DOCUMENT RESUME

ED 253 588 TM 850 138

AUTHOR Green, Kathy E.; Stager, Susan F.
TITLE Teachers' Attitudes toward Testing.
PUB DATE Mar 85
NOTE 25p.; Paper presented at the Annual Meeting of the National Council on Measurement in Education (Chicago, IL, April 1-3, 1985).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Attitude Measures; Elementary Secondary Education; Factor Analysis; Standardized Tests; *State Surveys; *Teacher Attitudes; Teacher Characteristics; Teacher Education; Teacher Made Tests; *Testing; Test Reliability; *Test Use
IDENTIFIERS Mail Surveys; Wyoming

ABSTRACT

This paper reports the development and testing of measures of teachers' attitudes toward testing and appropriate use of tests. A random sample of 555 practicing teachers in Wyoming were surveyed by mail (81 percent response rate). Five subscales assessing attitudes toward use of classroom and standardized tests were identified: (1) standardized test effects; (2) standardized test results-use; (3) value of classroom tests; (4) fairness of classroom tests; and (5) effectiveness of classroom tests. Their internal consistency reliabilities ranged from .54 to .75. Teachers using tests more often had somewhat more positive attitudes toward classroom test effectiveness. Males had more favorable attitudes toward classroom testing. Attitudes toward standardized testing tended to be indifferent to negative. Results are discussed in terms of instrument development, variable interrelationships, and implications for teacher training. (Author/BS)

TEACHERS' ATTITUDES TOWARD TESTING

Kathy E. Green, Ph.D.

and

Susan F. Stager, Ph.D.

University of Wyoming

Presented at the National Council on Measurement in Education Annual Meeting, Chicago, March, 1985. Appreciation is expressed to the State Department of Education for their support of this project and to the teachers who responded for their participation.

ABSTRACT

This paper reports the development and testing of measures of teachers' attitudes toward testing and appropriate use of tests. A random sample of 555 practicing teachers in the State of Wyoming were surveyed (81% response rate). Five subscales assessing attitudes toward use of classroom and standardized tests were identified with internal consistency reliabilities ranging from .54 to .75. An appropriateness of test use measure reflects the degree to which teachers used contemporary measurement techniques. Results are discussed in terms of instrument development, variable interrelationships, and implications for teacher training.

TEACHERS' ATTITUDES TOWARD TESTING

Testing in American schools has been and continues to be a subject of controversy from the local to the national level. Testing at all levels has increased with the accountability demands placed on district, state, and national educational institutions and the informational demands of objectives-based instructional systems and competency-based evaluation. The overwhelming majority of states have instituted some form of minimum competency testing--for high school graduation, for college matriculation, for teacher certification (Yeh, 1980). The modes and content of tests may be changing, but testing remains a fact of academic life.

Teachers' attitudes toward the tests they give and toward the practice of testing can influence many facets of education: the quality of tests given, the meaning in test scores, the way in which information from tests is used, the evaluations made by students (and parents) as well as by the teacher, and the students' perceptions of themselves, the school, and the instructional process. To some degree teachers also assess their own performance as educators and the effectiveness of instructional techniques on the basis of classroom and standardized test results. The primary purpose of this research was to develop a measure of elementary and secondary teachers' attitudes toward testing. A secondary purpose was to construct a measure of the use of contemporary measurement practices. Teachers' use of tests--the amount of time spent in testing-related activities, frequency of testing, and the degree to which students' grades are based on test results--was used as a corollary measure.

Background

Test scores are a major factor in determining whether a student enters a program, passes a course, is allowed to go to a higher level, or leaves a program.

Tests also have effects on aspects of students' lives other than their progress through an educational system. Kirkland (1971) reviewed studies reporting the effects of tests on students' self-concept, motivation, level of aspiration, study practices, and anxiety level. Also reviewed were studies reporting the effects of tests on teachers. Tests are felt to have shaped the curricula to an extent. Test results have been used to evaluate teacher effectiveness--rightly or wrongly. If teachers use tests extensively by choice and/or by mandate, and if test results influence teachers' behaviors and expectations about their students and about their own performance, then teachers' beliefs about tests, their knowledge of testing, and their opinions about how tests can be used bear strong implications for the educational process. Teachers' attitudes toward testing are part of their attitudes toward teaching as a whole and probably affect their behavior in the classroom.

Teachers have been found to use classroom tests extensively. The reported percentages of classroom tests which are teacher-constructed range from 0-100% (Gullickson, 1982: McKee & Manning-Curtis, 1982; Newman & Stallings, 1982) with an average of from 50 to 93%. Estimates of the percentage of their time teachers spend in test construction, administration, scoring, and return of tests ranges from 8 to over 20% (Carlberg, 1981; Fennessey, 1982; Gullickson, 1982; Newman & Stallings, 1982; Stager & Green, 1984). The estimated average percentage of students' course grades which are based on test scores is 40-50% (Gullickson, 1984; McKee & Manning-Curtis, 1982; Stager & Green, 1984), again with a range of 0-100%. In contrast, while often given annually and at most grade levels, the results of standardized tests are reported to be used very little by classroom teachers (Beck & Stetz, 1979; Fennessey, 1982; Goslin, 1967; Lazar-Morrison, Polin, Moy, & Burry, 1980; Stager & Green, 1984; Stetz & Beck, 1978).

Test results have been found to influence teachers' expectations and behavior (see Airasian, Kellaghan, Madaus, & Pedulla, 1977; Kirkland, 1971; Rosenthal & Fode, 1963; Rosenthal & Jacobson, 1968a, 1968b). Teachers given test score

information modified their expectations to accord with reported IQ or achievement levels. Salmon-Cox (1980) found that teachers used test information to verify their own perceptions. (It should be noted that while the Pygmalion effect has been found in some studies, it has not been found in others.) Teachers' expectations may be based on several sources of information. Arganbright (1983) suggests that teachers develop expectations from classroom encounters from knowledge of the child's family, and from records and tests. Teachers' beliefs in and use of classroom and standardized tests would seem to have both overt (e.g., ability grouping) and covert (e.g., expectations) effects on students. It is somewhat surprising that an area with such potentially far-reaching effects has received so little attention.

Attitudes toward testing have been specifically addressed in a few studies. Lambert (1980) surveyed the attitudes toward testing of three groups: legislators, education deans, and AFT-NEA officials. He used free-response questions. When characterizing the general attitudes of teachers toward standardized testing, the most frequent response was "negative, afraid of results, suspicious, a threat to job security" (p.14). Lambert found respondents to have more favorable attitudes toward criterion-referenced tests (about half having positive attitudes), though education deans commented that they didn't think teachers quite knew what criterion-referenced tests were. Almost all respondents felt it important for teachers to produce excellent classroom tests. In another study, Cramer and Slakter (1968) reported development of a 20-item scale measuring attitudes toward the use of aptitude tests.

Gullickson (1984) surveyed a stratified random sample of 391 third, seventh, and tenth grade teachers in South Dakota. He found teachers to agree that tests "...increase student effort, affect student self-concept, create competition, improve student interaction, and in general improve the learning environment" (p. 247). Teachers reported frequent use of tests but viewed tests as better measures of skills at the higher cognitive levels. Measures used in this study

were single items.

While teachers report spending significant amounts of time testing with substantive effects on students, there is little research on teachers' attitudes toward testing as a practice and little research on test construction prac ces. There is somewhat more research on test use and the use of appropriate testing practices by teachers (e.g., Fennessey, 1982; Goslin, 1967; Gullickson, 1982; Newman & Stallings, 1982). This study addresses the former concern by attempting to assess several aspects of teachers' attitudes toward testing. As stated previously, the major thrust of this effort was in measure development. Mean scores for attitude measures are reported, however, as well as the relationships to behavioral and other variables were:

1. Teachers who use tests more often have more positive attitudes toward classroom testing.

2. Teachers using contemporary measurement practices have more positive attitudes toward classroom testing.

3. Teachers having one or more courses in testing and measurement have more positive attitudes toward testing (both classroom and standardized).

4. Teachers' personal experiences as students with tests relate to their attitudes toward testing.

5. No differences exist by age group or by sex in attitudes toward testing (classroom or standardized).

METHODS

Instruments

The first author constructed a pool of items addressing the effects of classroom tests on students (motivation, self-concept, etc.), the value of tests (whether testing is worthwhile), and the effectiveness of test results in promoting change (n=15 items). Also constructed were items relating to the use of and effects on students of standardized tests (n=10 items). Items were rated

on a 1-6 point Likert scale, with higher numbers indicating stronger disagreement with the statement. Statements were phrased both negatively and positively. All items were reviewed by the second author and by two former teachers for clarity and appropriateness. Items w[illegible]sed and incorporated in a survey form containing a t[illegible]l of 49 questions. Ot[illegible]r items queried:

'demographics'- age, sex, grades/subjects taught, coursework in tests and measurement, recency of coursework, degree(s) held, years in teaching

use of tests - time, freqv[illegible]cy, proportion of grade based on tests

use of suggested testing techniques - table of specifications, Bloom's taxonomy, etc.

personal experiences with tests - fairness, value, liking for tests

overall attitude toward testing - classroom and standardized

Subjects

Our goal was to survey approximately 500 teachers--a sample size adequate to allow factor analysis of 25 items and to allow analyses by subject area and grade level. The size of the sample was based on expectations of a 70% return rate. A systematic random sample was chosen from the State Department of Education list of all Wyoming educators. During the spring semester, these teachers were sent a letter explaining the nature of the study, a survey form, and a stamped return envelope. A return rate of 55% was obtained from the first mailing. With two follow-ups, a total of 555 replies were received, or 81% of the deliverable envelopes. (Twelve were undeliverable, 4 refused, and 133 did not reply.)

Characteristics of the participating subjects are summarized in Table 1. The sample includes a greater percentage of females, primarily as a consequence of the over-representation of females among elementary school teachers. The greatest percentage of teachers in the total sample and at each of the three grade levels is in the 30-39 year-old age range. The average number of years of teaching experience is 12. All teachers in the sample hold bachelor's degrees,

with 23% holding master's. Subject area responsibilities seemed representative of Wyoming teachers. The majority of elementary teachers are responsible for all areas; at the junior and senior high levels the most frequently reported areas are in core subjects (English, math, science, social studies, physical education).

(Table 1 about here)

RESULTS

Construction of Measures

Attitude items were recoded so that higher numbers represented a more positive attitude toward testing. Items were then grouped using principal components analysis followed by varimax rotation. Pairwise deletion of cases with missing values was used since the percentage of missing data for any one item was small and since no combination of items was systematically skipped. Initially, all items were entered into the analysis. When this resulted in a solution which was very difficult to interpret (having a number of factors with only one item loading substantially on that factor), items relating to standardized tests vs. classroom tests were analyzed separately. Analysis of standardized test items resulted in four significant factors being found (68.5% of the variance accounted for.) When intercorrelated with the remaining two clusters. It was decided, therefore, to fit the items to a 2-factor solution since subscale scores were to be used and not factor scores. Table 2 presents items, factor loadings from the two-factor solution, item means, and standard deviations. No total score was calculated since the resulting two subscales were not conceptually related nor were they moderately correlated.

(Table 2 about here)

Principal components analysis of items relating to classroom testing resulted in three significant factors (45.2% of the variance accounted for). One item loaded marginally on several factors and was therefore dropped from further analysis. The factor-based groupings were named test value, effectiveness, and

fairness. When forced to a two-factor solution, value and effectiveness collapsed to the same factor. Items were also analyzed using oblique rotation instead of varimax. Results were parallel, though, of course, factor loadings differed.

Table 2 presents items with factor loadings from the three-factor solution. Using the factor-based item groupings as subscales, internal consistency reliability estimates (Cronbach's alpha) were calculated for the subscales. Table 2 also presents the item-total correlations and subscale alpha's. Internal consistency reliability estimates ranged from .54 to .75; subscales had from three to six items. Since a reliability estimate of at least .50 is recommended for use of a measure in making group comparisons, the subscales were deemed useful in experimental work. Clearly, increasing the length of the subscales by adding items of homogeneous content would increase the internal consistency reliability.

A total score was calculated for attitude to classroom testing by combining all items. The reliability of the total scale was .72.

Subscale intercorrelations are presented in Table 3.

(Table 3 about here)

Frequency of use of contemporary measurement practices was rated on a 1-6 Likert scale, with higher numbers indicating more frequent use of those practices. A measure of use was developed by summing responses across the following:

Frequency of use of--

- behavioral objectives
- table of specifications
- higher-level questions
- files of previously used test items
- Bloom's taxonomy
- test reliability
- descriptive test statistics
- item difficulty levels
- item analyses

The internal consistency reliability of this measure was .79. The average reported use of these measurement practices was between "rarely" and "sometimes", with test

statistics and tables of specification being the least frequently used and behavioral objectives the most frequently used.

Five questions asking about personal experiences with tests were also included in the survey. Three questions concerned personal experience/liking for tests. These three were summed as a single measure (alpha = .61). These three questions were:

- I did well on tests when I was in school. ($\bar{x}$ = 1.77, SD = 1.13)
- I personally dislike taking tests. ($\bar{x}$ = .3.46, SD = 1.16)
- At present I have no objection to taking tests. ($\bar{x}$ = 3.25, SD = 1.00)

Two additional questions asked about the value and fairness of tests:

- I have not found my <u>own</u> test results to be of much value to me. ($\bar{x}$ = 4.22, SD = 1.07)
- The tests I took as a student were generally good assessments of my knowledge of an area. ($\bar{x}$ = 3..9, SD = .99)

Concurrent validity estimates for subscales were obtained by correlation with self-rating statements of overall attitude toward testing (classroom and standardized). These correlations were significant ($p<.05$) and low to moderate. It should be noted that with a sample size of over 500, very low correlations will be significant.

(Table 4 about here)

Subscales were also related to frequency of giving tests, percentage of students' grades based on tests, time spent testing, and use of contemporary measurement practices. These correlations are also presented in Table 4. Total scale score correlated more highly with alternative measures than did subscale scores. Correlation with these measures were significant for the most part but low. These results suggest that the constructed measures do relate to another attitude measure but have low predictive validity for testing-related behavior.

<u>Hypothesis Tests</u>

In regard to our a priori expectations, it was found that teachers who use tests more often have somewhat more positive attitudes toward the effectiveness of classroom tests than do teachers who use tests less frequently. Correlation with

the value and fairness of classroom tests was lower. Teachers using suggested measurement practices have somewhat more positive attitudes toward classroom tests. Relationships with attitude subscales were not significant. Teachers who spend more time testing and who base more of students' grades on test scores have somewhat more positive attitudes toward testing. Again, it should be noted that while these relationships were significant, they were weak.

No overall differences were found in attitudes between teachers having had one or more tests and measurement course(s) vs. those not having had such training ($F_{5,532} = 1.50$, $p<.19$). Selected subscales were also related to personal experiences with tests, perceived value of tests, and fairness of tests. Correlations with these measures were mainly significant and moderate (Table 4). This confirms our a priori expectation that there would be significant relationships among these variables. Teachers who found their own tests results to be of personal value reported more positive attitudes toward the value and effectiveness of classroom tests. Teacher who thought tests they took were fair assessments of their knowledge reported more favorable attitudes toward the fairness of classroom tests. Reported liking for tests did not relate as strongly to attitude subscales.

Group differences in attitudes toward testing were investigated using multivariate analyses of variance followed by univariate tests. Contrary to expectation, significant overall differences were found for sex ($F_{5,508} = 7.67$, $p<.01$). Males had significantly more positive attitudes toward testing on all classroom testing subscales (see Table 5). Differences were not significant for attitudes toward standardized testing. Since grade level was found to be related to attitude, and since sex of teacher was related to grade level, a further analysis was conducted to determine if sex differences held within each grade level. No significant interaction was found between sex and grade level for any of the attitude subscales. Males had more favorable attitudes toward classroom testing across grade levels although differences were not significant at the junior high level. No significant

sex differences were found within content area taught (English, math/science, art/music, physical education). Overall differences for age groups were also significant ($F_{15,1463} = 1.70$, $p<.05$). Attitude differences were significant, however, only for test effectiveness and for effects of standardized tests (Table 5).

(Table 5 about here)

Significant overall differences were found for grade level taught ($F_{10,1008} = 3.07$, $p<.01$). Teachers' attitudes toward classroom testing were more positive as grade level increased. Differences in attitudes toward standardized tests were not significant (Table 5).

DISCUSSION

The results of this study should be interpreted in light of three facts: (1) the teachers who responded may not be representative of teachers in other areas, especially urban areas; (2) all data were obtained via self-report; and, (3) no causali can be implied from the design used in this study. The reliabilities of the measures developed seemed adequate for initial use. If these specific measures were to be used in subsequent work, it is recommended that the subscale "value" be lengthened. The conceptual clarity of the scales also needs to be examined. Scale items factored along negative and positive dimensions of attitudes toward testing. It may be preferable to try out items which are all positive and still written to assess several facets of attitudes (e.g., value, fairness, effects on students). While the measures developed show adequate reliability, further work needs to be done in content validation, especially if subscales are to be used rather than a total scale. Results of this study suggest that with respect to attitudes toward testing, negative and positive may not be endpoints of the same scale. Perceiving tests as bearing great potential benefit does not preclude seeing tests as potentially harmful. McKee and Manning-Curtis (1982) may have encountered a similar situation when defining "test wariness"

and "test confidence" as aspects of attitude. In fact, both negative and positive aspects may be indices of test awareness. If items were to be analyzed using strength of opinion as the criterion rather than direction and strength, the item groupings would probably shift.

Criterion-related validation also needs to be further investigated. Attitude scores were moderately related to other attitude measures but their relationship to behavioral variables was low. Either attitude (in this case as in others) does not predict behavior very well or behavior needs to be assessed as an aggregate rather than as a single-item, self-report measure. The authors feel that attitudes in and of themselves are interesting but that further investigation of relationship to aggregate measures of behavior would be useful. The relationship between teachers' and students' attitudes would also be of interest.

While relationships between attitude measures and reported use of classroom tests were low, they were consistently positive. This suggests that there may be a true relationship between attitude and test use. But it also suggests that testing practice is probably affected by numerous other factors such as accountability demands, student expectations, and peer expectations. A model of teacher use of tests would include factors such as teacher training, personal experiences with tests, school/district policy, and environmental factors as well as attitudes toward testing. Perceptions of test fairness (positive aspects of tests) related the most strongly to test use.

Attitudes toward testing bore scant relationship to use of contemporary measurement practices. Use of contemporary measurement practices had a somewhat stronger relationship to use of tests ($r = .15$ with frequency of testing, $r = .16$ with time spent in testing). It has been suggested that understanding of technical concepts should contribute to positive attitudes toward tests (McKee & Manning-Curtis, 1982). Use of contemporary measurement practices hopefully relates to understanding of contemporary measurement practices. Results here suggest little difference in

attitude for those persons reporting more skilled use of tests. These persons may be more aware of the positive and negative effects of using tests and may have more ambivalent attitudes. Or, the differences in use of contemporary measurement practices among respondents may be too minor to provide well-defined groups.

The relationships between attitude measures and reported personal experiences were stronger. This suggests that tests and measurement course instructors truly need to practice what they preach and present a model of how tests can be used effectively and fairly. And this may not be enough to counter years of experience with sub-standard testing practices. And, performance may have shaped attitudes independent of the quality of the tests taken. As one teacher commented: "It is my opinion that people who rant about the uselessness of tests never liked the results of their own; i.e., reality was tough on their ego...To do away with tests would be a subversion of our society to the student as 'real life' tests us everyday and we are expected to measure up." Further research with attitude toward testing and personal experience with tests needs to include a measure of achievement as a covariate.

Age differences in attitudes toward testing are consistent with Yeh et al. (1981): Younger teachers seem to be more skeptical of testing. Grade level differences in use of tests have been reported (Gullickson, 1982; Yeh et al., 1981), with teachers using fewer tests at the lower grade levels. Attitudes covary with use of tests for this variable. Teachers in this study reported more positive attitudes toward testing at the upper grade levels.

Reported attitudes toward classroom testing tended to be generally positive. Attitudes toward standardized testing tended to be indifferent to negative. While use of tests was extensive, use of contemporary measurement practices was lacking. Extensive use of tests with its concomitant demand on teacher time and failure to use testing principles held to be important by most textbook authors suggests a need for attention to the testing curriculum. These results, which are consistent with previous research, have led several authors (e.g.,

Gullickson, 1982, 1984; Yeh et al., 1981) to question the efficacy of training in tests and measurement. Newman and Stallings (1982) found teachers to be no more competent in their testing practices now than they were a decade or more ago. Fennessey (1982) argues that training in this area would ideally be focused on the student's area of the curriculum. Thus, multiple sections or blocks of a tests and measurement course would need to be offered, with each tailored to a subject area (e.g., physical education, English, math). The suggestion was also made that grade level be considered, with courses structured for elementary, junior high, and senior high levels. Change could also come in the application of microcomputer technology. Use of micros for testing would hopefully reduce demands on teacher time and would also efficiently provide test and item statistics useful in improving tests. Change could come via a centralization of testing resources (e.g., item banks) which could be made available to classroom teachers at different grade levels in different course areas.

A considerable number of teachers in this study expressed interest in upgrading their skills in tests and measurement. (For example, 47% expressed an interest in inservice training in assessing test reliability and validity, 44% were interested in learning how to use micros in testing, and 42% wanted to learn more about designing classroom tests.)

Restructuring tests and measurement course offerings to make them more compatible with the classroom situation and introduction of new technology as a time-saving measure along with continued professional development for practicing teachers may serve to improve testing practice (and attitudes). As Gullickson (1984) noted, there is a need to reach agreement on differences between what is being taught and what teachers actually need and to develop and implement strategies to meet those identified needs.

REFERENCES

Airasian, P.W., Kellaghan, T., Madaus, G.F., & Pedulla, J.J. (1977). Proportion and direction of teacher rating changes of pupils' attributable to standardized test information. Journal of Educational Psychology, 69, 702-709.

Arganbright, J.L. (1983). Teacher expectations--a critical factor for student achievement. NASSP Bulletin, 67(464), 93-95.

Beck, M.D., & Stetz, F.P. (1979). Teachers' opinions of standardized test use and usefulness. Paper presented at the American Educational Research Assn. meeting, San Francisco.

Carlberg, C. (1981). South Dakota study report. Denver, Co.: Midcontinent Regional Educational Laboratory.

Fennessey, D. (1982). Primary teachers' assessment practices: some implications for teacher training. Paper presented at the annual conference of the South Pacific Assn. for Teacher Education, Frankston, Victoria, Australia. (ERIC Document ED229376)

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

Gullickson, A.R. (1984). Teacher perspectives of their instructional use of tests. Journal of Educational Research, 77, 244-248.

--- (1982). The practice of testing in elementary and secondary schools. (ERIC Document ED229391)

Kirkland, M.C. (1971). The effects of tests on students and schools. Review of Educational Research, 41, 303-350.

Lambert, R.F. (1980-81). Teacher attitudes on testing: a multiple perspective. College Board Review, n118, 13-14,29-30.

Lazar-Morrison, C., Polin, L., Moy,R., & Burry, J. (1980). A review of the literature on test use. Center for the Study of Evaluation, California University, Los Angeles. (ERIC Document ED204411)

McKee, B.G., & Manning-Curtis, C. (1982). Teacher-constructed classroom tests: the stepchild of measurement research. Paper presented at the National Council on Measurement in Education meeting, New York.

Newman, D.C., & Stallings, W.M. (1982). Teacher competency in classroom testing, measurement preparation, and classroom testing practices. Paper presented at the American Educational Research Assn. meeting, New York. (ERIC Document ED22049

Rosenthal, R., & Fode, K. (1963). Experimenter bias. Behavioral Science, 83, 183-189.

Rosenthal, R., & Jacobson, L.F. (1968a). Self-fulfilling prophecies in the classroom: teachers' expectations as unintended determinants of pupils' intellectual competence. In M. Deutsch et al. (Eds.), Social class, race, and psychological development. New York: Holt, Rinehart, & Winston.

Rosenthal, R., & Jacobson, L.F. (1968b). Teacher expectations for the disadvantaged. Scientific American, 218, 19-23.

Salmon-Cox, L. (1980). Teachers and tests: what's really happening. Paper presented at the American Educational Research Assn. meeting, Boston.

Stager, S.F., & Green, K.E. (1984). Wyoming teachers' use of tests and attitudes toward classroom and standardized tests. Dept. of Educational Foundations and Instructional Technology, University of Wyoming, Laramie, Wyoming.

Stetz, F.P., & Beck, M.D. (1978). Comments from the classroom: teachers' and students' opinions of achievement tests. Paper presented at the American Educational Research Assn. meeting, San Francisco.

Yeh, J.P. (1980). A re-analysis of test-use data. Center for the Study of Evaluation, California University, Los Angeles. (ERIC Document ED 205590)

Yeh, J.P., Herman, J.L., & Rudner, L.M. (1981). Teachers and testing: a survey of test use. Center for the Study of Evaluation, California University, Los Angeles. (ERIC Document ED218336)

Table 1. Sample characteristics

	Total Sample (n=555)	Elementary (n=288)	Junior High (n=103)	Senior High (n=129)
Sex				
Male	36.5	21.0	50.0	59.5
Female	63.5	79.0	50.0	40.5
Age Group				
20-29	21	24	18	19
30-39	40	37	44	46
40-49	24	23	26	24
50-59	13	14	12	14
60+	2	2	1	2
When were Tests & Measurements Taken?				
Undergraduate	94	97	93	90
Graduate	62	57	70	67
Inservice	51	49	57	53
Required? (% yes)	92	92	93	91
Median year of most recent T&M training	1974	1974	1974	1973

Table 2. Subscale and item statistics

Subscale - Item	Factor Loading	Item Mean*	SD	Item-Total Correlation	Subscale Alpha/n
Effects of standardized tests on students and instruction					
- Standardized tests serve no useful purpose.	.71	4.07	.97	.48	
- Standardized tests assess only unimportant educational outcomes.	.73	3.95	.88	.51	
- Standardized tests force teachers to "teach to the test".	.56	3.11	1.22	.54	.75 537-546
- Low scores on standardized tests damage a student's self-concept.	.63	3.19	1.00	.49	
- Standardized tests generate harmful anxiety in students.	.69	3.42	1.05	.57	
Use of standardized test results					
- Standardized tests are the best way to evaluate a teacher's effectiveness.	.68			.44	
- Teachers whose students score higher on standardized tests should receive higher salaries.	.69			.47	
- All districts in the state should be required to use the same standardized testing program.	.72			.38	.68 538-551
- Requiring students to pass competency tests would raise educational standards.	.68			.44	
- Requiring teachers to pass competency tests would raise educational standards.	.75			.48	
CLASSROOM TESTING					
Value of classroom tests					
- Tests are of little value in identifying learning problems.	.74	4.24	1.05	.35	
- Tests tend to penalize the brighter, more creative students.	.65	4.22	1.06	.37	.54 542-550
- Test construction takes up more time than it's worth.	.46	3.97	.88	.36	

Table 2 (cont'd)

Subscale - Item	Factor Loading	Item Mean	SD	Item-Total Correlation	Subscale Alpha/n
Fairness of classroom tests					
- Testing has a favorable impact on student motivation.	.51	3.07	1.02	.33	
- Tests are effective ways to direct student learning.	.63	2.99	.93	.43	
- Tests are of great value in communicating with parents about a student's progress.	.68	2.97	.95	.47	.63 541-548
- It is relatively easy to construct tests in my subject area.	.58	2.94	1.09	.34	
- Test scores are a fair way to grade students.	.68	3.35	.99	.35	
Effectiveness					
- Tests measure too many things besides knowledge of content.	.49	3.40	1.04	.32	
- Tests tend to discriminate against minority students.	.51	3.71	1.07	.34	
- Teachers' testing practices are often ineffective.	.59	3.20	.95	.45	
- Tests measure only superficial aspects of what students know and can do.	.63	3.24	1.00	.47	.68 535-545
- Tests tend to create too much anxiety in students.	.67	3.29	1.03	.40	
- Too many tests are given to students already.	.66	3.67	.96	.49	

Table 3. Subscale intercorrelations

Subscale:	Standardized Test Results - Use	Value	Fairness	Effectiveness
Standardized Test - Effects	.15	.29	.07	.39
Standardized Test Results - Use		.03	.03	.04
Value of Classroom Tests			.27	.42
Fairness of Classroom Tests				.17

Table 4. Validity coefficients*

Subscale	1**	2	3	4	5	6	7	8	9	10	11
Standardized Test Effects	.55	.56									
Standardized Test Results- Use	.27	.27									
Value of Classroom Tests			.28	.31	.07	.11	.06	.05	.17	.43	-
Fairness of Classroom Tests			.27	.25	.09	.14	.12	.12	.17	-	.35
Effectiveness of Classroom Tests			.32	.35	.13	.18	.07	.06	.19	.37	-
Total Attitude To Classroom Testing			.39	.42	.13	.19	.11	.10	.24	.51	.24

*Correlations greater than .07 are significant at $p<.05$; correlations greater than .10 are significant at $p<.01$.

**Variables are:

1 - Attitude toward the usefulness of standardized tests (single item)
2 - Disfavor/favor using standardized tests (single item)
3 - Attitude toward the usefulness of classroom tests (single item)
4 - Disfavor/favor using classroom tests (single item)
5 - Frequency of giving tests (single item)
6 - Percentage of students' grades based on tests (single item)
7 - Amount of time spent testing (single item)
8 - Use of suggested measurement techniques (scale)
9 - Personal experience with/liking for tests (scale)
10 - Own test results have been of value (single item)
11 - Tests taken were good assessment of knowledge (single item)

Table 5. Means and standard deviations for group differences in attitude scores

Variable	n	Attitude Subscales: Means (SD) Value	Effectiveness	Fairness	Stan-Effects	Stan-Results
Sex:						
Male	190	4.26	3.61	4.30	3.59	2.81
		(.73)	(.60)	(.54)	(.68)	(.86)
Female	324	4.08	3.30	4.15	3.51	2.71
		(.72)	(.62)	(.52)	(.76)	(.83)
		$p<.01$	$p<.01$	$p<.01$	-	-
Age:						
20-29	110	4.18	3.37	4.15	3.48	2.81
		(.70)	(.61)	(.51)	(.69)	(.81)
30-39	219	4.10	3.35	4.18	3.47	2.71
		(.68)	(.64)	(.50)	(.78)	(.83)
40-49	130	4.18	3.51	4.21	3.58	2.70
		(.81)	(.64)	(.60)	(.71)	(.76)
50+	79	4.19	3.55	4.34	3.76	2.84
		(.73)	(.57)	(.54)	(.62)	(1.01)
		-	$p<.02$	-	$p<.02$	-
Tests & Measurements Course?						
Yes	443	4.16	3.43	4.23	3.54	2.77
		(.74)	(.62)	(.53)	(.74)	(.83)
No	95	4.11	3.37	4.08	3.54	2.67
		(.63)	(.64)	(.47)	(.67)	(.86)
		-	-	$p<.05$	-	-
Grade Level Taught						
Elementary	285	4.09	3.33	4.16	3.59	2.75
		(.71)	(.58)	(.54)	(.71)	(.82)
Junior High	102	4.17	3.48	4.23	3.50	2.81
		(.73)	(.66)	(.49)	(.69)	(.86)
Senior High	124	4.25	3.56	4.30	3.46	2.72
		(.75)	(.65)	(.54)	(.82)	(.88)
		-	$p<.01$	$p<.06$	-	-