DOCUMENT RESUME

ED 379 326 TM 022 697

AUTHOR Marso, Ronald N.; Pigge, Fred L.
TITLE Testing Directors', Supervisors', Principals', and Teachers' Perceptions of the Effectiveness of Their Schools' Practices Related to the Management of Standardized Testing.
PUB DATE Apr 94
NOTE 19p.; Paper presented at the Annual Meeting of the National Council on Measurement in Education (New Orleans, LA, April 5-7, 1994).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150) -- Tests/Evaluation Instruments (160)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Administrator Attitudes; Beliefs; Comparative Analysis; Educational Administration; *Educational Practices; Evaluation Methods; *Principals; Rating Scales; School Districts; *Standardized Tests; *Supervisors; *Teacher Attitudes; Testing; Test Use
IDENTIFIERS *Test Directors

ABSTRACT

This study was designed to collect and then to compare various educators' perceptions (N=484) of the effectiveness of selected standardized testing program management practices in their schools. It was found that these educators perceived their schools' performance of testing-related practices to be more effective than the schools' overall performance in meeting other district responsibilities. Testing practices rated highest were use of quality tests and materials, maintenance of pupil records, and use of understandable scores and reports. Testing practices rated lowest were use of test results to evaluate instruction, availability of written policies, and use of publisher instructional guides accompanying achievement batteries. Comparatively, the teachers' ratings of the effectiveness of their schools' testing practices tended to be lower than were the administrators's ratings, and the testing directors' ratings of their testing practices tended to be lower than were the principals' and the supervisors' ratings. (Contains 12 references, 4 tables, and 4 figures.) (Author)

Testing Directors', Supervisors', Principals', and Teachers' Perceptions of the Effectiveness of Their Schools' Practices Related to the Management of Standardized Testing

Ronald N. Marso and Fred L. Pigge
Educational Foundations and Inquiry
Bowling Green State University
Bowling Green, Ohio 43403

A paper presented at the annual conference of the National Council for Measurement in Education

New Orleans, Louisiana
April 5-7, 1994

Abstract

This study was designed to collect and then to compare various educators' perceptions (N=484) of the effectiveness of selected standardized testing program management practices in their schools. It was found that these educators perceived their schools' performance of testing related practices to be more effective than their schools' overall performance in meeting other district responsibilities. Testing practices rated highest were use of quality tests and materials, maintenance of pupil records, and use of understandable scores and reports. Testing practices rated lowest were use of test results to evaluate instruction, availability of written policies, and use of publisher instructional guides accompanying achievement batteries. Comparatively, the teachers' ratings of the effectiveness of their schools' testing practices tended to be lower than were the administrators' ratings, and the testing directors' ratings of their schools' testing practices tended to be lower than were the principals' and the supervisors' ratings.

Testing Directors', Supervisors', Principals', and Teachers'
Perceptions of the Effectiveness of Their Schools' Practices Related
to the Management of Standardized Testing

Educators generally do not have a high regard for standardized testing despite the increased use of these tests in recent school reform efforts (Haney & Madaus, 1989). For example, many classroom teachers appear to have an unfavorable-indifferent attitude toward standardized testing (Borg, Worthen, & Valcarce, 1986), and school administrators tend to view standardized testing as being a relatively unimportant administrative function in their schools (Sproull & Zubrow, 1981). Additionally, assessments of the research literature reveal that testing and evaluation practices receive less attention from educational researchers than many other aspects of education (Crooks, 1988).

This less than positive regard for standardized testing is also revealed in what many educators believe about testing. Classroom teachers commonly believe that standardized testing skills are less needed than are other testing skills (Marso & Pigge, 1988); many teachers perceive the primary benefits of their school districts' standardized testing programs accrue not to themselves but to the school administration (Salmon-Cox, 1981); building principals typically do not perceive the need for testing specialists to be involved in the selection of standardized tests (Kinney, Brickell, & Lynn, 1988); and school counselors frequently feel testing services dominate too much of their time (Miller, 1977).

Furthermore, this less than positive attitude of educators toward standardized testing may be having an undesirable impact upon standardized testing practices in the K-12 schools. For example, many teachers report very limited use of the results from standardized testing in their classroom instruction (Linn, 1990), and educational administrators frequently do not convey the results from standardized testing to their teachers (Wood ,1982). Further curtailing the effective use of the results from standardized testing, the results of this testing, if made available, typically are not available to educational staff until six or eight or more weeks after test administration (Hall, Carroll, & Comer, 1988).

In brief, the existing research literature suggests that educators do not hold standardized testing in high regard, that limited management attention is typically provided for standardized testing activities, that the results of standardized testing may have little impact upon classroom instruction, and that testing and evaluation receive less research attention than many other aspects of education. In light of these research findings and the increased use of standardized tests in recent years, it appears prudent that we know more about typical testing practices in the K-12 schools and the effects of these testing practices upon the schooling process (Paris, Lawton, Turner, & Roth, 1991). The present study was designed to identify and then to compare and contrast various educators' perceptions of the effectiveness of selected practices related to the management of their schools' standardized testing programs. More specifically, this study was designed to answer the following types of questions. To what extent do testing directors, teacher supervisors, building principals, and classroom teachers perceive their schools' standardized testing practices to be effective? Do these various groups of educators differ one from the other in their perceptions of the effectiveness of these practices? Do groups of educators assigned to elementary and secondary grade levels differ one from the other in their perceptions of the effectiveness of these practices? To what extent do educators perceive that their schools' testing practices encourage the use of results from standardized tests in classroom instruction?

Methods and Procedures

The data gathered for this paper were one component of a larger state-wide assessment of the management and operation of standardized group testing programs in the K-12 public schools of Ohio. Each of the 616 nonvocational school districts were contacted regarding their willingness to participate in an extensive investigation of standardized testing practices and of the uses of standardized testing results. This inquiry resulted in 171 superintendents indicating a willingness to have their school districts participate in the study. From these 171 school districts, 106 districts were randomly selected

using type of administrative organizations (city, county local, and exempted village) of the school districts as strata in the selection process. Of these 106 randomly selected districts, 97 districts (92%) ultimately did participate in the study.

The survey assessment instruments were mailed directly to the participating superintendents who in turn were asked to forward the sealed packets of materials to selected teacher supervisors and elementary and secondary school principals. The criterion provided to the superintendents for these selections was that they select one of their elementary and one of their secondary principals and one of their teacher supervisors who would be most knowledgeable about and who could best inform the researchers about the practices and procedures of their school districts' standardized group testing program.

The elementary and secondary school principals receiving the survey packets from their superintendents, in addition to completing their own assessment of their district's testing practices, were also directed to select and forward enclosed survey materials to classroom teachers. The elementary principals were directed to select and to forward designated survey packets to one teacher assigned to grades one through four and to one teacher assigned to grades five or higher who were most knowledgeable about and who could best inform the researchers about the practices and procedures associated with their school district's standardized group testing program. The secondary principals were asked to follow these same procedures but were asked to select one teacher from the math-science and one from the English-social studies subject areas.

The goals of this subject selection procedure were first, to solicit assessment responses just from those educators knowledgeable about their districts' testing practices and secondly, to insure responses from educators who were representative of the instructional, administrative, grade, and subject diversities found in the K-12 schools. The variety of instructional and administrative responsibilities of educational personnel within school districts as well as the variations in testing practices from school district to school district result in considerable diversity in the extent of experiences a particular educator might have with his/her school's standardized testing program. For example, as a consequence of test scheduling decisions, standardized tests might not be scheduled in fourth, sixth, and eleventh grades in a particular school district over a period of years. Consequently, teachers at these grade levels may have few direct experiences with standardized testing; whereas their cohorts assigned to other grades may have frequent and direct involvement with their school's standardized testing program.

These subject selection and contact procedures resulted in the return of usable survey assessment forms from 82 testing directors, 155 principals, 47 supervisors of teachers, and 200 classroom teachers. Just those individuals designated as testing directors by their superintendent and who, themselves, acknowledged that title were included in the testing directors group, and just those teacher supervisors employed by the selected school districts were included in the supervisors group. Several school superintendents reported either that no formal teacher supervisor positions existed in their district or that teacher supervisory services were provided through their county offices of education. These respondents were employed in schools organized by city district (42%), local county district (44%), and exempted village district (14%), in schools located in geographic settings described as rural (37%), suburban (57%), and urban (6%), and in small schools (11% with fewer than 1,000 pupils), moderately sized schools (34% with 1,000 to 2,000 pupils), moderately large schools (34% with 2,001 to 4,000 pupils), and large schools (21% with more than 4,000 pupils). These proportions of respondents representing different types of school settings were judged to be approximately similar to the composition of all such schools as reported in the Ohio Education Directory.

The focus of the present report is upon these educators' responses to 10 survey items related to their school district's practices associated with the management of standardized testing. They responded to each of the 10 testing practices by rating the "relative effectiveness" of their school

district's testing practices or procedures during the past year or two. This latter reference to time was provided to give the educators a common reference period for their ratings.

In rating the relative effectiveness of their school district's testing practices and uses, the educators also were provided with a common comparative rating reference by being directed to rate their school's effectiveness on each of the 10 items in terms of their perceptions of their school's performance of each practice compared to their school's overall performance as an educational institution. A five-point scale with narrative descriptions at each scale point and with an accompanying "DK" response option, defined as "I really do not know," was provided with each of the 10 testing practices items. This scale ranged from we perform well below our average '1' to we excel '5'. The 10 testing practices items and the scale response codes are presented in the appendix.

One- and two-way ANOVA procedures were used to identify possible rating mean differences (with alpha level adjusted via the Bonferroni approach and post-hoc pair wise tests done by the Scheffe procedure at $\alpha \leq .10$) among the groups of educators. The one-way ANOVA procedures were used to identify significant ($\alpha \leq .05$) rating mean differences among the teacher, principal, supervisor, and test director respondent groups. One-way ANOVA procedures were used also to examine possible differences among the directors and the other three groups when the latter three groups included first, just those with secondary school assignments and second, just those with elementary school assignments. The two-way ANOVA procedures were used to identify significant ($\alpha \leq .05$) differences among the ratings of the teacher, supervisor, and principal groups when each was classified by assignment to either elementary or secondary schools. These ANOVA procedures were completed on the data derived from respondent ratings of each of the 10 testing practices.

Findings

Each of the four groups of educators, testing directors, teachers, supervisors, and principals, rated their school's performance of the selected 10 testing practices about average or somewhat higher (3 or higher on the five-point scale) compared to their school's performance in meeting their overall responsibilities as educational institutions (see Table 1). Only when the teachers, principals, and supervisors were classified by elementary and secondary school assignments were any rating means found below the "about average performance for us" or '3' level. Two of the rating means of the secondary teachers, one of the rating means of the secondary supervisors, and none of the mean ratings of the testing directors and the teachers with elementary school assignments were below the "about average" or '3' level (see Tables 2 and 3).

Insert Tables 1, 2, and 3 about here

More explicitly, the testing practices rated relatively more effective (see Table 1) by the total groups of educators were management of pupil records, use of quality tests and materials, selection and use of tests, and use of understandable scores and reports (items 8, 3, 1, and 5, respectively). Practices rated relatively less effective were use of achievement battery results to evaluate district classroom instruction, provision of instructional guides to relate test results to instruction, availability of school policy regarding access/dissemination/ storage of test results, and prompt availability of test results after testing (items 10, 6, 7, and 4, respectively).

Used descriptively, the one-way ANOVA procedures revealed group differences in ratings of the effectiveness of each of the 10 testing management practices for the combined elementary and secondary educators (see Table 1), the directors and elementary educators (see Table 2), the directors and secondary educators (see Table 3). The combined, elementary, or secondary groups of teachers' rating means differed significantly between one or more of the groups of educators on each of the 10 items. The testing directors' ratings of the effectiveness of district testing practices differed from the

combined, elementary, or secondary groups of teachers' ratings of eight of the testing practices; the principals' and teachers' ratings differed from the combined, elementary, or secondary groups on eight practices; and the supervisors' ratings differed on two testing practices from the teachers' ratings for the combined, elementary, or secondary groups. No significant differences were identified between any of the rating means for any of the possible mean pairs for the administrator groups of principals, directors, and supervisors.

More specifically, the elementary and secondary teachers as a collective group (Table 1) rated test selection/administration/scheduling (item 1) lower than did the supervisors, principals, and directors; rated test scheduling to aid decision-making and prompt testing results (items 2 and 4) lower than did the supervisors and directors; but they rated the provision of criterion-referenced data (item 9) higher than did the testing directors. When the directors' ratings along with teachers', principals' and supervisors' ratings who had elementary school assignments were analyzed (Table 2), the teachers' ratings were lower than the directors' ratings of practices related to test selection-administration, test scheduling, and prompt return of test results (items 1, 2, and 4), but the teachers' ratings were higher than the directors' ratings of the provision of criterion-referenced data and handling of pupil permanent records (items 8 and 9) and higher than were the directors' and principals' ratings of the provision of instructional guides and the availability of written school policies regarding pupil records (items 6 and 7).

The analysis of the ratings of the directors, along with the secondary (Table 3) principals, teachers, and supervisors, revealed that the secondary teachers' ratings were higher than the secondary principals' ratings of the use of understandable scores and reports and of the use of achievement batteries to evaluate district instruction (items 5 and 10), but the secondary teachers' ratings were lower than the directors' and principals' rating of the practices of test selection-administration, test scheduling, test and materials quality, and promptness of test results (items 1, 2, 3, and 4) and lower than were the directors' and supervisors' ratings for the provision of instructional guides to aid instruction (item 6).

Omitting the testing directors and with the use of the job assignment classification of elementary or secondary school assignment in the two-way ANOVA procedures revealed that the elementary educators (combined principals, supervisors and teachers) rated higher the practices of using understandable scores and reports, the provision of instructional guides for instruction, and use of scores for evaluation of district instruction (items 5, 6, and 10) than did their secondary cohorts.

Insert Table 4 about here

The two-way ANOVA procedures also revealed significant job-group and grade-level interactions for items five, six, seven, and nine. The mean patterns for each of these four interactions indicate a greater discrepancy between the ratings of the elementary and secondary teachers compared to the discrepancies between the elementary and secondary supervisors and principals. For each of these four testing practices, understandable scores and reports, availability of instructional guides, presence of school policies, and provision of criterion-referenced test data, the secondary teachers' ratings were sharply lower than were the elementary teachers' ratings (see Figures 1, 2, 3, and 4). Additionally, the ratings of the elementary and secondary supervisors differed sharply regarding the effectiveness of the provision of criterion-referenced achievement battery results as well as norm referenced scores. Elementary supervisors rated this practice much lower than did their secondary cohorts as well as much lower than the elementary teachers.

Insert Figures 1-4 about here

Summary and Discussion

Somewhat contrary to what might be expected from the studies revealing that educators tend to have a less than positive attitude toward standardized testing (Haney & Madaus, 1989) and that school administrators view standardized testing as a relatively unimportant management function (Sproull & Zubrow, 1981), these teachers, supervisors, principals, and testing directors rated the performance of their schools relative to each of the 10 selected standardized testing program practices as being average or above average in comparison to their schools' performance in meeting their overall responsibilities as institutions. This contrast between present and prior findings may in part be due to the selection of the educators for this study; for at each distribution level of the survey materials, superintendents and principals, the distributor was directed to forward the materials to individuals most knowledgeable about and who could best inform the researchers about their school's standardized testing program. Additionally, the respondents were provided with, and many used, a "do not know" category on the rating side. These research procedures were used to increase the assurance that the raters were knowledgeable about their school's testing practices; consequently, the findings from this study are likely to more accurately reflect the performance levels of the schools, but the present findings may not be as representative of less knowledgeable or typical educators' perceptions of the effectiveness of their schools' testing practices.

The high relative consistency between the ratings (Rho's in .90's, see Table 1) of the principals, supervisors, and testing directors, adds further confidence to the interpretation that these findings accurately reflect the perceptions of those educational administrators most knowledgeable of testing practices in Ohio schools. The frequent differences found between the magnitudes of the ratings of the teachers and the three administrative groups of educators probably reflect differences between their job experiences and responsibilities. For example, the supervisors', with responsibilities more predominantly linked to classroom instruction as compared to principals and testing directors, rating means were much more similar in a relative sense to those of the classroom teachers than were the rating means of the principals and the testing directors. Just two testing practices revealed differences between the mean ratings of the teachers and supervisors; whereas eight practices revealed differences between teachers and either the principals or the directors. Also relative to the teachers' ratings, there was less consistency found between the ratings of the elementary teachers and the various administrator groups (Rho = .49 to .60, see Table 2) than was found between the ratings of the secondary teachers and the various administrator groups (Rho = .87 to .94, see Table 3). This discrepancy might reflect the more similar background of the secondary teachers and the testing directors than between the elementary teachers and the testing directors as most of the directors either had prior assignments in or concurrently held assignments in secondary schools.

The teachers tended to rate the effectiveness of their schools' testing practices lower than did the three groups of administrators. Few differences were found between the ratings of the testing directors, supervisors, and the principals. Also, few differences were identified between the respondents when grouped as secondary and elementary educators, and when these few differences were noted they were limited primarily to teachers' ratings.

The main effect differences noted between the educators assigned to elementary and the secondary schools centered around the availability and interpretability of scores and instructional related uses of tests (items 5, 6, 7, and 9) and seemed to be precipitated primarily by differences in the ratings of teachers (see Figures 1-4). These rating differences between the elementary and secondary teachers may simply reflect the differences between the testing practices and the focus of standardized testing in the elementary and secondary grades. In the elementary grades, the typical goal of standardized testing focuses upon the guidance of pupil instruction with frequent scheduling of achievement batteries and scholastic aptitude tests; whereas in the secondary grades standardized testing is typically focused upon the goal of pupil career and academic counseling with scheduling of aptitude batteries, vocational interest inventories, and college admission tests.

Similarly, the statistical interaction effects identified in the present study might also be explained by this difference in the focus of standardized testing between the secondary and elementary levels. For example, the nature of score reports, the practices related to the retention of testing records, availability of test remedial instruction guides, and the provision of criterion-referenced data are likely to vary considerably between the elementary and secondary grades. The elementary grade test reports are likely to be less complex; remedial instruction guides would not typically be available for tests administered at the secondary level; score records are more likely to be readily available to elementary teachers and frequently are stored within the self-contained classroom, in contrast test scores are generally less readily available to secondary teachers due to multiple class group responsibilities and test results are typically stored in a central location; and typically only the achievement batteries which are most frequently administered in the elementary grades would likely provide criterion-referenced data.

The pattern of high and low rating means across the 10 testing practices noted in the present study suggests possible implications for the management of standardized testing programs. Certainly, first and foremost, the ratings of these educators suggest that standardized testing programs, at least in Ohio, are perceived to be well managed as compared to the effectiveness of the overall performance of the school districts participating in this study. Each of the groups of educators in the present study appeared to be satisfied with the quality of the tests, materials, report forms, and management of pupil records, but they appeared to be skeptical about the effectiveness of the use of achievement scores to evaluate classroom instruction. The teachers, as compared to the administrators, appeared to be less satisfied with test selection, administration and scheduling, and the prompt availability of the results from testing. Conversely, elementary school teachers appeared to be more satisfied with the effectiveness of the guides for remedial instruction and criterion-referenced data accompanying achievement batteries than were the other groups of educators.

The wise testing directors might prudently build on the present satisfactions of their educational cohorts but strive to work more closely with classroom teachers. In particular, it appears that these testing directors, and perhaps other educational administrators, ought to work more closely with teachers in the selection, administration, and scheduling of tests, in the dissemination of test results, in preparing written testing policy, and in making remedial instruction guides accompanying achievement batteries available to facilitate the use of test results in the instructional process. The present findings also suggest that collaborative efforts might be more essential in the elementary as compared to the secondary schools. Lastly, it would seem to behoove testing directors to investigate the major discrepancy identified between elementary teachers' and elementary supervisors' perceptions of the effectiveness of the use of achievement battery criterion-referenced data in linking testing results with instructional planning. The elementary supervisors in the present study rated the provision of criterion-referenced data from achievement batteries as being less effective than did the teachers, principals, and secondary supervisors. Measurement specialists would anticipate that elementary supervisors would be the strongest advocates of the provision of achievement test criterion-referenced data due to its potential close link with instructional planning.

M:2/6

References

Borg, W. R., Worthen, B. R., & Valcarce, R. W. (1986). Teachers' perceptions of the importance of educational measurement. *Journal of Experimental Education*, 55(1), 9-14.

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

Hall, B. W., Carroll, D., & Comer, C. B. (1988). Test use among classroom teachers and its relationship to teaching level and teaching practices. *Applied Measurement in Education*, 2, 145-156.

Haney, W., & Madaus, G. (1989). Searching for alternatives to standardized tests: Whys, whats, and withers. *Phi Delta Kappan*, 70, 683-688.

Kinney, D., Brickell, J., & Lynn, M. A. (1988). Testing programs in Illinois elementary school districts: The relationships between standardized achievement, competency-based, and teacher-made tests. Paper presented at the American Education Research Association annual meeting, New Orleans.

Linn, R. L. (1990). Essentials of student assessment: From accountability to instructional aid. *Teachers College Record*, 91(3), 422-436.

Marso, R. N., & Pigge, F. L. (1988). Standardized and competency testing programs: Classroom teachers' needs and proficiencies as viewed by principals and supervisors. (ERIC Document Reproductive Service No. ED TM 012854).

Miller, G. M. (1977). After testing is over. *Elementary School Guidance and Counseling*, December, 139-143.

Paris, S. G., Lawton, T. A., Turner, J. C., & Roth, J. L. (1991). A developmental perspective on standardized achievement testing. *Educational Researchers*, 20(5), 15-20.

Salmon-Cox, L. (1981, May). Teachers and standardized achievement tests: What's really happening. *Phi Delta Kappan*, 631-634.

Sproull, L., & Zubrow, D. (1981). Standardized testing from the administrative perspective. *Phi Delta Kappan*, 62, 631-794.

Wood, C. M. (1982). The role of standardized achievement tests in the management of instruction: A survey of teacher and administrator practices and attitudes. (ERIC Document Reproductive Service No. ED 219 446).

M:2/6

Table 1

Analyses of mean differences between testing directors', teachers', principals' and supervisors' ratings of the effectiveness of testing practices.

	(1) Directors			(2) Teachers			(3) Principals			(4) Supervisors				
Practices	N	M	SD	N	M	SD	N	M	SD	N	M	SD	F	p
1. Test selection & administration	77	4.01 A	.75	193	3.56 B	.99	151	3.92 A	.80	47	4.00 A	1.00	7.75	.00
2. Test scheduling to aid decisions	80	3.90 A	.98	188	3.40 B	1.15	152	3.77 A	.91	46	3.70 A,B	1.11	5.66	.00
3. Quality tests, materials, reports	81	4.17 A	.89	194	3.96 A	.94	151	4.17 A	.86	47	4.04 A	1.14	1.80	.15
4. Results available promptly	80	3.70 A	.99	200	3.15 B	1.18	152	3.48 A	.99	47	3.43 A,B	1.04	5.87	.00
5. Understandable scores, reports	80	3.98 A	.87	196	3.76 A	1.04	155	3.89 A	.93	43	3.77 A	1.07	1.21	.31
6. Instructional guides available	74	3.20 A	1.12	195	3.23 A	1.49	148	3.09 A	1.32	42	3.40 A	1.25	.67	.57
7. Written test records policies	72	3.10 A	1.15	129	3.41 A	1.29	141	3.18 A	1.28	40	3.58 A	1.17	2.05	.11
8. Student permanent records	78	4.03 A	1.02	198	4.26 A	1.01	154	4.29 A	.88	43	4.23 A	.87	1.46	.23
9. 'CR' available data	73	3.29 B	1.16	149	3.86 A	1.16	136	3.68 A,B	1.12	39	3.59 A,B	1.25	4.05	.01
10. Evaluate class instruction	71	3.08 A	1.14	148	3.07 A	1.19	146	3.25 A	1.15	42	3.05 A	1.38	.68	.57

* Unique letters indicate significant mean differences, similar letters indicate nonsignificant mean differences, Scheffe $\alpha = .10$. N's varied as respondents were provided with the option of "I really do not know" in rating each practice.

** $Rho_{1,2} = .73$, $Rho_{1,3} = .93$, $Rho_{1,4} = .93$, $Rho_{2,3} = .77$, $Rho_{2,4} = .88$, $Rho_{3,4} = .94$

Table 2

Analyses of mean differences between testing directors', elementary teachers', elementary principals' and elementary supervisors' ratings of the effectiveness of testing practices.

	(1) Directors			(2) Teachers			(3) Principals			(4) Supervisors				
Practices	N	M	SD	N	M	SD	N	M	SD	N	M	SD	F	p
1. Test selection & administration	77	4.01 A	.75	112	3.57 B	.90	76	3.86 A,B	.86	17	4.00 A,B	1.22	4.48	.00
2. Test scheduling to aid decisions	80	3.90 A	.98	106	3.40 B	1.14	77	3.68 A,B	1.01	17	3.53 A,B	1.18	3.53	.02
3. Quality tests, materials, reports	81	4.17 A	.89	113	4.05 A	.91	76	4.08 A	.96	17	4.12 A	1.27	..27	.85
4. Results available promptly	80	3.70 A	.99	117	3.13 B	1.29	76	3.32 A,B	1.09	17	3.41 A,B	1.06	3.98	.01
5. Understandable scores, reports	80	3.98 A	.87	116	4.02 A	.97	79	3.90 A	.97	14	3.79 A	1.12	.41	.74
6. Instructional guides available	74	3.20 B	1.12	112	3.79 A	1.28	77	3.23 B	1.35	16	3.31 A,B	1.40	4.51	.00
7. Written test records policies	72	3.10 B	1.15	74	3.65 A	1.21	73	3.00 B	1.33	15	3.60 A,B	1.12	4.36	.01
8. Student permanent records	78	4.03 B	1.02	112	4.41 A	.94	79	4.37 A,B	.83	14	4.21 A,B	.89	2.91	.04
9. 'CR' available data	73	3.29 B	1.16	92	4.05 A	1.15	69	3.62 A,B	1.19	12	3.42 A,B	1.44	6.03	.00
10. Evaluate class instruction	71	3.08 A	1.14	82	3.32 A	1.16	73	3.26 A	1.16	13	3.31 A	1.49	.55	.65

* Unique letters indicate significant mean differences, similar letters indicate nonsignificant mean differences, Scheffe $\alpha = .10$. N's varied as respondents were provided with the option of "I really do not know" in rating each practice.

** $Rho_{1,2} = .494$, $Rho_{1,3} = .927$, $Rho_{1,4} = .545$, $Rho_{2,3} = .553$, $Rho_{2,4} = .597$, $Rho_{3,4} = .802$

Table 3

Analyses of mean differences between testing directors', secondary teachers', secondary principals' and secondary supervisors' ratings of the effectiveness of testing practices.

	(1) Directors			(2) Teachers			(3) Principals			(4) Supervisors				
Practices	N	M	SD	N	M	SD	N	M	SD	N	M	SD	F	p
1. Test selection & administration	77	4.01 A	.75	78	3.51 B	1.10	62	3.97 A	.70	12	3.92 A,B	1.00	4.94	.00
2. Test scheduling to aid decisions	80	3.90 A	.98	78	3.36 B	1.17	63	3.87 A	.81	11	3.91 A,B	1.14	4.72	.00
3. Quality tests, materials, reports	81	4.17 A	.89	77	3.81 B	.96	63	4.25 A	.76	12	4.25 A,B	1.14	3.64	.01
4. Results available promptly	80	3.70 A	.99	79	3.14 B	1.01	63	3.67 A	.88	12	3.25 A,B	1.06	5.62	.00
5. Understandable scores, reports	80	3.98 A	.87	77	3.32 B	1.02	63	3.92 A	.92	12	3.75 A,B	1.06	7.37	.00
6. Instructional guides available	74	3.20 A	1.12	79	2.42 B	1.38	59	2.92 A,B	1.21	10	3.60 A	1.35	6.33	.00
7. Written test records policies	72	3.10 A	1.15	51	3.00 A	1.34	57	3.32 A	1.17	12	3.67 A	1.44	1.33	.27
8. Student permanent records	78	4.03 A	1.02	83	4.05 A	1.07	62	4.21 A	.85	12	4.08 A	1.00	.45	.71
9. 'CR' available data	73	3.29 A	1.16	54	3.50 A	1.13	56	3.71 A	1.02	11	4.00 A	1.18	2.31	.03
10. Evaluate class instruction	71	3.08 A,B	1.14	63	2.78 B	1.16	60	3.32 A	1.14	12	2.67 A,B	1.56	2.60	.05

* Unique letters indicate significant mean differences, similar letters indicate nonsignificant mean differences, Scheffe $\alpha = .10$. N's varied as respondents were provided with the option of "I really do not know" in rating each practice.

** $Rho_{1,2} = .87$, $Rho_{1,3} = .95$, $Rho_{1,4} = .79$, $Rho_{2,3} = .94$, $Rho_{2,4} = .92$, $Rho_{3,4} = .84$

Table 4

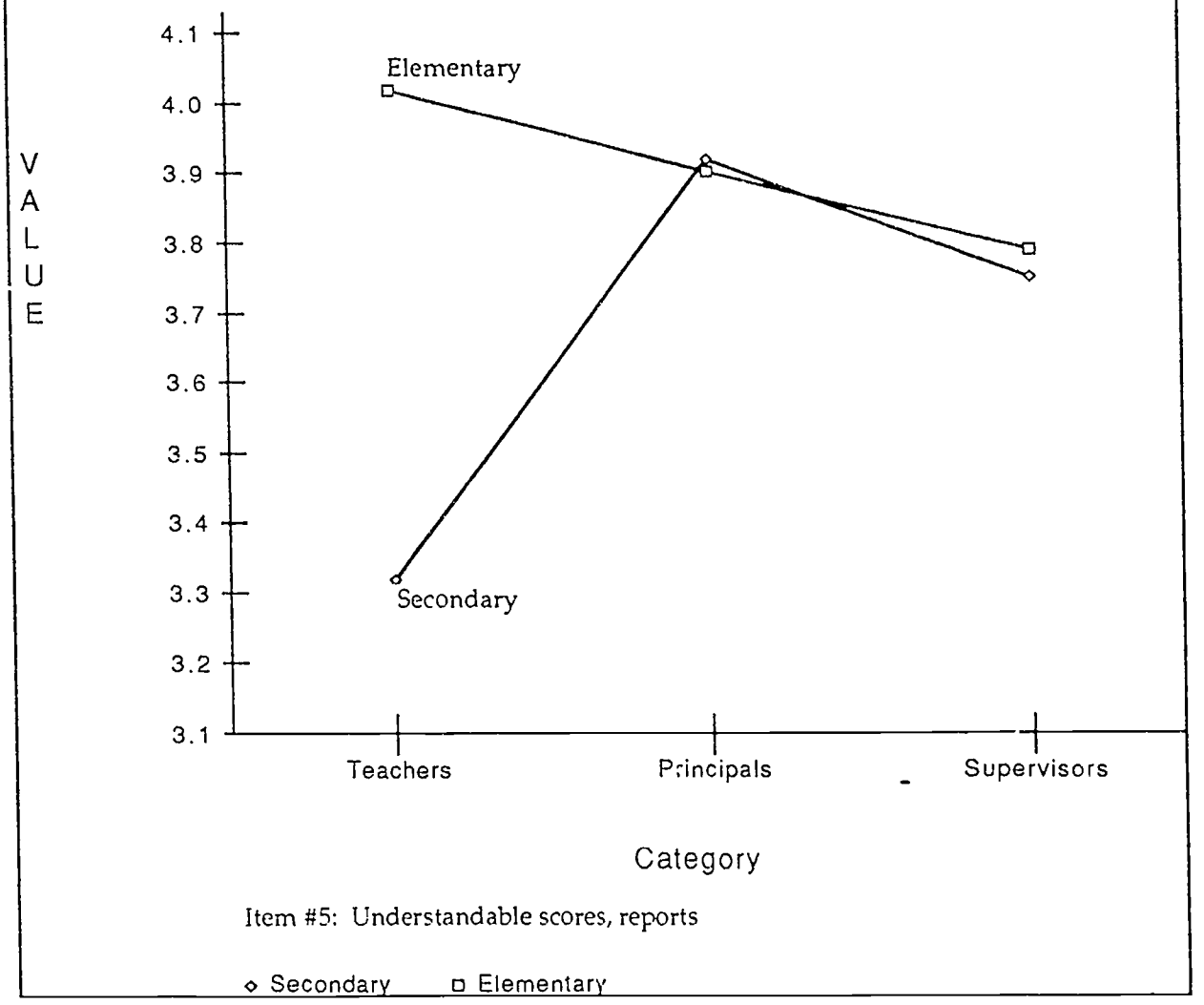
Analysis of teachers', principals', and supervisors' ratings of the effectiveness of testing practices when classified by elementary and secondary grade levels.

Practice or Procedure	Means (N's) Job Assignments Teach	Princ	Supr	F	p	Means (N's) Grade Level Elem.	Sec.	F	p	Job x Grade F	p
1. Test selection & administration	(190) 3.55 B	(138) 3.91 A	(29) 3.97 A	7.19	.00	(205) 3.71 A	(152) 3.73 A	.01	.94	.37	.69
2. Test scheduling to aid decisions	(184) 3.38 A	(140) 3.67 A	(28) 3.68 A	5.68	.00	(200) 3.52 A	(152) 3.61 A	1.27	.26	.74	.48
3. Quality tests, materials, reports	(190) 3.95 A	(139) 4.16 A	(29) 4.17 A	2.86	.06	(206) 4.07 A	(152) 4.03 A	.02	.89	2.15	.12
4. Results available promptly	(196) 3.13 A	(139) 3.47 A	(29) 3.34 A	4.17	.02	(210) 3.22 A	(154) 3.36 A	.17	.68	1.20	.30
5. Understandable scores, reports	(193) 3.74 A	(142) 3.91 A	(26) 3.77 A	2.37	.10	(209) 3.96 B	(152) 3.61 A	2.53	.11	5.67	.00
6. Instructional guides available	(191) 3.23 A	(136) 3.10 A	(26) 3.42 A	.90	.41	(205) 3.55 B	(148) 2.70 A	5.36	.02	8.75	.00
7. Written test records policies	(125) 3.38 A	(130) 3.14 A	(27) 3.63 A	1.71	.18	(162) 3.35 A	(120) 3.22 A	.21	.65	4.61	.01
8. Student permanent records	(195) 4.26 A	(141) 4.30 A	(26) 4.15 A	.31	.73	(205) 4.38 A	(157) 4.11 A	2.36	.13	.55	.58
9. 'CR' available data	(146) 3.85 A	(125) 3.66 A	(23) 3.70 A	.29	.75	(173) 3.84 A	(121) 3.64 A	.05	.83	3.95	.02
10. Evaluate class instruction	(145) 3.08 A	(133) 3.29 A	(25) 3.00 A	1.66	.19	(168) 3.29 A	(135) 3.01 B	4.09	.04	2.46	.09

* Unique letters indicate significant mean differences, similar letters indicate nonsignificant mean differences, Scheffe $p < .10$. N's varied as respondents were provided with the option of "I really do not know" in rating each practice.

** Rho between ranks of elementary and secondary personnel rating means = .77.

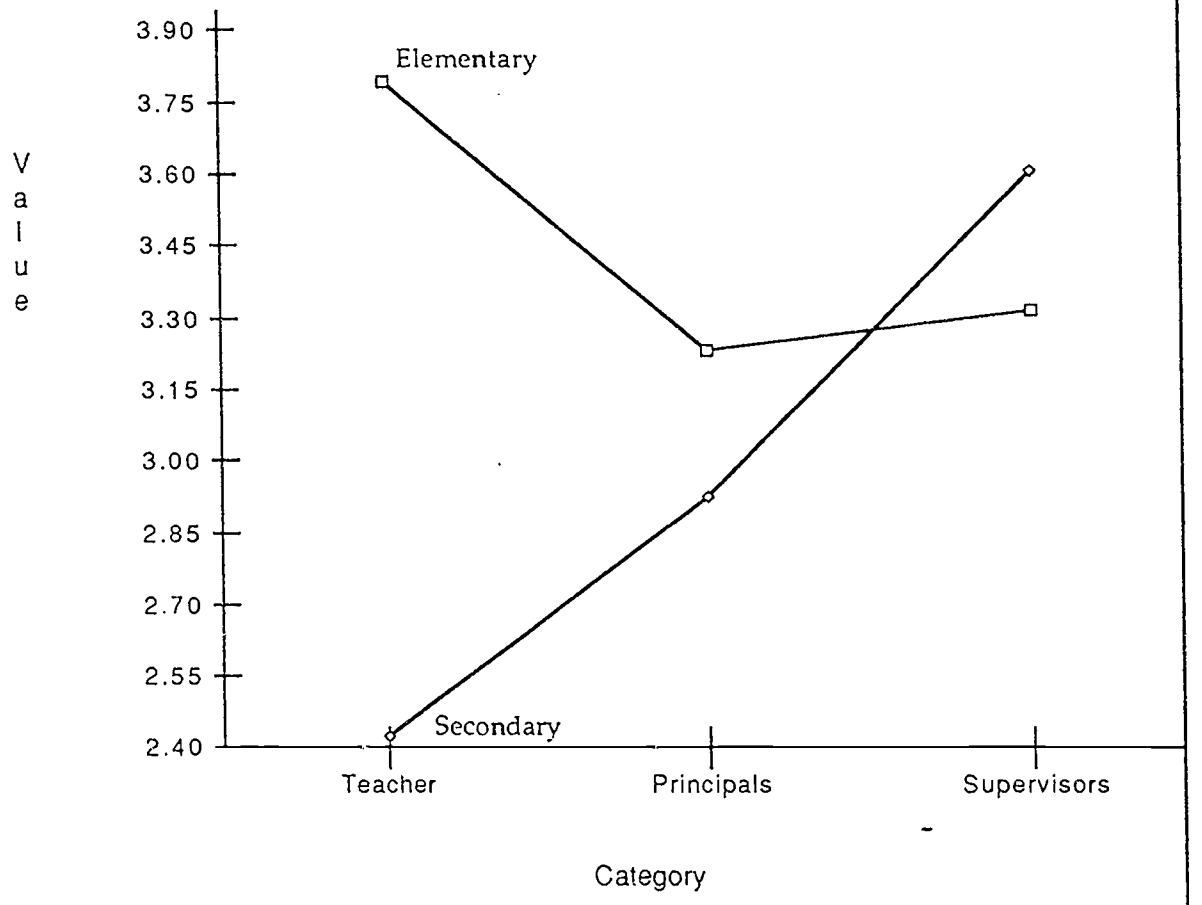
Figure 1



Item #5: Understandable scores, reports

◇ Secondary □ Elementary

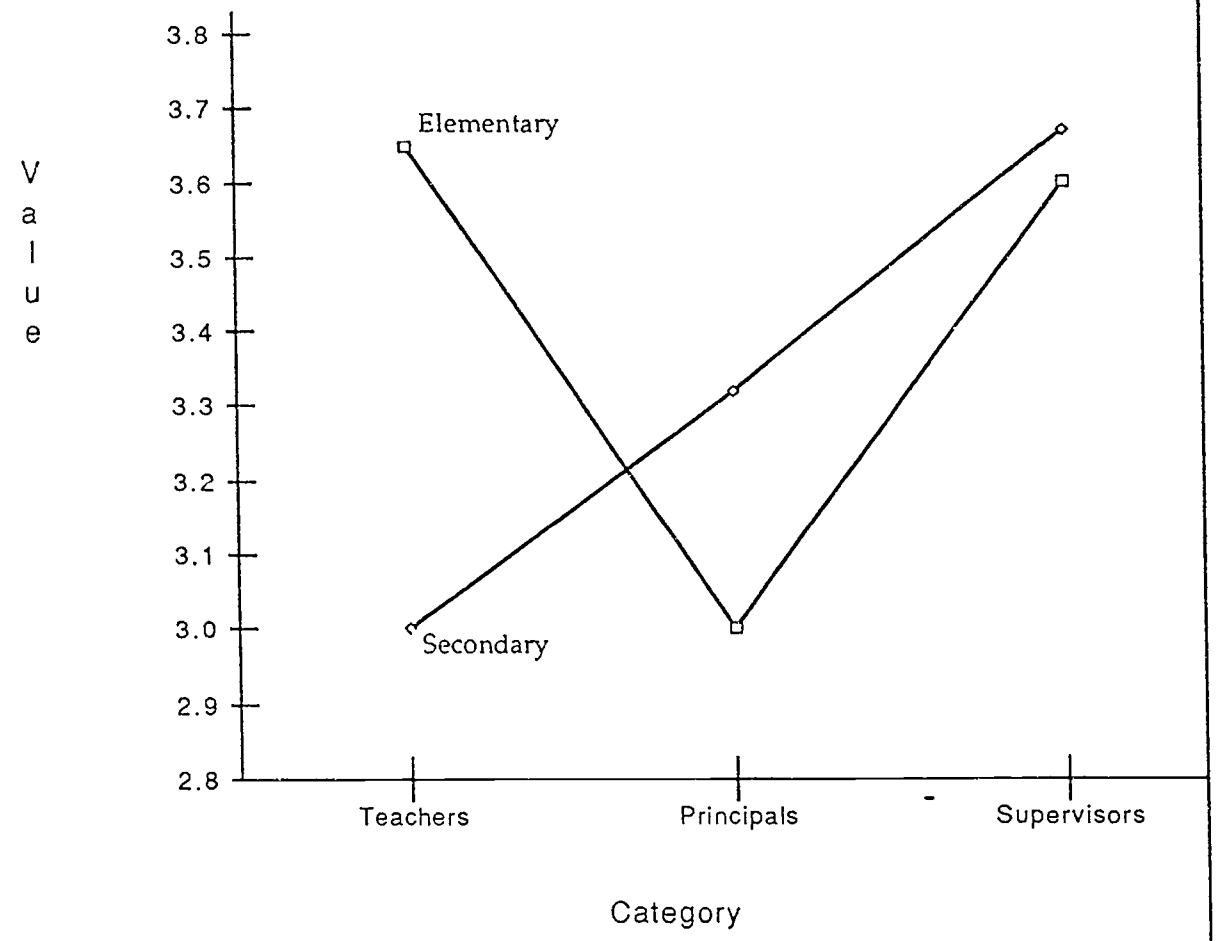
Figure 2



Item #6: Guides for remedial instruction available

◇ Secondary □ Elementary

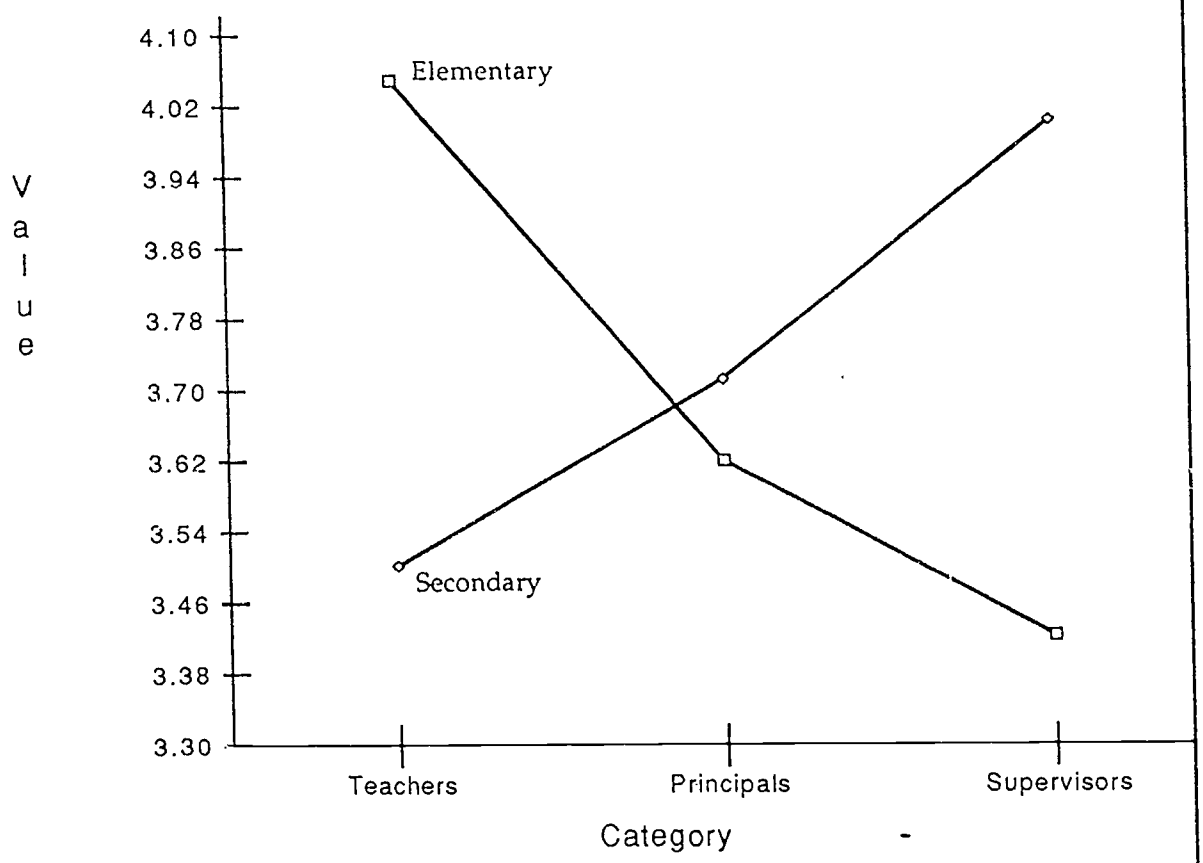
Figure 3



Item #7: Written policies related to test records

◇ Secondary □ Elementary

Figure 4



Item #9: Criterion-referenced data available

◇ Secondary □ Elementary

Appendix

Survey Instrument

SECTION IV. School Standardized Group Testing Program Practices or Procedures.

Please rate each of the following group testing practices or procedures in terms of the relative effectiveness of what happens in your school(s). Please respond to each item the best you can although you may be more or less informed about some of these practices. Please circle your rating of effectiveness using the code below.

		Relative Effectiveness*
Response Codes	'1'	We perform well below our average* here
	'2'	We perform below our average here
	'3'	About average performance for us
	'4'	We perform somewhat above average here
	'5'	We excel here
	'DK'	I really do not know

* Your perception of your school's (s') performance on this practice relative to its overall performance as an educational institution.

	Practice or Procedure	Relative Effectiveness					
		low				high	(?)
1.	Effective test selection/administration/scheduling for standardized testing program (overall)	1	2	3	4	5	DK
2.	Tests are scheduled at times to aid decision-making	1	2	3	4	5	DK
3	Quality tests, materials, and reports are used	1	2	3	4	5	DK
4.	Results of tests are available promptly to aid use of results	1	2	3	4	5	DK
5.	Understandable scores, narrative reports and pupil profiles are used to report performance	1	2	3	4	5	DK
6.	Teachers' instructional guides are made available to all teachers to aid instructional use of achievement battery results	1	2	3	4	5	DK
7.	Written school policies are available for access/dissemination/ storage of test results	1	2	3	4	5	DK
8.	Student permanent records are updated periodically (dated information removed, new added, etc.)	1	2	3	4	5	DK
9.	Criterion-referenced achievement battery results are provided as well as norm-referenced scores	1	2	3	4	5	DK
10.	Achievement battery scores are used in part to evaluate district classroom instruction	1	2	3	4	5	DK