LINKING TESTING WITH INSTRUCTIONAL DECISION MAKING: SOME MODELS AND GUIDELINES FROM RESEARCH

Donald W. Dorr-Bremme

CSE Report No. 209
1983

CENTER FOR THE STUDY OF EVALUATION
Graduate School of Education
University of California, Los Angeles

The work reported herein was supported in part under a grant from the National Institute of Education. However, the opinions expressed do not necessarily reflect the position or policy of the National Institute of Education and no official endorsement should be inferred.

TABLE OF CONTENTS

Introduction

The nation's investment in school achievement testing is enormous; the amount and variety of testing have continued to grow. Unfortunately, however, much of this testing remains under-utilized. Achievement-test results can be extremely useful in school-wide and district-wide planning and decision making. They can provide information for analyzing curriculum, diagnosing instructional strengths and weaknesses, and determining directions for school improvement. Test-score patterns can also serve as the basis for projecting student achievement goals and monitoring progress toward their attainment. But, as four years of research at UCLA's Center for the Study of Evaluation shows, it is only the rare district in which schools regularly and systematically use testing for these or similar purposes. In most school districts across the country, testing and instructional decision making are not routinely and coherently linked. As a result, few schools and districts are getting a maximum return on the dollars, staff time, and student learning time that they invest in testing. At the same time, the quality of the educational planning in their schools suffers from a lack of precise and useful information.

What distinguishes the relatively small number of schools in which principals and teachers regularly pay close attention to test scores? How do some districts structure on-going links between testing and instruction in their schools? How can others do so, achieving both fuller use of testing and more informed instructional decision making in the process? This article addresses each of these questions, drawing answers from the Center for the Study of Evaluation's recent, extensive research.

After a brief review of the research base, the discussion below opens with a look at current test-use patterns in the nation's schools. These patterns verify that much achievement testing is indeed under-utilized, and they suggest that there is little use of test information in systematic, school-wide planning. The data also indicate that leadership by district administrators is a key factor in how closely testing is linked to instructional decision making and reveal two general approaches that district leaders can follow in building such links. Attention then turns to some specific ways that each of these general approaches can be pursued by district leaders. Models and concrete guidelines are presented, all based upon intensive case studies of districts that have systems in place for linking assessment with instructional planning and decision making. A summary highlights the main features of these systems.

The Research

From 1979 through the present, the Center for the Study of Evaluation (CSE) at UCLA has been conducting two, complementary studies of testing and test use. One, the Test Use in Schools Project, has studied current nation-wide patterns. Inquiry has focused on many types of tests and other assessment techniques, especially in the basic skills areas of reading/English and mathematics. The project's central effort was a 1981 survey of assessment practices, test uses, and contextual factors that influence them. Questionnaires were sent to the principal and four classroom teachers in elementary schools and high schools randomly chosen from within a nationally representative sample of school districts. Returns were received from 1,058 respondents: 220 principals, 475 upper-elementary-

grade teachers, and 363 high school teachers of English and mathematics. Fieldwork took place before and after this survey; a total of eleven schools (four secondary, seven elementary) from five districts in different geographic regions of the country were visited. The two phases of field study produced over 100 in-depth interviews with school- and district-level educators, including 12 principals, 69 classroom teachers, various instructional specialists, and other administrative personnel.

While the Test Use in Schools Project has focused on what is, the second CSE effort, the Evaluation Design Project, has yielded a detailed look at what can be. It has conducted fieldwork in districts which currently have testing-instruction linkage systems in place. In order to locate these districts, recommendations were widely sought and between 40 and 50 of the districts nominated were screened in phone interviews with appropriate district-level administrators. Ultimately, eight districts were selected for intensive exploration in on-site interviews and observations. A first phase of the Evaluation Design district organization study (1979-1982) was directed at understanding how the testing-instruction linkages in these districts worked and at identifying the district characteristics that contributed to their creation and maintenance. More recent work has examined the costs and impacts of two of these eight systems, each of which represents one general approach to tying test results with instructional decision making.

While the Test Use Project establishes the need for linking testing with instruction and indicates some general approaches for doing so, study of these eight districts by the Evaluation Systems Project provides some clear guidelines for effective testing-instruction linkage systems.

The Status Quo: Two Tiers of Testing and Limited Use

The Test Use in Schools Project provides a portrait of the contemporary status quo. It describes central tendencies in public-school test use across the nation. This portrait reveals that there are two tiers or layers of student-achievement assessments in our schools today. These are consistently distinguishable from one another in their proprietorship, characteristics, and functions. One tier of assessment is internal or local to the schools. It is "owned," and for the most part produced, by teachers themselves. This local or internal tier includes two main types of assessment: (1) the tests, quizzes, and other measures that teachers construct and administer in the course of their teaching, and (2) the clinical judgments of students' achievement that teachers form as they interact with students and observe their work in various classroom situations day after day. A third kind of measure also figures in this tier, but it is especially imporant for elementary-school teachers. These are the tests included with commercial curriculum materials used in the classroom. While these are not produced in the school, teachers in the elementary grades most often have an investment in them. Teachers often have a say in choosing (and choosing how much to use) them and the materials they accompany; teachers can time their administration and adapt their content to fit the pace and emphases of instruction.

The second tier of assessment is external to the school: mandated by the district, state, and/or suggested by federal program requirements (e.g., for placement in compensatory education programs). Norm-referenced, standardized test batteries are the most common among these. Other types

of measures used for minimum competency (or functional literacy) testing or as part of state assessment programs are also included here. In some cases, too, tests constructed or purchased by districts and referenced to their curricular objectives fall in this second category. Tests of these kinds are also included there. In some cases, too, tests constructed or purchased by districts and referenced to their curricular objectives fall in this second category. Tests of these kinds are developed beyond the schools. Their administration is called for primarily to meet organizational needs and concerns at higher levels of public-education governance. Those who work at those levels may have a sense of ownership in these tests; educators in the schools rarely do.

These two tiers of assessment function quite differently in most schools and districts. Teachers and principals rely heavily on the results of internal assessment strategies and consider them important as they go about routine instructional planning and decision making. At the same time, they generally treat information from external testing as of minor importance, using it only occasionally and idiosyncratically. These patterns are obvious in both CSE's fieldwork findings and survey data.

Forty-three classroom teachers were interviewed during pre-survey fieldwork in a systematic-but-open-ended format. They discussed all the information they had throughout the year on students' academic capabilities, performance, and progress; they described whether and how they used that information. Collectively, they cited far more uses for the information that came from assessment strategies that were local to the school and classroom. (See Table 1.)

Teachers surveyed across the nation were asked to rate the importance of diverse types of assessment results in four routine, decision-making

tasks. Again, the pre-eminence of the internal tier of assessment was apparent. (See Table 2.) Principals in CSE's national survey were asked to rate how important a role data from various sources played in eight regular school-level administrative activities. Here, the separate functions of the two tiers of achievement assessment were especially apparent. Principals reported counting internal assessment data more heavily in making instructionally relevant decisions, e.g., allocating funds, assigning students, evaluating teachers. But they indicated that results of external measures were more important in reporting to those beyond the school, e.g., to district administrators and the public.

Taken together, the research findings just cited show that there are notable quantitative differences in the ways the external and internal tiers of assessment are used by educators in the schools. They reveal that the results of externally mandated testing serve fewer purposes (Table 1) and are not counted as heavily in planning or decision making (Tables 2 and 3). But fieldwork clearly suggests that there are also significant qualitative differences in how the two tiers of assessment are typically utilized by teachers and principals. The results of external tests are most often examined briefly, casually, and asystematically. Do principals consider the results of standardized and district-objectives-based tests in curriculum evaluation? Table 3 suggests that they do. But interviews indicate that this often means that they merely glance over the scores, mention them in a faculty meeting, and point out the areas in which the school did especially well or poorly. Do teachers use standardized test results in planning? Apparently they do to some extent (Tables 1 and 2).

Table 1

Types of Tests and the Uses of Their Results by Teachers (Interview Data)

(Cells show the number of times the 44 interviewed teachers freely cited each use for each type of test)

USES	TEST TYPES									
	A	B	C	D	E	F	G	H	I	Total
Planning Instruction	24	21	10	3	2	3	13	4	2	82
Referral/Placement	3	6	0	2	0	0	11	1	0	23
Within Classroom Grouping & Individual Placement	6	14	18	4	6	5	4	3	1	61
Holding Students Accountable for Work, Discipline	8	2	3	0	0	0	0	0	0	13
Assigning Grades	32	8	17	5	1	1	1	1	0	66
Monitoring Students' Progress	18	12	17	2	1	1	0	1	0	51
Counseling & Guiding Students	10	6	0	0	1	2	3	0	0	22
Informing Parents	0	1	0	0	0	1	0	0	0	2
Reporting to District Officials, School Board, etc.	0	3	1	0	0	2	0	0	0	6
Comparing Groups of Students, Schools, etc.	0	1	0	0	0	1	1	0	0	3
Certifying Minimum Competency	0	0	0	0	0	0	0	1	0	1
TOTAL USE CITATIONS	101	74	63	16	11	19	33	10	3	330
Explicit Statements of Non-use	0	1	0	1	0	0	10	2	7	21

Legend:

A = Teacher Constructed
B = Teachers' Other Major Assignments
C = Curriculum Embedded
D = School/Department/Grade Level
E = Commercial Diagnostic
F = District-Objectives Based
G = Standardized
H = Minimum Competency
I = Statewide Assessment

Table 2

Importance of Test Results for **Teacher** Decision-Making in Elementary and Secondary Schools (Survey Data)

[mean ratings on 4-point scale: 4 = Crucial; 3 = Important; 2 = Somewhat Important; 1 = Unimportant or not used]

			ELEMENTARY		
Decision Area:	A	B	C	D	E
Planning teaching at beginning of the school year	----	3.39	----	2.53	2.60
Initial grouping or Placement of students	3.12	3.58	2.91	2.51	2.59
Changing a student from one group or curriculum to another, providing remedial or accelerated work	3.12	3.66	3.04	2.52	2.52
Deciding on report card grades	3.38	3.69	2.89	1.62	1.81
			SECONDARY		
Planning teaching at the beginning of the school year	----	3.59	----	2.22	2.38
Initial grouping or placement of students	3.04	3.84	2.48	2.28	2.46
Changing students from one group or curriculum to another, providing remedial or accelerated work	3.27	3.61	2.67	2.52	2.59
Deciding on report card grades	3.65	3.68	2.29	1.36	1.45

Legend:
A = Teacher-Made Tests
B = Teacher Observations/Opinions
C = Tests Included with Curriculum
D = Standardized Test Batteries
E = District Continuum or Minimum Competency Tests

Table 3

Importance of Test Results for **Principal or Schoolwide** Decision-Making in Elementary and Secondary Schools (Survey Data)

[mean ratings on 4-point scale: 4 = Crucial; 3 = Important; 2 = Somewhat Important; 1 = Unimportant or not used]

	ELEMENTARY					
Decision Area:	A	B	C	D	E	F
Curriculum Evaluation	2.94 (.84)	3.27 (.64)	3.01 (.67)	2.91 (.75)	3.04 (.87)	2.99
Student Class Assignments	2.93 (.79)	3.12 (.71)	2.50 (.81)	2.35 (.91)	2.46 (.99)	2.44
Teacher Evaluation	2.12 (.97)	----	1.70 (.76)	1.53 (.78)	1.80 (.93)	1.68
Allocating Funds	----	3.08 (.71)	1.91 (.87)	1.89 (.90)	1.94 (1.01)	1.91
Student Promotion	3.05 (.70)	3.29 (.67)	2.65 (.81)	2.31 (.96)	2.38 (.94)	2.45
Public Communication	2.31 (1.05)	----	2.77 (.90)	2.47 (.99)	2.34 (1.00)	2.52
Communicating to Parents	3.43 (.55)	3.45 (.57)	2.91 (.60)	2.64 (.98)	2.67 (.95)	2.74
Reporting to District	2.62 (.91)	----	3.12 (.68)	2.78 (1.10)	2.74 (1.10)	2.88
	SECONDARY					
Curriculum Evaluation	2.76 (.75)	3.14 (.70)	2.83 (.67)	3.27 (.64)	2.95 (.82)	3.02
Student Class Assignments	2.98 (.73)	2.99 (.79)	2.77 (.77)	2.98 (.87)	2.78 (.87)	2.84
Teacher Evaluation	2.39 (.83)	----	1.63 (.74)	1.77 (.71)	1.84 (.78)	1.75
Allocating Funds	----	3.34 (.54)	1.73 (.81)	2.20 (1.13)	2.06 (1.08)	2.00
Student Promotion	3.33 (.85)	3.46 (.75)	1.61 (.78)	2.58 (1.28)	2.05 (1.13)	2.08
Public Communication	2.24 (1.05)	----	2.84 (.80)	2.92 (1.03)	2.30 (1.07)	2.69
Communicating to Parents	3.56 (.55)	3.38 (.76)	2.91 (.58)	3.03 (1.00)	2.55 (.99)	2.83
Reporting to District	2.53 (.88)	----	3.10 (.64)	3.12 (.97)	2.92 (.95)	3.04

Legend:

A = Results of Teacher and Curriculum Tests
B = Teacher Opinions/Recommendations
C = Standardized, Norm-referenced Test Batteries
D = Minimum Competency Tests
E = District Objective-based or Continuum Tests
F = Average External Tests (C,D,F)

Fieldwork suggests, however, that more often than not, this means a once-a-year visit to the office for quick look at their students' cumulative files. Are standardized test batteries and minimum competency scores consulted in student placement? Each table indicates that they are. But visits to schools make clear that they are most often consulted as part of an automatic or cursory gate-keeping procedure. Law or policy guidelines direct that students with scores below a certain cut-off point be placed in a compensatory program or remedial class. Alternatively, as one high-school teacher put it, describing a procedure reported by many offices:

> They give me each kid's standardized-test score on my class roster. If one stands out, I usually check with the counselor to be sure the kid should really be assigned to geometry.

Such uses contrast sharply with teachers' recurrent and systematic use of assessment that are local to the classroom and school in an on-going process of intructional planning and decision making. They contrast markedly with principals' serious consideration of teachers' advice, recommendations, and grades on teachers' assignments in making budgetary decisions or next year's class assignments. And they certainly do not constitute thorough utilization of external testing data in a systematic process of school-wide analysis of curriculum and instruction, decision-making, and planning.

Why do the two tiers of achievement assessment function in the different ways that they commonly do? The reasons are not hard to find. They lie in the interplay of several factors: characteristics of the measures themselves, circumstances surrounding their availability, educators' training in assessment, and the organization of educational planning in schools,

districts, and beyond. Examining these factors not only discloses why internal and external assessments are used differently in the schools. It also begins to reveal how each tier of assessment can be used more fully toward improving instruction -- and why district leadership is the key to this process.

Our system of schooling is organized such that teachers routinely do a great deal of instructional planning. They have a major role in planning what to teach (and/or emphasize) and how to teach it, in diagnosing individual students' learning needs, and in assuring that students are working at appropriate levels in the curriculum. As the school year unfolds, they need to monitor their students' progress, to consider whether and how to adjust the pace and emphases of their teaching, to grade students and inform parents of achievement-to-date, and so on. To do all this and do it well, teachers need assessment tools with three basic characteristics: (1) Validity -- they must assess what the teacher believes he or she has actually taught in a way that seems consonant with the way he or she has taught it; (2) Suitability -- their intended purposes must fit the tasks the teacher needs to accomplish, (thus teachers seek placement tests for placement, chapter and unit tests for monitoring progress and grading, etc.); and (3) Immediate Availability -- the teacher must be able to employ them whenever it seems appropriate to do so and have the results back promptly. In short, the assessment tools that teachers need must be sensitive to local conditions, to the array of particular circumstances in their particular classrooms at the moment. And, in order to function throughout the year as the instructional leaders of their schools, principals need

measures of the same kind. It is not surprising, then, that both teachers and principals rely heavily on assessment strategies that are internal to the school and its classrooms; teacher-made tests and assignments, teachers' observations and clinical judgments, and the adaptable, readily available tests that come with the commercial curriculum materials they are using. From their points of view, these internal measures have all three of the characteristics listed above. Externally mandated measures, on the other hand, usually do not. They are not designed primarily to provide data for routine classroom decision making. The fit between their contents and format and a particular teacher's curriculum is problematic. Often, their scores are not returned until weeks or months after administration. Often too, the results come back in a format teachers and many principals find unfamiliar and/or cumbersome. For any or all these reasons, the results of standardized tests, minimum-competency measures, and many district-objectives-based tests can seem remote and irrelevant to teachers and principals. In addition, teachers and principals generally have limited formal training in testing and measurement or the use of test data (Herman & Dorr-Bremme, 1983; Yeh, Herman, & Rudner, 1978). This limits also the accessibility of external testing data to educators in the schools. CSE's Test Use Project fieldwork found teacher and principals voicing these very concerns as drawbacks of external testing (Burry, et al., 1982; Dorr-Bremme, 1983).

But the very characteristics that make internal assessment tools ideal for use in individual teachers' and principals' routine work severely restrict their utility for systematic school- and district-wide planning.

Their content and the timing of their administration is idiosyncratic, variable from classroom to classroom. Aggregating the data they provide in order to see achievement patterns across grade levels, a department, or the entire school, therefore, is difficult if not inappropriate and impossible. This is especially true of teacher-made tests and assignments, but it also often applies to tests embedded in texts and other commercial materials. (Teachers time their administration differently; they sometimes adapt their contents. The same materials or text series are not always used throughout the school.) And while teachers' cumulative observations and experience-based judgments are valuable sources of information, they cannot be readily synthesized into a precise, detailed, picture of specific curricular or teaching strengths and weaknesses across many classrooms or schools.

It is these problems with local or internal assessment strategies that have made standardized, minimum-competency, and special district-objectives-based tests attractive to local school districts -- and make similar measures a virtual necessity for states and other educational agencies. By providing standard and consistent data across settings, such tests facilitate comparisons among classrooms, schools, and/or districts; they permit year-to-year monitoring of performance. They are likely to be more sound psychometrically than teachers' own tests; in most circumstances they are sufficiently valid to indicate broad patterns and trends. Tests of these kinds can take time to administer, score, and analyze comprehensively, but comprehensiveness is important to district and state planning, especially if data are gathered only annually or biannually. Coming full circle,

however, the same features that make these types of measures useful to districts and larger education agencies generally limit their usefulness for teachers and principals. Thus, two tiers of achievement testing, largely distinct in their functions, are maintained in public schooling.

Both of these tiers could be used more fully than they now are as information sources in school-wide decision making and planning for instructional improvement. How? The data and analysis presented above suggest two approaches that districts can follow: One approach is to build from the inside out: to construct district tests that have the characteristics of internal assessment tools -- the validity for local curricula, suitability for routine classroom purposes, and immediate availablity that appeal to teachers -- and at the same time provide consistent, reliable data that can be aggregated in ways useful for school and district decision making. The second approach is to build from the outside in: to analyze information from externally mandated measures currently given in the district and deliver it to schools at times and in formats that maximize its utility in planning for curricular and instructional improvement.

These approaches are not mutually exclusive; both can be followed simultaneously. But the effectiveness of either depends upon more than the proper handling of testing and test scores. It also depends upon a district system that structures and supports the use of testing information in an on-going planning process -- systems of a type that are not widely present in most districts today. On the whole, most districts do not routinely return test results to schools in ways that facilitate their use in decision making. Admnistrators review scores for the faculty in most

schools, but rarely on a periodic basis as part of routine procedures. Follow-up to assure that teachers are giving attention to the content area, skills, etc., that test scores indicate need emphasis is rarely routine either. (See Table 4.) Survey data show that the majority of teachers are instructed in how to administer tests and that they are informed about test results. Yet it appears that few receive training in how to link teaching and testing or in how to use test results in improving instruction. (See Table 5.)

These are only some very general indicators that not many districts are closing the testing-instruction loop with systematic planning mechanisms. They are supported, however, by fieldwork from both CSE projects. Furthermore, even though efforts of the kinds shown in Tables 4 and 5 are only the most elemental in a district testing-instructional decision making linkage system, they can make a difference in how teachers view and use testing. Analyses of survey data show that where there is more support by district and school leaders for the use of test results in planning, and where there is more staff development in assessment, teachers have a significantly more positive view of testing and its uses and tend to treat the results of district-objectives-based, standardized, and even minimum-competency tests as more important in instructional decision making. With this in mind, let's examine some ways that districts can create successful links between testing and planning for instructional improvement in their schools.

Building Links From the Inside Out

Districts that follow this approach build outward from classroom assessment needs to those of the school and districts. They also build

Table 4

Making and Holding Teachers Accountable for Test-score-Based Curricular Decisions (Survey Data)

	Principals' Reports*		Teachers' Reports*	
	Elementary	Secondary	Elementary	Secondary
SCHOOL ADMINISTRATOR(S) . . .				
Meets with teachers to review scores and identifies areas that need extra emphasis	3.09	2.94	2.84	2.05
Observes teachers, reviews their plans to ensure areas indicated by tests are being emphasized	3.23	3.07	2.66	2.31
Takes test scores into account in evaluating teachers and/or establishes test-score goals for teachers to meet	1.57	1.55	1.46	1.27
DISTRICT ADMINISTRATOR(S) . . .				
Returns test results such that they can be used in school's curricular decision making	2.63	2.03	Not Asked	
Observes, reviews school plans and/or requires reports to assure school is emphasizing skills that test scores show need work	2.84	2.67	"	
Establishes specific test-score goals for school	2.12	2.33	"	

*Mean ratings on four-point scale: 4 = happens regularly, routinely; 3 = not regular or routine but happens fairly often; 2 = not regular or routine and happens rarely; 1 = does not happen at all.

Table 5

Percentages of Teachers Reporting Recent Participation in Staff Development

	Topic	Elementary	Secondary English	Secondary Math
(1)	How to administer tests required by my state, district, and/or school (procedures to follow, etc.)	78	54	46
(2)	Analysis and explanation of state, district, or school test results	84	70	60
(3)	How to tie what is taught more closely to the skills, content covered on required tests	50	37	25
(4)	Training in the use of test results to improve instruction	35	21	19

from what should be taught to what should be tested. First they construct district curricula, then district tests to match.

Two of the districts studied closely by CSE's projects were especially successful in taking this approach. Their slightly different testing-instruction linkage systems are useful models for others.

The Central City Model *

Located in the rural midwest, Central City School District serves about 5,000 students in seven elementary schools, three junior highs, and a high school. It has a long history of innovation and commitment to curriculum development. It also has a group of teachers who pioneered use of the high school's main-frame computers (originally purchased and used for computer-assisted instruction) in the scoring and analysis of teacher-made tests. These factors, and an energetic leader, joined in the creation of Central City's system for linking test information with instructional planning.

The test information. Each summer in recent years, the district has sponsored curriculum development projects. But while the district initiated, compensated, and guided, it was teachers who did the work. Several representatives from the faculties of each school were selected by their peers to participate.

*District names used in this paper are pseudonyms. Any resemblance between these names and those of actual districts and communities is unintended.

Efforts began with the construction of an elementary-grade media (or library) skills module and continued through the development of complete mathematics and social science curricula for the elementary grades. Later, the mathematics curriculum was extended through grade 8 and work began on a reading program. In each case, development was done unit by unit in several stages. First, teachers decided on instructional objectives and selected and/or wrote materials and learning activities for achieving them. Then, pre-and post-tests referenced to the objectives of each unit were designed and "mastery levels" for each objective were specified. Units and accompanying tests were piloted the next year; objectives, materials, and test items were revised in light of teachers' criticisms and suggestions. Further revisions incorporating teachers' feedback were made after the units went into general use in schools across the district.

Testing materials were designed such that all the unit tests could be scored and analyzed by computer and returned to the teachers in a day or two. Results came in the form of a set of easy-to-read sheets, one for each student. The sheet listed each objective covered on the test, the number of items that measured the particular objective, the number of these items the student had correct and incorrect, and whether the number correct equaled "mastery." At the top of each sheet appeared a paragraph that described the types of errors the student had made and summarized the types of difficulties the student seemed to be having with the skills or content covered.

In mathematics, the district had selected a sample of items from the unit tests and combined these to create mid-year and end-of-the-year

summary measures given to students in all schools. Teachers received summary sheets of the type described above for these tests, too. (The district was considering developing similar tests in other subject areas once the process of curriculum and test-item revision was considered complete.)

All this applies to the lower grades, but similar developments had begun in the high school mathematics department. These were initiated by the teachers, who had worked toward common curricula and devising computer-scored tests for various courses. In line with a general district attitude, other departments were encouraged, but not required, to follow this example.

The end results of the district-wide effort were several: (1) curricula that were consistent across the district, that teachers were invested in, and which teachers actually used; (2) a system of tests that fit the curricula and provided timely information in a form appropriate for a variety of routine instructional decisions; and (3) a body of test information that was valid and consistent from classroom to classroom and could thus be aggregated and compared in school and district planning.

The structure of school decision making. Within the schools, these test data came into play in two main ways. First, they were routinely used by teams of teachers in regular "unit" meetings. Elementary-school "units" included several teachers (one of whom was chosen as unit leader) a cluster of students across two or three grades, and occasionally an instructional aide. Students were often divided among unit teachers in different groupings for different subjects based on their current level of achievement and rate of learning. (Some schools, however, tended to use the self-contained classroom approach for most students).

Unit teams met at least weekly during release time at the end of an abbreviated school day. At the beginning of the year, they discussed students' placement and planned instructional emphases and pacing. Later on, they routinely examined students' progress, reviewed their placements, re-evaluated and altered their teaching, and discussed individual learner's problems and how best to address them. Data from district tests, as well as other available information, were routinely examined as these matters were considered. Unit meetings, then, were the primary setting for linking test data with instructional decision making. (Where classrooms were self-contained, teachers reported using the district tests individually, as well as in unit meetings. And similar procedures were followed in the junior high and high school math departments.)

A second use of district test data occurred periodically as principals established school goals and agendas for school in-service activities.

District support systems. The linkage effort described above was supported by the Central School District in a number of ways.

First, district leaders initiated and provided resources for the curriculum-and-test development. They also gave release time for weekly unit meetings in which test data are used for instructional planning.

Second, district administrative leaders provided staff development in curriculum writing and test development. Originally, these semester-long, weekly "courses" were led by professors from a state university. Later, however, the district encouraged teachers to take over the classes: to adapt them to be more practical and relevant and to serve as instructors. Credit on the district's pay scale was given for participation in these

classes. A district administrator also maintained close contact with the nearby office of the local Intermediate Educational Agency (IEA). IEA help was routinely sought in problems in test development and scoring-and-analysis issues. The IEA also provided some staff development in instruction.

Third, the district maintained media centers staffed by instructional specialists in each school. Specialists helped unit teams and individual teachers locate supplementary teaching materials to address learners' needs. They also offered training in such areas as instructional diagnosis and prescription.

Fourth, a district administrator worked with teacher committees in piloting curriculum units and tests, eliciting teachers' critiques, and revising objectives, materials, and test items.

It was this same administrator who encouraged continuing and broadening the use of the computer-scoring-and-test-analysis process.

The Shelter Grove Model

The Shelter Grove Unified School District is located in the southwestern region of the country. Until three years ago, Shelter Grove was an elementary school district. The recent merger has brought Shelter Grove's enrollment to about 5,700. These students are distributed through four elementary schools, two middle schools (grades 6-8), and a four-year high school.

Shelter Grove's system for linking testing with instruction is similar to Central City's in several ways. Yet it is different enough to be worth description as a second "inside out" model.

The test information. Like Central City, Shelter Grove administers tests of several types. But those that have the greatest power to influence instruction in Shelter Grove schools are those developed by the district and referenced to its continua (or sequences) of instructional objectives in reading, mathematics, and writing (composition).

Shelter Grove initially contracted with a commercial firm which promised to write test items for district-selected objectives and to provide computer printouts of scores. Introduced in the early 1970's, these tests failed to win teacher support. Teachers complained that they were not coordinated with anything that was taught, that they did not know what to do with the results.

Teacher committees were appointed to try to revise test items. They responded to the need for coordinating testing and curriculum by beginning to work on a district-level continuum of objectives. From then on Shelter Grove's experience paralleled the more recent history of Central City. By the late 1970's, teacher committees had devised continua of objectives and accompanying criterion-referenced tests for reading and math, as well as similar tests for language arts. More recently, a district writing continuum was established.

Unlike the Central City materials, Shelter Grove's tests do not serve as unit pre-tests or post-tests. And except in written composition, district objectives are not accompanied by district-designed materials or recommended learning activities. Rather, the continua are aligned with commercial reading and math text series used district-wide.

The district tests at the appropriate level were routinely administered to students by classroom teachers at two or three points between October and February. Scores were aggregated by the district's Testing Coordinator for individual students, instructional groups, entire classes, and the school. These profiles were sent to the schools in time for planning days that occur regularly at several points through the year.

In addition, proficiency tests composed of various segments of the district's criterion-referenced tests were administered to children in grades 4, 5 and 6 each year in April and May in accordance with state requirements.

The structure of school decision making. District tests were routinely used in each elementary and middle school during planning days that occur at several points in the school year. (The system has yet to be introduced in the district's high school.) Two of these days were in June. On the first, the program of the school was routinely evaluated by the entire school staff looking at the group, classroom, and total school scores. These sessions functioned as a needs assessment for the next school year. On the second June planning day, individual teachers placed students in appropriate learning groups for the coming year using the test-result profiles on each student.

In September of each year, test information was updated; information on students new to the district was added. In October, teachers met with their principals to set learning goals -- benchmarks on the continuum that, based upon past performance profiles, they expected the children in each instructional group to meet.

A mid-year evaluation took place each February. Summary reports on current-year testing were run, distributed, and examined. Principals met with teachers, as well as with the Superintendent and Assistant Superintendent for Instruction, to discuss students' progress. Plans for modifying the instructional program were made at this time. Then, in June, the cycle began anew with reference to the again-updated test-score profiles.

Individual teachers also used criterion-referenced test information in reporting to parents each October and again each spring. Report cards listed continuum skills on one side and noted students' progress toward each objective. And each May, letters were sent to the parents of children who were two grade levels behind expected performance; special conferences with these parents were also arranged.

District support systems. As was the case in Central City, a number of district activities and programs helped to sustain the linking of test data with instructional planning in Shelter Grove. In addition to the district's leadership and resources in developing the instructional-objectives continuua and criterion-referenced tests, these included the following.

The district maintained a Professional Development Program (PDP) that provided teachers with the skills necessary to act upon the test results. Coordinated by a full-time specialist, the PDP had evolved over time based upon the Madeline Hunter orientation to teaching. Level One activities (for all new teachers, aides, and substitutes) dealt with such basic teaching skills as understanding goals and objectives, motivation and reinforcement, and task analysis and diagnosis. Level Two activities (which were not required but encouraged, and which many teachers took) extended those

of Level One with emphasis on individualizing instruction. Strategies for meeting affective needs using inquiry skills and teaching specific curriculum content were also covered. (Prior to the general implementation of this PDP program, all principals had been required to take the Level One course plus courses in clinical teacher supervision.)

PDP sessions require teachers to apply the skills taught back in their own classroom, with supervision and feedback from the PDP coordinator.

Learning specialists conducted demonstration lessons, recommended materials, conducted diagnoses of new students, and assisted teachers in planning and placement when new criterion-referenced test scores arrived in the schools. The learning specialists were considered master teachers, and regularly played an important role in helping teachers use test information. They also functioned to communicate changes in the continuum or changes in district policy to the faculty. With the PDP, learning specialists were perceived as critical supports to the district's linkage effort.

A Testing Advisory Committee composed of a principal and several teachers continually updated and improved the district's tests in light of teacher criticisms. This group also handled whatever administrative and technical problems arose in testing, scoring, and reporting back results.

Ad hoc continuum revision committees made up of teachers and learning specialists were paid during the summer to revise sections of the continua as seemed appropriate.

In addition to these formal organizational features, a variety of other networking activities (e.g., principal observations, learning specialists' visits to classrooms, monthly meetings of a district communications

council) helped district personnel work closely together in maintaining links between test data and instructional planning in the Shelter Grove schools.

Guidelines

The experiences of Central City and Shelter Grove, especially in contrast to those of two other districts with similar but less successful linkage systems, suggest a number of guidelines for other districts interested in linking testing with instruction from the inside out.

1. Build curriculum and assessment measures together "in-house."

Administrators and teaching staff in both districts believed very strongly in the district development process. They felt that it helped assure teacher "ownership" and confidence in both curricula and tests; these, in turn, seemed important as prerequisites to teacher use. Shelter Grove's unhappy experience with tests built outside the district, even when they were coordinated to district specifications, supports this wisdom.

2. Assure a close fit between test items and curricular objectives and materials.

This can best be done by designing curriculum first and then the tests, as was done in Central City and, ultimately, in Shelter Grove too.

Teachers are inclined to see district objectives-based or criterion-referenced tests as a burdensome irrelevancy if this condition is not met. New Branford, an urban district with 30,000 enrollment in the northeastern United States, attempted to devise criterion-referenced tests keyed to its district reading and math objectives. But when Test Use Project researchers visited New Branford schools, they found that few teachers used

them. Continuum objectives were intended to fit with the five or six math and reading series used across the district. In fact, according to teachers, they fit well with none. Thus, teachers continued to use the tests included with these commercial series to get the information on achievement they needed -- and they also had to give district tests to comply with district requirements. But information from the latter was rarely consulted, and teachers resented the requirement. For similar reasons, Central City teachers neglected their district's objectives-based reading tests, although they were generally enthusiastic about those in the other subjects. Developed years earlier with little teacher participation, and no accompanying curriculum materials, teachers complained that the reading tests were not valid for the content of the two basal series used in Central City.

3. Strive for maximum teacher involvement.

To help build curriculum and tests that teachers own and use, teachers' participation in the development process must be more than nominal. Both Shelter Grove and Central City included many teachers on their development committees; these teachers did the real work of constructing the curricula (or continua) and the test items. Mechanisms were provided that allowed all district teachers to offer feedback on a regular basis. Their criticisms were taken seriously in the revision process.

In contrast, New Branford (mentioned just above) and Metro District (another urban district studied by the CSE Test Use Project) had a small number of teachers on district advisory committees as they constructed continua of objectives and accompanying tests. These teachers did not

participate in the actual development process; their presence was not visible to district faculty; they had little impact on the results. And in neither district did teachers feel the objectives or tests were completely suitable. New Branford teachers' response has been described. Teachers' response to Metro District's tests was quite mixed.

4. Make tests that cover the entire range of skills in the curriculum and/or continuum of objectives.

The district tests of Central City and Shelter Grove included items that assessed students' performance on skills and content from the most elemental to the most advanced in the subject areas tested. Metro District (enrollment over 100,000), in contrast, purchased tests for each grade level in reading, math, and language arts that covered only the most simple skills to be taught in the grade. In the economically disadvantaged neighborhoods where more students had trouble with these skills, test results did help teachers locate areas in which individuals and class groups needed remediation. But in these schools, the tests also functioned to push the actual curriculum in the direction of the most elemental skills. Teachers and principals wanted students (and their schools) to do well on the tests each spring. Thus, they spent much time drilling and re-drilling children on the simple skills tested. Simultaneously, they gave shorter shrift in their teaching to other skills specified for the grade level, which were included on the test. Elsewhere in the district, where students routinely obtained 90 percent to 100 percent correct on the same tests, the tests yielded little diagnostic or placement information for teachers.

One moral of these contrasting stories, then, is test what you want teachers to teach, because teachers will place their teaching emphasis on what you test.

Several other "do's" and "don'ts" can be abstracted from the Central City, Shelter Grove, and similar but less successful models. These, however, are equally pertinent to the "outside in" linkage approach discussed next. Thus, they will be omitted here and mentioned in the concluding summary.

Building Links From the Outside In

Districts that follow this approach adapt information from externally mandated tests to suit schools' planning needs. In so doing, they support school-level planning structures and procedures, just as districts taking the inside out path do.

The testing-instruction linkage systems of two districts that followed the outside in approach are described below. They provide very different, but equally instructive models.

The St. John Model

The St. John School District covers a wide geographic area of suburban and semi-rural municipalities in a Western state. Its 72 schools serve between 40 and 50 thousand students in grades K-12.

Linking testing with instructional planning began in St. John during the mid-1970's when the state legislature enacted a program intended to stimulate planning for school improvement at local school sites. Participation in the program was voluntary, but over the years most St. John elementary schools, along with two of its junior high schools and one high school, elected to participate. The district encouraged this involvement;

in turn, schools' participation occasioned district efforts to provide test data for use in local site planning.

The test information. Long before the advent of the state-sponsored school improvement program, St. John School District had required administration of the Iowa Test of Basic Skills. Students were tested each January in grades 2-6. The purposes this information had served previously are not germane here. But once numerous St. John schools joined the state program, test data became especially important for them. Guidelines for the state school-improvement planning process required that in establishing improvement plans schools specify: (1) the "existing level of performance" in a particular area, (2) the "needed program changes or additions," (3) improvement objectives, and (4) activities to measure these objectives. Major activities to be undertaken in order to achieve the objectives also had to be described, along with budgets and other improvement-program features. But the four requirements enumerated here were those that called for "hard data" such as test results.

It seemed reasonable to use ITBS results in developing these improvement plans, yet district administrators realized that these came back from the test publisher in a form that was cumbersome. Computer printouts presented the results for each sub-test area for each grade for each year on a separate page. Principals and teachers found these reports complicated as well as overwhelming in size. Consequently, the district undertook development of what it now calls the Academic Performance Profile (APP).

In summary, the APP gave each district elementary school an annual overview of its ITBS test results for all years and all grades for a particular subtest (e.g., reading comprehension, math concepts, etc.) on a single page. This reduces fifty pages of computer printout to approximately six ordinary $8^1/_2$ by 11 inch pages.

In addition, the APP simplified the format in which the information appeared. Simple graphs were devised to visually display : (1) the scores of student groups as they moved through the grades (1982 first graders as second graders in 1983, etc.); (2) the performance at various grade levels in various years (the fourth grade in 1981, 1982, 1983, etc.); and (3) the gains (indicated in terms of grade-level growth) realized from one year to the next for the various grade levels (the gains made from second grade in 1982 to third grade in 1983). Two simple tables on each page (i.e., for each sub-test) supplemented the three line graphs.

Since the state program guidelines also call for annual needs assessment, the St. John District also created survey questionnaires for staff, parents, and students. These solicited respondents' perceptions of the effectiveness of the schools' various programs, as well as their perceptions of how strong the need is to improve in the same areas. Each school can add up to 20 questions to those common across the district. Surveys are administered annually in the spring of each year. The district's evaluation office tabulates those for each school and returns their results in a concise form.

The structure of school decision making. The state's school improvement program mandated the creation of a School Planning Council (SPC) in

each participating school. SPC members had to include the principal and representatives of the teachers, other school staff, parents and other community members, and (at the secondary level) students. This group was assigned central responsibility for establishing needs, goals, and activities for school improvement, as well as allocating the state program funds provided to the school for improvement activities.

However, St. John's district evaluation specialists elaborated on these state requirements. They urged their schools to also create "component committees," smaller groups (including SPC members and others) who were charged with planning for improvement in particular areas -- in subject areas, in school environment, in human relations, in staff development, etc.

Component committees review the ITBS/APP summary forms, survey results, and other information. They specify and document needs, set objectives, and develop school and classroom activities to realize them. They also state how achievement of the objectives will be evaluated and propose a budget suitable for their plan. In a next step, various component committees present their particular plans to the School Planning Council. The SPC accepts or suggests changes in each improvement-plan component and makes decisions regarding final allocation of state program dollars among the various components. The SPC also monitors implementation of the plan through the coming school year.

While plans were routinely developed for a three-year period, revisions were made each spring based on information gathered during the current school year. Thus, school improvement planning was an annual process

centered in the spring. But implementation in classrooms and SPC monitoring recurred throughout each school year.

Interviews with participants and observation of planning meetings indicated that test data (and survey results) were used in deciding upon and substantiating needs, specifying objectives, evaluating implementation, and revising the plans. SPC members also referred to this information in making and justifying budgetary decisions.

District support systems. The St. John School District supported its testing-instruction linkage system in many of the same ways that Shelter Grove and Central City supported their quite different ones.

Staff development in the organization and process of planning, including the use of the APP test summaries, was conducted for 600 district personnel during their first year in the state program. Others received this introductory training as they entered the program. Furthermore, teachers, principals, and parents agreed that the regular availability of the districts' two evaluation specialists was a key to the program's maintenance. They routinely provided staff development and answered ad hoc questions regarding planning and test-data use.

St. John also maintained a comprehensive staff-development program in instructional techniques. This was a major factor in facilitating the realization of school plans.

The Bayview Model

Bayview is a community of 100,000, and is located about 50 miles from a major Western metropolitan area. The Bayview Unified School District's

sixteen elementary schools, four junior highs, and three senior highs enroll 14,000 students.

Bayview's six-year-old effort at testing-instructional linkage was more diffuse than that in most of the other school districts visited by CSE researchers. Interest in testing and evaluation was relatively new, and many in the district were skeptical of their value. Nonetheless, the need to comply with externally mandated testing programs stimulated a small group of district administrators to try to make greater local use of them. Only one of these uses will be discussed here. It offers an example of "outside in" testing-instruction linkage that is quite different from the St. John School District's.

The test information. Three different achievement testing programs figured in the Bayview linkage endeavor to be described here. The first of these was the State Assessment Program (SAP). This half-hour test was administered each spring to students in grades 3, 6, and 10 in accordance with state requirements. The test was devised by the state and referenced to objectives common to many state-approved text series. Items were matrix sampled; not every student was asked to respond to identical questions. Thus, data for individual students were not reported. Results focused on grade level and school patterns.

A second test used by Bayview was the norm-referenced, standardized Comprehensive Test of Basic Skills (CTBS). The district had just begun to require this test in all schools for grades 1-9 when CSE fieldwork was conducted. Formerly, it had been given only in schools with Title I (now Chapter 1) compensatory education programs.

The district's proficiency (or minimum competency) testing program was also used in testing-instruction linkage. Forms for grades 5, 9, 10, and 11 had been developed with the help of consultants to meet the state's mandate. These measures covered reading, writing, and mathematics skills deemed essential for life coping. The current forms of the test were introduced in 1978.

The decision-making structure. The data from these three tests were brought to bear on instructional planning in several ways by Bayview district leaders. Chiefly, however, they had begun to use the three test programs mentioned above as content for staff development course work in task analysis and diagnostic-prescriptive teaching.

District leaders had won grant funds from the state to create a Professional Development Center (PDC). The primary focus of the PDC's program was the continuing development of effective teaching strategies. A Teacher Center funded by a federal grant augmented the PDC. Curriculum development and the translation of educational research for practical, instructional applications were the central thrusts of the Teacher Center's program. The very presence of these two centers testified to Bayview's emphasis on teaching-effectiveness skills. In addition, principals were required to attend workshops dealing with supervision, and these focused on the elements of effective teaching.

It was in the context of increasing external test mandates and the emphasis on staff development that Bayview's linkage system began to take shape. From the perspective of District leaders, Bayview teachers and principals were avoiding facing the issues raised by the District's

relatively poor performance on the external measures. In response, said the Director of Staff Development:

> We [at the central office] tried to model a problem-solving way of looking at it so principals could do similarly in their schools. The Director of Instruction worked with principals in the way he wanted them to work with teachers. Also, we asked teachers if they were addressing areas of the test. They said they were. When we observed, we found teachers had difficulty defining the skills to be taught as well as diagnosing for these skills. As a result, we built task analysis cycles into our Professional Development Center programs focusing on the low scoring skill areas identified by the State Assessment Program.

The district's cadre of leaders began by training principals to examine SAP (and later the other tests) to see what specific skills they assessed. Once these were identified, the next step was for principals and their faculties to examine school curricula in order to determine whether these skills were being taught and if so at what grades and with what emphasis. Staff development provided principals, and later teachers, with the information and techniques they needed to do this.

This was taking place with varying degrees of thoroughness in different Bayview schools when CSE's Evaluation Design Project staff made its several visits. At the same time, areas of curricular and instructional weakness district-wide had been identified by district administrators. These areas were then targeted for sessions on diagnostic-prescriptive teaching and other instructional skills.

Analysis of test results also suggested areas for emphasis in the development of continua. Citing the impact of proficiency-test skill and score analysis, for example, the Bayview Coordinator of Curriculum said:

> The proficiency exam has helped the district focus on curriculum... [We learned that] in math we teach computation but the test tests applications through story problems.

Thus, in the Bayview Unified School District, task analysis of tested skills served as the basis for a comprehensive examination of the district's curricula and suggested areas of curricular weakness. Simultaneously, analysis of test results led to the identification of teaching weaknesses. Links between testing and instruction were generated through the development of district-wide objectives and in Professional Development Center and Teacher Center programs.

Guidelines

The St. John and Bayview districts had put in place very different kinds of systems for linking the results of externally mandated testing with instructional planning in their schools. Nevertheless, it is possible to abstract a number of guidelines from their "outside in" models. Other districts would be well advised to bear these in mind should they follow similar approaches.

1. Make test score data comprehensible for teachers and principals.

Providing test results in a format that facilitates their use is obviously a key to testing-instruction linkage. That professional educators working in the schools can be bewildered and intimidated by reports of scores from externally mandated measures was clear in Test Use Project fieldwork (cited early on in this paper). It was equally apparent in the early experiences of district administrators in both Bayview and St. John. The latter addressed this problem by translating the scores into succinct, easy-to-read, relevant tables and graphs. Bayview dealt with it by teaching principals and teachers to dissect the tests and test results.

2. Train teachers and principals to use test scores as diagnostic tools.

As noted near the outset of this paper, the results of externally mandated tests are commonly used in a brief and casual way to get a general, comparative reading on group performance. The essence of their use in the St. John and Bayview systems was diagnostic. They played a role in identifying patterns of strength and weakness in particular content areas and skills. They served to stimulate questions such as "Why are we scoring as we are scoring in this area?" and "How can we improve?" Diagnostic uses are not routine in most schools. Simply presenting test scores in clear, readable format does not mean that they will occur. Teachers need teaching and practice in hypothesizing the different factors that underlie test performance. They need instruction and help in abstracting meaning from scores. Survey findings suggest that most districts do not provide this. In different ways, both St. John and Bayview did.

3. Expect that results of externally mandated tests will serve as only one source of information in planning and decision making.

Wisely, neither Bayview's cadre of leaders nor St. John's district evaluation specialists tried to make test results the sole basis for educational decisions. Human values and priorities do and should influence decisions about what objectives to pursue in school improvement or to build into district continua. The day-to-day experiences with students and observations of their work that teachers and principals rely upon so heavily are hardly irrelevant in making instructional decisions. These factors were routinely accepted, along with test data, as bases for decision making by St. John administrators as they assisted School Planning Councils and

reviewed their plans. Bayview's Coordinator of Staff Development, too, recognized that test data needed to be examined in light of other factors. "When we see through our task analysis and curriculum review what we are and are not teaching, the next step is to ask, 'Do we or don't we want to teach this? How important is it for our students.'"

Data from externally mandated tests can serve to identify problems, to support or disconfirm experience-based judgments, and to stimulate questions. They can be used to justify or rationalize decisions that have already been made. But as the separate experiences of St. John (recall their needs assessment questionnaires) and Bayview (recall their juxtaposition of multiple measures to district curricula) indicate, test data in themselves are only one important source of information for educational planning.

Summary and Conclusions

CSE's national survey and its fieldwork in two research projects suggest that both testing that is internal to the school and that which is externally mandated can be used more fully in systematic educational decision making. Districts can build a curriculum and tests that can serve teachers' routine classroom needs and simultaneously provide consistent, reliable, and valid data for school and district decision making. Districts can also capitalize upon data from externally mandated testing by adapting it to local needs. No single approach or model will be appropriate to every setting. But whether a district chooses to pursue linkage from the inside out or from the outside in, there are several factors that will be necessary for success.

One of these is district leadership. In each district studied by CSE, there was an individual or a small group in the district office -- idea champions and supporters -- who were vitally interested in using test data in instructional planning and decision making. CSE's national test use survey substantiates that such leaders make a difference in school-level uses of test information.

A second element in district success is an organizational arrangement -- a setting and set of procedures -- for decision making. In Central City schools there were the weekly meetings of unit teams; in St. John, regular sessions of the School Planning Councils. Shelter Grove held its principal-teacher planning days in June, October, and February each year. In Bayview, the locus of linkage was staff development workshops, continuum-building committees, and regular school faculty meetings. These organizational arrangements motivated and structured the use of test results by creating (1) real needs for information, and (2) procedures by which the implications of test-score patterns could be discussed and acted upon. None of the field study districts with successful linkage systems simply offered schools test data and left their use to chance.

Third, each of the districts managed testing and/or test results such that they increased the marginal utility of test information for teachers and principals. Teachers routinely receive data on student achievement as they watch their students in class, review their assignments, and grade classroom tests. These data are immediate, rich, and compelling. So too is the information principals regularly gather as they talk with staff and visit their classrooms. To be equally useful and compelling, external test

information must add "something new" to what teachers and principals already know. Each of the four models described above did this. Central City's computer-scoring-and-analysis system for unit tests summarized individual students' mastery of objectives, as well as their errors and weaknesses. Shelter Grove compiled data on the progress of individuals and instructional groupings toward benchmark goals. St. John's Academic Performance Profiles charted year-to-year trends and annual gains. Bayview's task analysis projects based on tested skills and test scores helped to reveal why and how students' performance was what it was. In each case, test data were configured in ways that told teachers and principals something more than "your students are doing well in this and not so well in that" -- which is information teachers and principals typically feel they already have.

A fourth and final element in successful district linkage is the maintenance of on-going resource and support systems. In the districts studied, these centered in the area of staff development: training in test development and use, training in how to realize instructional goals derived from test information, or both. Frequently, too, instructional support staff -- learning specialists, media specialists, evaluation specialists -- were routinely available to provide help and answer questions. Support also took the form of adaptability and flexibility on the part of district administrators. Clear channels were open for Central City and Shelter Grove teachers to participate in the development of and criticize the quality of district curriculum and tests. St. John's evaluation specialists revised district needs-assessment surveys in light of teachers' feedback;

local schools could add survey items suitable to their particular concerns. Bayview district leaders showed patience and understanding in encouraging principals and teachers to take a "problem-solving approach" to low test scores. And of course, each district supported its testing-instructional linkage system with release time and other resources.

The models and guidelines suggested here will not answer all the questions and concerns school districts will encounter as they work to link testing and instruction in systematic ways. But they do indicate productive directions to the more efficient use of testing and the improvement of education planning in American schools.

References

Burry, J., Catterall, J., Choppin, B., & Dorr-Bremme, D. Testing in the nation's schools: How much? What kinds? To what ends? At what costs? CSE Report No. 194. Los Angeles: UCLA Center for the Study of Evaluation, 1982.

Dorr-Bremme, D. Assessing students: Teachers' routine practices and reasoning. Evaluation Comment, 1983, 6(4), 1-12.

Herman, J.L., & Dorr-Bremme, D.W. Testing in the nation's schools: A national profile. In W.E. Hathaway (Ed.), New directions for testing and measurement: Testing in the schools. San Francisco: Jossey-Bass, 1983, 7-17.

Yeh, J.P., Herman, J.L., & Rudner, L.M. Teachers and testing: A survey of test use. CSE Report No. 166. Los Angeles: UCLA Center for the Study of Evaluation, 1981.