DOCUMENT RESUME

ED 244 987 TM 840 310

AUTHOR Herman, Joan L.; Dorr-Bremme, Donald W.
TITLE Teachers and Testing: Implications from a National Study. Draft.
INSTITUTION California Univ., Los Angeles. Center for the Study of Evaluation.
SPONS AGENCY National Inst. of Education (ED), Washington, DC.
PUB DATE Apr 84
NOTE 38p.; Paper presented at the Annual Meeting of the American Educational Research Association (68th, New Orleans, LA, April 23-27, 1984).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Administrator Attitudes; Educational Needs; *Educational Testing; Elementary Secondary Education; *National Surveys; Principals; Teacher Attitudes; *Teacher Behavior; Teacher Education; *Teacher Made Tests; Test Selection; *Test Use
IDENTIFIERS *Curriculum Embedded Tests

ABSTRACT

This paper presents findings from a study of teachers' and principals' testing practices. The research included a nation-wide survey, exploratory fieldwork in preparation for the survey, and a case study inquiry on testing costs. Teachers and principals share misgivings with some of the research community about the appropriateness of required tests for some students, and about their quality and equity. Teachers seem to use test results temperately--as one of many sources of information. As a result of required testing, more time is spent in teaching basic skills and less attention can be paid to other subject areas. The survey also suggests that those in the education and testing communities have paid far too little attention to the matter of teachers' assessment skills. Teachers essentially receive neither training nor any kind of supervision nor any supporting resources in the development of their own tests. Given their frequency and importance at the elementary school level, the findings also suggest curriculum-embedded testing as another neglected area of inquiry. Finally, formal measures should have three important qualities: a close match to curriculum, immediate availability and accessibility, and feelings of ownership. (BW)

DRAFT

Teachers and Testing: Implications from a National Study

Joan L. Herman
and
Donald W. Dorr-Bremme

Center for the Study of Evaluation
UCLA

The research presented herein was performed pursuant to a grant from the National Institute of Education. However, the opinions expressed herein do not necessarily reflect the position or policy of the National Institute of Education, and no official endorsement by the National Institute of Education should be inferred.

Paper Presented at the Annual Meeting
of the American Educational Research Association
New Orleans, LA
April 1984

Teachers and Testing: Implications from a National Study

Abstract

Joan L. Herman and Donald Dorr-Bremme
Center for the Study of Evaluation
UCLA

This paper presents findings from a national survey of teachers' and principals' testing practices. Implications are drawn for staff development and training in test development and selection, clinical decision-making, and assessment of higher level skills; for quality control in curriculum-embedded testing, and for structuring district and school testing programs to facilitate their use by teachers.

Introduction

Fueled by school board accountability concerns, minimum competency mandates, evaluation requirements for federal, state and local programs, and the growth of curriculum-embedded and continuum-based assessment systems, achievement testing in American schools has become both an enterprise of significant scope and visibility and the subject of considerable public discussion and debate. Critics have attacked the arbitrariness of current testing practices (Baker, 1978), have expressed concerns about their validity and bias (Perrone, 1978), have accused testing of narrowing the curriculum and have questioned the value of traditional testing amidst changing functions of education (Tyler, 1978). The quality of available tests continues to be controversial (CSE, 1979; The Huron Institute, 1978), at least one major teachers' organization has called for a moratorium on the use of standardizxed tests, and vigorous legal battles have been launched.

Responding to these various challenges, advocates of testing have reaffirmed its importance and reasserted the variety of purposes that current tests can and do serve. Supporters have maintained, for example, that testing promotes accountability, facilitates more accurate placement and selection decisions, and yields information useful for curricular and instructional improvement.

The testing controversy rages on while the nation's considerable investment in achievement testing continues. Although the stakes in the debate are high, public policy in this arena has plodded on without

the benefit of basic information about the nature of testing as it actually occurs and is used in schools. How much testng really goes on? How are test results used? What functions do tests serve for teachers and principals? What are the effects on schools of various local, state and federal mandates? These and similar questions have gone largely unaddressed. A few studies have indicated teachers' reservations about the limited use of one type of achievement measure -- the norm-referenced standardized test (Airasian, 1970; Body et al, 1975; Goslin, 1965; Goslin, Epstein and Hilloch, 1965; Resnick, 1981; Salmon-Cox, 1971; Statz and Beck, 1979). Beyond this, however, the landscape of testing practices and test used in American schools have remained unexplored.

In this context, the UCLA Center for the Study of Evaluation's (CSE) three year study provides educational policy-makers with basic, new information on classroom achievement testing across the United States. Conducted from 1979 through 1983, CSE's research was designed to take a comprehensive picture of national testing practices. It investigated a wide range of types of formal assessment measures (e.g., commercially produced norm- and criterion-referenced tests and curriculum embedded measures, tests of minimum competency and functional literacy; district-, school-, and teacher-developed tests) as well as some less formal means for gauging student progress and achievement (teachers' observations of and interactions with learners). Within this broad range, inquiry focused on achievement testing practices in reading/English and in mathematics, basic skills areas which are the subject of continuing public concern. Teachers and principals at both elementary and secondary grade levels served as primary subjects for the nationwide survey, addressing those grade levels

which had been identifed in prior research as important transition points and the targets of frequent testing.

A nation-wide survey of teachers and principals was central to the study, and results of this survey form the basis of the report that follows. The research also included exploratory fieldwork in preparation for the survey and, following the survey, case study inquiry on testing costs. During these phases of the project, intensive interviews were conducted with approximately 100 school-level educators in five school districts across the country.

Below, we first provide a brief description of the survey sample, then continue with survey findings on three major questions:

1. How much and what kinds of achievement testing take place in the nation's schools?

2. How important are the results of different types of assessment in teachers' routine tasks?

3. What are schools' and districts' administrative practices with regard to testing and test use?

We conclude by considering the findings in light of the current testing controversy and explore the study's implications for teacher training, quality control, and for structuring district and school testing programs to facilitate their use by teachers in the classroom.

The Survey Sample*

The survey addressed a nation-wide sample of principals and teachers drawn through a successive, random-selection procedure. First, a nationally representative probability sample of 114 school districts was drawn, stratified on the basis of district size, minimum competency testing policy, socioeconomic status, urban-suburban-rural locale, and geographic region of the country. (A lattice sampling technique was used to select cells from the matrix defined by these five stratifying variables, and then random sampling to select districts within a cell.) Next, from within these districts, size permitting, two elementary schools and two high schools were randomly selected using a procedure that facilitated (where possible) inclusion of schools at levels serving both higher- and lower-income populations. Finally, in each of these schools, principals received directions for randomly drawing four teachers for inclusion in the study. Directions for elementary principals guided the random selection of two fourth-grade and two sixth-grade teachers; those for high school principals, the random selection of two teachers of tenth-grade English and two of tenth-grade mathematics. The principal and each of the four participating teachers received received questionnaires that elicited detailed information on their individual and school testing practices, as well as related contextual and attitudinal data.

*A detailed description of the sampling procedure and results is contained in a separate report (Choppin, et. al, 1981). This information has not been reproduced here in order to avoid redundancy. Readers interested in more information regarding the sample and procedure used to draw it are referred to that earlier work.

Returns were obtained from 220 principals, 475 elementary school teachers, and 363 high school teachers in 91 of the 114 districts sampled. Return rates from all principals and from teachers at the elementary level were approximately 60%. About 50% of the high school teachers in the sample responded. To correct for differential return rates by sampling cell and to approximate a nationally representative distribution of respondents, weightings were applied in all descriptive analyses. The results reported below, therefore, represent weighted estimates of national testing practices, test use patterns, and principal and teacher perceptions on testing-related issues.

How Much Testing Goes on in Schools?

Survey results show that the typical student in the upper elementary grades spends, on the average, about 10 hours a year taking reading tests and somewhat more than 12 hours a year taking mathematics tests.[1] (See Table 1.) Test-taking time, then, seems to comprise a little over five percent of the time often allocated annually to formal instruction in each of these subjects. (This figure assumes one hour of daily instruction in each subject for 177 school days per year.)

The typical tenth-grade student enrolled in English, survey results indicate, spends about 26 hours a year completing English tests. This constitutes in the neighborhood of twenty percent of his or her annual time in English class. For the typical tenth grader enrolled in mathematics, taking math tests consumes a little over 24 hours each year -- roughly eighteen percent of the time spent annually in mathematics

1 It is likely that survey results underestimate actual time. The survey asked teachers to fill in all the tests they give over the year and to estimate the student time required for each. It is moot whether they consistently includes all tests.

class. (Here, the percentages given assume daily classes of 45 minutes in each subject, over 177 days per school year.) Clearly, on the average nationally, the frequency and duration of testing in the high school subjects exceed those in the equivalent upper-elementary-school subjects. (Refer again to Table 1.)

The annual times for testing reported are estimates of students' test-taking times. They can probably only serve as rough indicators of the times that the teachers in question spend giving tests in the classroom. On-site interviews (Dorr-Bremme, 1982) suggest that elementary teachers spend only about a quarter to a third of their total time on testing actually giving tests in the classroom. That is, for each hour they devote to giving a reading or math test, they typically spend another two or three hours in such activities as preparing for testing (e.g., constructing and dittoing the test, reviewing directions for standardized testing), correcting and grading tests (or checking over students' standardizedtest answer sheets), recording scores, etc. (Time spent consulting test results and otherwise "using" them is not included here.) Thus, elementary-school teachers' annual time on testing far exceeds the typical student's. (Case studies in two elementary schools found that teachers spent on the average of 200 to 250 hours per year, in and out of class, in achievement testing in all subject areas--or roughly 12 to 15 percent of their reported annual work time.) Resources were not available for detailed case studies in high schools, but pre-survey interview data indicate that the average testing time per year of high-school teachers is also much greater than their students'.

Table 1

Time Devoted to Testing in Typical Classes

	Total Amount of Class Time Spent on Testing per Annum	No. of Test Sessions for Typical Student	Average Length of Session
Elementary School (Grades 4-6)			
--Reading Tests	9 hrs. 56 min.	22	27 min.
--Mathematics Tests	12 hrs. 28 min.	23	32 min.
10th Grade English Class	26 hrs. 34 min.	49	32 min.
10th Grade Mathematics Class	24 hrs. 18 min.	45	33 min.

Table 2

Time Devoted to Required Testing,
As a Percentage of Total Testing Time
For Typical Classes

	Percentage Time on Testing Required by State	Percentage Time on Testing Required by Local School District	Percentage Testing Time Devoted to Non-Required Tests
Elementary School (Grades 4-6)			
--Reading	30	29	41
--Mathematics	21	25	54
10th Grade English Class	12	13	74
10th Grade Mathematics Class	9	14	77

How much of the testing just described is required by the educational hierarchy beyond the school? How much is undertaken at the discretion of teachers? Table 2 provides data to answer these questions. Elementary teachers in the sample report that about half the testing they conduct both in reading and in math is required by their state or school district. At the high school level, about one quarter of the classroom assessment in both English and mathematics results from state or school-district mandates. Notice, however, that since high school students on the average spend twice as much time annually being tested as elementary students do, these percentages suggest that the actual number of hours spent in required testing is quite similar at both levels of schooling. Notice, too, that a greater proportion of assessment in the high school subjects is voluntary: conducted at the discretion of the individual teacher.

What types of tests are used most heavily? Which types consume larger proportions of classroom testing time? As Table 3 shows, tests developed by individual teachers and schools and, at the elementary level, those which accompany curriculum materials, occupy the great majority of classroom testing time. Of all the test types listed, these are the types over which teachers have most control. They can administer them when they deem appropriate; they can design (or readily adapt) the content to suit their own teaching emphases. Most teachers interviewed said that these types of tests fit best with their instructional schedules and curricula. And, from their points of view, these are the most valid instruments of those listed for such routine tasks as grading, on-going planning of teaching, etc. The predominance of

locally developed tests at the secondary level supports the notion that high school teachers have more control over classroom assessment than do elementary school teachers. But heavy use of locally developed tests in the high schools may also reflect that they have fewer suitable commercial testing materials available. Comprehensive curricular programs -- including texts with coordinated workbooks, tests, etc. -- are more widely available for teachers of the elementary grades.

Finally, note that the two types of testing most often generated by state policy -- minimum competency testing and state assessment -- consume on the average very small proportions of classroom testing time.

How are Test Results Used?

Long lists of tests' purposes have been provided in almost every test and measurement text in education. Lists of such purposes usually include selection, placement, remediation, instructional improvement, teacher assessment, accountability, and so on. But to what extent do these ideals represent reality? The survey questionnaires sampled a variety of potential purposes and examined the extent to which the results of particular types of tests and other methods of assessment actually serve each.

Teachers also were asked to rate the importance of a variety of assessment types for activities in which they routinely engage. The results in Table 4 show that both elementary and secondary teachers do see test results of various types as useful in making a variety of decisions. Clearly, however, teachers accord the highest importance to their own observations of students' work and to their own clinical

Table 3

Types of Test Used,
As a Percentage of the Total Time
Devoted to Testing

TYPE OF TEST	Elementary Teachers		10th Grade English Teachers	10th Grade Mathematics Teachers
	Reading	Math		
Tests which form part of a statewide assessment program	3	3	5	1
Required Minimum Competency Tests	1	2	1	1
Tests included with curriculum materials	28	35	8	17
Other commercially published tests	17	18	6	3
Locally developed and district adopted tests	13	8	5	2
School or teacher developed tests	37	35	74	76

judgments. For initially grouping or placing students in a curriculum, for changing students from one group or curriculum to another, and for assigning grades, nearly every teacher respondent reported that their "own observations and students' classwork" is a crucial or important source of information. The great majority of respondents also indicate that the results of the tests they themselves develop also figure as crucial or important in these decisions. Many elementary school teachers also responded that the "results of tests included with the curriculum being used" are quite influential in their instructional decision-making.

These results indicate that while teachers do not attribute heavy importance to the results of required tests, they do view them as somewhat useful sources of data for decisions about initial planning and placement of students in groups or curriculum, and even for decisions about reassigning studentsto different instructional groups or curricula throughout the year. In this last process, they probably serve as a kind of benchmark for judging individual student's "capabilities." For example, imagine a situation where a student is performing poorly in his or her instructional group. A teacher might examine standardized test results to determine whether the problem is "low ability" or whether other factors such as motivation seem a more likely explanation, and then base instructional decisions accordingly.

It is apparent from these results that teachers use a variety of sources to make each kind of decisions listed; they do not rely only upon a single information source. As one teacher stated:

Table 4

Importance of Test Results for Teacher Decision-Making in Elementary and Secondary Schools*

Decision Area:	Standardized Test Batteries	District Continuim or Minimum Competency Tests	Tests Included with Curriculum	Teacher-Made Tests	Teacher Observations/ Opinions
			ELEMENTARY		
Planning teaching at beginning of the school year	2.53 (0.74)	2.60 (0.79)	----	----	3.39 (0.76)
Initial grouping or Placement of students	2.51 (0.74)	2.59 (0.82)	2.91 (0.74)	3.12 (0.83)	3.58 (0.78)
Changing a student from one group or curriculum to another, providing remedial or accelerated work	2.52 (0.79)	2.52 (0.81)	3.04 (0.74)	3.12 (0.84)	3.66 (0.72)
Deciding on report card grades	1.62 (0.76)	1.81 (0.81)	2.89 (0.79)	3.38 (0.74)	3.69 (0.72)
			SECONDARY		
Planning teaching at the beginning of the school year	2.22 (0.84)	2.38 (0.93)	----	----	3.59 (0.60)
Initial grouping or placement of students	2.28 (0.92)	2.46 (0.98)	2.48 (0.92)	3.04 (0.87)	3.84 (0.85)
Changing students from one group or curriculum to another, providing remedial or accelerated work	2.52 (0.95)	2.59 (0.86)	2.67 (0.93)	3.27 (0.76)	3.61 (0.66)
Deciding on report card grades	1.36 (0.66)	1.45 (0.64)	2.29 (0.96)	3.65 (0.62)	3.68 (0.65)

* [4-point scale: 4 = Crucial Importance - 1 = Unimportant or not used]

> "You can't count a score on one test too heavily. The kid could be sick or tired or just not feel up to doing it that day. Maybe his parents had a fight the night before. Maybe he doesn't try. Maybe he doesn't test well." (Choppin , et al, 1981)

Not only do survey respondents indicate that they consult several sources of information about students' achievement in making particular instructional decisions, respondents -- and particularly those at the elementary school level -- also report thinking that many kinds of assessment techniques give them crucial and/or important information. The data in Table 5 are illuminating here: over half the elementary school teachers surveyed report giving heavy weight to each of many sources of information in planning their teaching, in making initial groupings and placements, and in modifying instruction throughout the year.

What are Schools' and Districts' Administrative Practices in the Area of Testing and Test Use?

A growing literature suggests that district and/or school leadership is a significant determinant of whether and how educational innovations and practices are sustained (Berman & McLaughlin, 1978; Bank & Williams, 1982; Edmonds, 1979). Thus, the Test Use in Schools survey examined the practices of school and district administrators in: (1) making, and holding teachers accountable for curricular decisions based on test scores; (2) monitoring and/or supporting school and classroom testing practices; and, (3) providing information and staff development on testing.

Making and holding teachers accountable for test-score-based curricular decisions. The school and district administrative practices in this area that were included on the survey appear in Table 6.

Table 5

Proportion of Teachers who Report Considering Many Types of Assessment Information Critical/Important for Given Activities

	Planning Teaching at Beginning of School Year	Initial Grouping or Placement of Students	Changing Grouping or Placement	Deciding on Report Card Grades
Number of Sources of Information Given in Question on Survey	4	7	6	6
Number of Sources Defined as "Many" for Purposes of this Analysis	3	4	4	4
Proportion of Elementary Teachers who Indicated That at Least this many functioned as Critical and/or Important for the Given Activity	50%	71%	62%	40%
Proportion of High School Teachers	33%	47%	49%	20%

As the table shows, school and district administrators hardly ever establish specific test-score goals for individual schools or teachers. However, district administrators occasionally do check to see that areas in the curriculum that test scores indicate need improvement are in fact being emphasized in their schools; principals monitor their staff members' teaching fairly often toward this same end, particularly in lower SES schools. Often, too (but not, on the whole, as a matter of routine), school administrators meet with teachers in groups or individually to review test scores and highlight their implications for curricular emphases.

Table 6 also indicates that test scores function in making and holding teachers accountable for decisions on curricular emphases less frequently at the secondary-school level than they do in elementary schools. Perhaps this occurs in relation to districts' practices in returning test results. Secondary principals find that scores are only rarely returned by their district such that they can be used in curricular decision making. In elementary schools, the curriculum-embedded tests that accompany basal reading and math series can be used as a basis for cross-classroom analysis of achievement patterns when standardized-test results and other scores are not forthcoming from the district office. (Recall that the use of commercial, curriculum-embedded tests is more prevalent in the elementary grades.)

Monitoring and supporting testing practices. Table 7 displays those school and district practices examined in this area. Of all the practices examined, only one seems to occur more than occasionally: district monitoring of the district testing program. Release time for

Table 6

Making and Holding Teachers Accountable for Test-score-Based Curricular Decisions

	Principals' Reports*		Teachers' Reports*	
SCHOOL ADMINISTRATOR(S) . . .	Elementary	Secondary	Elementary	Secondary
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.

teachers to develop tests is on the whole a rare phenomenon. So, too, are administrative reviews of (a) teacher-constructed tests and (b) student performance on such instruments as unit and chapter tests. (Although not specified in Table 8, the latter test types were mentioned explicitly in the questionnaire item.) These results suggest that there is little monitoring of teachers' classroom testing schedules. They also indicate that one type of measure upon which teachers rely heavily -- tests that they themselves construct -- is most often written individually and with no supervisory review.

Providing staff development and information about testing and test results. Principals were asked to comment on the frequency with which they and district administrators provided in-service experiences germane to testing and test results. In addition, teachers were asked to report on the occurrence of particular types of staff development over the last two years. The responses of principals and teachers to these questions are shown in Tables 8 and 9.

According to principals, staff development for teachers in the area of assessment occurs occasionally, i.e., with a frequency that on the average falls about midway between survey categories "very often" and "rarely." It appears that such staff development is generally initiated slightly more frequently by district administration than by principals.

Of all the topics listed, more teachers report participating in sessions devoted to: (a) analysis and explanation of test results, (b) directions for administering required tests, and (c) how to interpret and use the results of different types of tests. Staff

Table 7

Monitoring and Supporting Testing Practices

	Principals' Reports*		Teachers' Reports*	
	Elementary	Secondary	Elementary	Secondary
SCHOOL ADMINISTRATOR(S) . . .				
Requires teachers to turn in test scores/ grades on classroom tests and/or assignments	2.30 (1.10)	2.32 (1.10)	1.78 (1.17)	2.43 (1.02)
Requires teachers to turn in copies of tests they construct	1.62 (0.92)	2.17 (1.07)	Not Asked	
DISTRICT ADMINISTRATOR(S) . . .				
Conducts observations and/or requires reports to see that all aspects of district testing program are properly carried out	3.09 (0.95)	2.85 (1.07)	Not Asked	
Provides release time and/or extra pay for teachers to develop tests or curricular materials including tests	2.12 (1.03)	2.33 (0.98)	"	

*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 8

Providing Staff Development and Information About Testing

Principals' Reports on Frequency*	Elementary	Secondary
SCHOOL ADMINISTRATOR(S) . . .		
Brings in speakers, workshops, printed material to update teachers' assessment skills	2.62 (0.87)**	2.48 (0.77)
DISTRICT ADMINISTRATOR(S) . . .		
Brings in speakers, workshops, printed material to update teachers' assessment skills	2.73 (0.98)	2.71 (0.90)

* 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.

** Numbers in parentheses are standard deviations.

Table 9

Percentages of Teachers Reporting Participation in Staff Development

	Topic	Elementary	Secondary English	Secondary Math
(1)	Analysis and explanation of state, district, or school test results	84	70	60
(2)	How to administer tests required by my state, district, and/or school (procedures to follow, etc.)	78	54	46
(3)	How to interpret and use results of different types of tests (e.g., norm-referenced and criterion-referenced tests and their applications)	59	35	34
(4)	Alternative ways (other than tests) to assess student achievement	54	25	21
(5)	How to tie what is taught more closely to the skills, content covered on required tests	50	37	25
(6)	Presentation of published materials designed to prepare students for particular tests or to improve test-taking skills	41	32	29
(7)	Training in the use of test results to improve instruction	35	21	19
(8)	How to construct or select good tests	20	23	18

development devoted to increasing teachers' routine classroom assessment skills, these data indicate, occurs much less frequently. Thus, for example, only about a fifth of the teachers in each category report receiving instruction in "how to construct or select good tests," an area in which teachers see a critical need. (See Ward, 1983) Information on other means of assessment (alternatives to testing) was equally rare for secondary teachers, although some 54% of the elementary teachers did report staff development on this topic. Training in the use of test results to improve instruction was evidently provided for 35% of the elementary teachers and about 20% of the secondary teachers sampled.

Finally, it is worth noting that secondary teachers, overall, report receiving staff development in topics related to testing less often than elementary teachers do.

Resources in support of testing. In a set of questionnaire items separate from those discussed just above, teachers were asked to comment on the availability and use of four resources which could support their classroom testing efforts. Teachers' responses to these items (Table 10) are presented in this section since the availability of each of these resources can be interpreted as due, at least in part, to the initiatives of school or district administrators. This is particularly true for item banks of test questions and computerized scoring and analysis of tests. In the case of the other two items included (other teachers with whom I plan and develop tests, someone to help grade tests and assignments), administrators can structure organizational arrangements that facilitate their availability and use.

The list of resources included in the survey instrument was selected on the basis of considerable fieldwork and piloting. Nevertheless, each resource was unavailable to a large proportion of respondents. The exception, of course, was "other teachers with whom I plan and develop tests or other evaluation assignments," but only about a quarter of the elementary-school teachers and a similar fraction of the secondary-school teachers reported taking advantage of this resource frequently. Some 45% of the secondary teachers reported constructing tests with others a few times a year, and fieldwork suggests that this often occurs as teachers in the same department conjointly devise mid-term and final exams.

Computerized test scoring and analysis was reported as used a few times annually by a quarter to a third of both the elementary and secondary teachers sampled. Fieldwork indicates that these reports may reflect the use of optical scanning machines for certain standard (including norm-referenced, standardized) tests. Some districts, however, have developed computer programs for scoring unit and chapter tests and simultaneously analyzing individual students' strengths and weakness on the skills they cover.

A final point: in general, nearly all those teachers who have access to the resources listed report using them at least sometime during the school year.

Table 10
Available Resources for Testing Percentages of Teachers Reporting

Resource	NOT AVAILABLE	AVAILABLE: Not Used	AVAILABLE: Used Once To Several Times/Year	AVAILABLE: Used at Least Once/Month	
Item banks of test questions upon which I draw in making up my tests.	71	4	8	16	Elementary
	51	8	24	16	Secondary
Other teachers with whom I plan and develop tests or other evaluation assignments.	37	12	26	24	Elementary
	21	10	45	24	Secondary
Someone who helps me read, grade, or correct tests and assignments.	69	6	4	21	Elementary
	70	5	4	21	Secondary
Quick, computerized scoring and analysis of tests	64	2	30	4	Elementary
	58	16	22	4	Secondary

Conclusions

We began this discussion by noting the public controversy over the quality and usefulness of testing, a controversy which has been marked by more rhetoric than empirical evidence and one which has centered primarily on standardized tests and large scale assessments. What do the survey results have to say about these concerns and, more particularly, about concerns for the potential misuse and abuse of test results? Teachers and principals do share misgivings with some in the research community about the appropriateness of required tests for some students, and about their quality and equity. Survey findings here, however, allay some concerns about the inappropriate use of tests by classroom teachers. Teachers (and principals, according to findings not reported here) seem to use test results temperately -- as one of many sources of information. They do not give undue weight to any single source, but rather evaluate available data in combination with their own observations to reach decisions. Test results, according to the findings presented here, are thus being used, but not abused.

The influence of test results on school and classroom decision-making is one direct impact of tests, but another impact is felt in the very presence of required testing in the schools. As a result of required testing, school personnel agree that more time is spent in teaching basic skills -- English and math -- and less attention can be paid to other subject areas, and principals and teachers, particularly in lower SES schools, are strongly encouraged to emphasize those skills which are included on required tests. The findings thus confirm the validity of some concerns about the effect of testing on the

curriculum. Admittedly, tests alone have not caused the curriculum to narrow. Rather, the narrowing is a consequence of the importance ascribed by society at large to test scores and of a societal emphasis on basic skills. Nonetheless, it might be well both for public and policymakers to consider whether the limited sample of skills assessed by most standardized tests represents an adequate curriculum and whether test developers, rather than teachers, administrators, school boards and the public, ought to be defining the curriculum.

What else does the CSE research have to tell us? First, the survey suggests that those in the education and testing communities have paid far too little attention to the matter of teachers' assessment skills. For the most part, as mentioned above, the debate on testing has been played out in exchanges about the relative merits of normed and criterion-referenced measures, in discussions of cultural and linguistic biases in standardized tests, in sociopolitical controversy over proficiency testing and so on. It has focused on measures employed nationwide or statewide that generally have been developed by commercial testing concerns or by other large agencies that employ psychometricians. It is appropriate for us to be concerned about the qualities and social implications of such tests. Although they figure less heavily in principals' and teachers' decisions and they consume only small proportions of classroom time, tests of this type do exert significant influence in major educational gate-keeping decisions. However, the quality of teachers' assessment skills, their skills as test developers and as clinical diagnosticians, have largely escaped attention. Yet the cumulative record of teacher-made tests, the grades in which they result, as

well as the teachers' informal judgments of children's competence clearly influence students' educational careers in major ways, perhaps to a degree exceeding that of more formal testing. What is more, students, particularly secondary students, spend large proportions of their testing time taking teacher developed and teacher-scheduled tests.

What do we know about the quality of teacher-developed tests? Very little. And the little we know is far from encouraging. Almost twenty years ago, Ebel (1967) identified common errors in teacher-developed tests and urged better training for teachers in this area. More recent research indicates that teachers remain poorly prepared in assessment (Rudman and others, 1980; Yeh and others, 1981), a finding which is not surprising in light of preservice and inservice requirements and opportunities for teachers. Few states explicitly require competence in testing for teacher certification (Woellner, 1979), and studies have indicated that while most teachers have had at least one measurement course, attention to teacher-developed tests and clinical assessment skills is virtually non-existent (Gullickson, 1984; Ward, 1983). The results reported here indicate that inservice training does little to fill the gap. Only about one-fifth of the teachers in our survey received inservice experience related to the selection and construction of good tests or in the use of testing for classroom decisionmaking and to improve instruction; according to other studies, these are two areas which teachers rate as most important and in which they agree they need help (Gullickson, 1984; Ward, 1983). Clearly, teachers need training opportunities if they are to be competent test developers, skilled analysts, and literate consumers of test information.

Although the study reported here did not directly address the issue of the quality of teacher-made tests, its findings combined with those cited above give cause for some pessimism. Teachers essentially receive neither training nor any kind of supervision nor any supporting resources in the development of their own tests. One of the few studies which have examined explicitly the quality issue raises additional concern. Fleming and Chambers (1983) analyzed teacher-developed tests in Cleveland schools and found that teachers can deal with many of the technical requirements for classroom tests, such as arrangement of test questions, format of test questions, and the avoidance of obvious technical flaws; however, almost one-fifth exhibited errors in mechanics and technical conventions. More disturbing is the fact that the vast majority of test questions reviewed focused on lower-level skills, requiring recall of terms, factual knowledge, rules and principles; test items requiring synthesis and higher level applications accounted for only a very small percentage of the questions. Many have noted that tests communicate expectations to students and identify for them the important knowledge and skills that are required for particular courses; the objectives that really matter for students are those embedded in the tests on which their grades are based (Bloom, 1981). Concerns were expressed earlier, and appropriately so, about curricular narrowing associated with required tests: an equally important issue may be the extent to which the curriculum is being narrowed to memory and rote learning as a function of teacher-developed tests. Teachers, in short, not only need training in test development, but they apparently also need particular assistance in assessing (and perhaps in teaching) higher level skills.

Given their frequency and importance at the elementary school level, the findings reported here also suggest curriculum-embedded testing as another neglected area of inquiry. Like teacher-developed tests, we know very little about the quality of these measures, and, again, what we do know does not give cause for optimism. For example, analyses of commonly-used basal series have criticized their failure to utilize common research-based design principles (Quellmalz and Herman, 1978), and informal perusal of some recent tests indicates some serious flaws, e.g., tests which claim to be diagnostic on the basis of one item per objective. It may well be that some quality assurance mechanisms are needed.

As we think about training requirements for teachers and quality control for commercial tests, it might be well also to explore other testing supports that might be provided for teachers. When taken seriously, test development is an arduous and time consuming process. One might wonder whether teachers, in fact, have the time and energy to produce good tests or whether a better approach might be to explore ways to better enable them to capitalize on and use the efforts of others. Item banks are one possibility, either representing the pooled efforts of teachers within a school/district or commercially available options (although they currently exist, both are likely to have quality control problems). With micro-computers on almost every school campus, the technological requirements are in place for easily accessible tests that can be customized to teachers' unique needs and classroom instructional programs. These same computers can be used to facilitate onerous test scoring, recording, grading and management tasks.

While we work to improve the quality of teacher and curriculum embedded tests, we must also strive to improve the usefulness of more formal measures. CSE's study suggests three general but highly important qualities that more formal measures should have, qualities which are inherent in the teacher-developed and curriculum-embedded tests that teachers use most frequently: a close match to curriculum, immediate availability and accessibility, and feelings of ownership. That is, formal measures must reflect what is being taught in class, and they must be sensitive to teachers' intentions and emphases as teachers themselves perceive them. Moreover, teachers must be able to administer these measures to students when they feel it appropriate, and the results must be both understandable and available promptly. Finally, the content, format and timing of the measures must be under the control and discretion of individual teachers and teachers must feel their needs and input have been influential. Many commercial, state, district, and school testing programs do not reflect these characteristics, and the results are predictable: elaborate systems that are of little use to teachers and that teachers little use. Counter-examples, however, also can be identified, and where these occur we have found that teachers routinely use more formal measures, representing more sophisticated technology and higher technical quality, rather than their own tests.

In summary, our research suggests several complementary avenues for improving the quality and use of tests in schools. First, given the time devoted to teacher-developed tests, it seems well worth considering teachers' preparation for the role of achievement assessor and their competence in that role. Similarly, given the time and importance

accorded curriculum embedded tests, we would do well to examine and better assure that quality of those tests. Finally, we need to investigate ways to provide teachers with tests which they can use routinely, which reflect sound test procedures, and which meet their needs.

Tyler, R. What's Wrong with Standardized Testing. Today's Education, 1977, 66(2), 35-58.

Woellner, R.S. Let's Use Test for Teaching: Standardized Test Results Can Provide the Basis for a Program of Instruction. Teacher, 1979, 90(2), 62-63, 179-181.

Yeh, J.P., Herman, J.L., & Rudner, L.M. Teachers and Testing: A Survey of Test Use. Report No. 166. Los Angeles: Center for the Study of Evaluation, 1981.

REFERENCES

Airasian, P.W. The Effects of Standardized Testing and Test Information on Teachers' Perceptions and Practices. Paper presented at the annual meeting the American Educational Research Association, San Francisco, 1979.

Baker, E.L. Is Something Better than Nothing? Metaphysical Test Design. Paper presented at the 1978 CSE Measurement and Methodology Conference, Los Angeles, 1978.

Boyd, J., Jacobsen, K., McKenna, B.H., Stake, R.E. & Yashinsky, J. A Study of Testing Practices in the Royal Oak (Michigan) Public Schools. Royal Oak: Royal Oak Michigan School District, 1975.

Center for the Study of Evaluation. CSE Criterion-Referenced Test Handbook. Los Angeles, CA: Center for the Study of Evaluation, 1979.

Ebel, R.L. Improving the Competence of Teachers in Educational Measurement. In J. Flynn and H. Garber (Eds.), Assessing Behavior: Readings in Educational and Psychological Measurement. Reading, MA: Addison-Wesley, 1967.

Goslin, D.A., Epstein, R., & Hilloch, B.A. The Use of Standardized Tests in Elementary Schools. Second Technical Report. New York: Russell Sage Foundation, 1965.

Huron Institute. Summary of the Spring Conference of the National Consortium on Testing. Cambridge, MA: Horon Institute, 1978.

Peronne, V. Remarks to the National Conference on Achievement Testing and Basic Skills. Paper presented at the National Conference on Achievement Testing and Basic Skills, Washington, D.C., March 1978.

Resnick, L.B. Introduction: Research to Inform a Debate. Phi Delta Kappan, 1981, 62(9), 623-624.

Rudman, H.C., Kelly, J.L., Wanous, D.S., Mehren, W.A., Clark, C.M., & Porter, AC. Integrating Assessment with Instruction: A Review 1922-1980. East Lansing, MI: Institute for Research on Teaching, 1980.

Salmon-Cox, L. Teachers and Tests: What's Really Happening? Phi Delta Kappan, 1981, 62(9), 631-634.

Stetz, F., & Beck, M. Teachers' Opinions of Standardized Test Use and Usefulness. Paper presented at the annual meeting of the American Educational Research Association, San Fracisco, 1979.

Tyler, R. What's Wrong with Standardized Testing. Today's Education, 1977, 66(2), 35-58.

Woellner, R.S. Let's Use Test for Teaching: Standardized Test Results Can Provide the Basis for a Program of Instruction. Teacher, 1979, 90(2), 62-63, 179-181.

Yeh, J.P., Herman, J.L., & Rudner, L.M. Teachers and Testing: A Survey of Test Use. Report No. 166. Los Angeles: Center for the Study of Evaluation, 1981.