). *Teachers and testing.* New York: Russell Sage.

Hoge, R.D., & Coladarci, T. (1989). Teacher-based judgments of academic achievement: A review of literature. *Review of Educational Research, 59*(3), 297-313.

Hoover, H.D., & Hill, R. (1991, June). *Authenticity, validity, and efficiency: When must we use authentic measures? When needn't we? When are we best and least well served by conventional measures?* Paper presented at the annual Alternative Assessment Conference of the Education

Commission of the States and the Colorado Department of Education, Breckenridge, CO.

Jackson, P.W. (1968). *Life in classrooms.* New York: Holt, Rinehart & Winston.

Jaeger, R.M. (1989). Certification of student competence. In R.L. Linn (Ed.), *Educational measurement* (3rd ed.) (pp. 485-514). New York: Macmillan.

Linn, R.L. (1986). Educational testing and assessment: Research needs and policy issues. *American Psychologist, 41,* 1153-1160.

Linn, R.L., Baker, E.L., & Dunbar, S.B. (in press). Complex performance-based assessment: Expectations and validation criteria. *Educational Researcher.*

Linn, R.L., Graue, M.E., & Sanders, N.M. (1990). *Comparing state and district test results to national norms: Interpretations of scoring "above the national average"* (CSE Tech. Rep. No. 308). Los Angeles: UCLA, Center for the Study of Evaluation, Standards, and Student Testing.

Madaus, G. (1987). *Testing and the curriculum.* Chestnut Hill, MA: Boston College.

Natriello, G., & Dornbush, S.M. (1984). *Teacher evaluative standards and student effort.* New York: Longman.

Oakes, J. (1985). *Keeping track: How schools structure inequality.* New Haven, CT: Yale University Press.

Office of Educational Research and Improvement. (1988). *Youth indicators 1988: Trends in the well-being of American youth.* Washington, DC: U.S. Department of Education.

Popham, W.J. (1987). The merits of measurement-driven instruction. *Phi Delta Kappan, 68,* 679-682.

Romberg, T.A., Zarinnia, E.A., & Williams, S.R. (1989). *The influence of mandated testing on mathematics instruction: Grade 8 teachers' perceptions.* Madison, WI: University of Wisconsin, National Center for Research in Mathematical Science Education.

Rosenbaum, J.E. (1976). *Making inequality: The hidden curriculum of high school tracking.* New York: Wiley.

Sarason, S. (1971). *The culture of the school and the problem of change.* Boston: Allyn & Bacon.

Sarason, S. (1982). *The culture of the school and the problem of change* (2nd ed.). Boston: Allyn & Bacon.

Shavelson, R.J., & Stern, P. (1981). Research on teachers' pedagogical thoughts, judgments, decisions, and behavior. *Review of Educational Research, 41*(4), 455-498.

Sirotnik, K., & Oakes, J. (1986). *Critical perspectives on the organization and improvement of schooling.* Boston: Kluwer Nijhoff.

Smith, M.L., Edelsky, C., Draper, K., Rottenberg, C., & Cherland, M. (1989). *The role of testing in elementary schools* (Final Report). Los Angeles: UCLA, Center for Research on Evaluation, Standards, and Student Testing.

Snow, R.E. (1980). Aptitude and achievement. In W.B. Schrader (Ed.), *Measuring achievement: Progress over a decade* (pp. 39-59). San Francisco: Jossey-Bass.

Stiggins, R.J. (1988). Make sure your teachers understand student assessment. *The Executive Educator, 10*(8), 24-30.

Stiggins, R.J., Frisbie, D.A., & Griswold, P.A. (1989). Inside high school grading practices: Building a research agenda. *Educational Measurement: Issues and Practice, 8*(2), 5-14.

Wittrock, M. (1990). Generative processes of comprehension. *Educational Psychologist, 24*(4), 345-376.

Yin, R.K. (1984). *Case study research: Design and methods.* Beverly Hills, CA: Sage Publications.