A STUDY OF STANDARDIZED TEST KNOWLEDGE
AND INTERPRETATION BY ELEMENTARY
CLASSROOM TEACHERS

DISSERTATION

Presented to the Graduate Council of the
North Texas State University in Partial
Fulfillment of the Requirements

For the Degree of

DOCTOR OF EDUCATION

By

Elouise G. Pugh, B.S., M.Ed.
Denton, Texas
December, 1980
Pugh, Elouise G., *A Study of Standardized Test Knowledge and Interpretation by Elementary Classroom Teachers*. Doctor of Education (Administrative Leadership), December, 1980, 80 pp., 8 tables, bibliography, 56 titles.

This study surveys a sample of second-, fourth-, and sixth-grade teachers' proficiencies in three areas of standardized achievement test information—knowledge of standardized test terminology, interpretation of standardized tests, and application of standardized test results in program planning. A comparison is also made of teacher knowledge of standardized tests and public school administrators' expectations of teachers' skills.

The purposes of the study were to determine elementary teachers' knowledge of standardized achievement test terminology, interpretation and application of test score information. This determination was made across the three variables, grade level taught, highest degree earned, and the number of years teaching experience.

The researcher-developed instrument, "Standardized Test Proficiencies for Elementary Teachers", was administered to 122 teachers in two metroplex, suburban school districts. The population included forty-three second-grade teachers, thrity-nine fourth-grade teachers, and forty sixth-grade teachers.
A researcher-developed semantic differential test was also administered to the participants to determine teacher attitude toward standardized achievement tests in the elementary school.

Fifteen hypotheses were tested across the three variables—grade level taught, years teaching experience, and highest degree earned. Hypotheses I, V, and IX predicted there would be no statistically significant difference in knowledge scores across the three variables. Hypotheses II, VI, and X predicted there would be no statistically significant difference in interpretation scores across the three variables. Hypotheses III, VII, and XI predicted there would be no statistically significant difference in application scores across the three variables. Hypotheses IV, VIII, and XII predicted there would be no statistically significant difference in total scores across the three variables. Hypotheses XIII, XIV, and XV predicted there would be no statistically significant difference in attitude scores across the three variables.

The analysis of the data revealed there was no statistically significant differences among the mean scores in any of the stated hypotheses. None of the null hypotheses were rejected.

The findings in the study demonstrated that teacher knowledge, interpretation, application, and attitude toward
standardized achievement tests are not significantly influenced by grade level taught, years teaching experience, nor highest degree earned. It is also shown that elementary teachers' knowledge and use of standardized tests are far below the expectations held by public school administrators.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>LIST OF TABLES</th>
<th>v</th>
</tr>
</thead>
<tbody>
<tr>
<td>LIST OF ILLUSTRATIONS</td>
<td>vi</td>
</tr>
</tbody>
</table>

Chapter

I. INTRODUCTION 1

- Statement of Problem
- Purposes of Study
- Hypotheses
- Background and Significance of Study
- Definitions
- Basic Assumption
- Delimitation
- Procedures for Collection of Data
- Procedures for Analysis of Data

II. REVIEW OF RELATED LITERATURE 17

- Use of Standardized Tests
- Teacher Attitude
- Curriculum Effects
- Standardized Achievement Test Practices

III. METHODS AND PROCEDURES FOR COLLECTION OF DATA 37

- Description of the Population
- Construction of Initial Skills Inventory
- Statistical Procedures

IV. PRESENTATION OF DATA 44

- Analysis of the Differences in Mean Scores of Grade Level Taught, Years Teaching Experience, and Highest Degree Earned Versus Knowledge of Standardized Tests
- Analysis of the Difference in Mean Scores of Grade Level Taught, Years Teaching Experience and Highest Degree Earned Versus Interpretation of Standardized Tests
<table>
<thead>
<tr>
<th>Table</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>I.</td>
<td>Analysis of Variance for Teacher Knowledge Scores on the &quot;Standardized Test Proficiencies for Elementary Teachers&quot;</td>
<td>45</td>
</tr>
<tr>
<td>II.</td>
<td>Analysis of Variance for Teacher Interpretation Scores on the &quot;Standardized Test Proficiencies for Elementary Teachers&quot;</td>
<td>47</td>
</tr>
<tr>
<td>III.</td>
<td>Analysis of Variance for Teacher Application Scores on the &quot;Standardized Test Proficiencies for Elementary Teachers&quot;</td>
<td>48</td>
</tr>
<tr>
<td>IV.</td>
<td>Analysis of Variance for Teacher Total Scores on the &quot;Standardized Test Proficiencies for Elementary Teachers&quot;</td>
<td>49</td>
</tr>
<tr>
<td>V.</td>
<td>Analysis of Variance for Teacher Attitude Scores on the &quot;Standardized Test Proficiencies for Elementary Teachers&quot;</td>
<td>51</td>
</tr>
<tr>
<td>VI.</td>
<td>Frequency Distribution of Teacher Knowledge Scores on the &quot;Standardized Test Proficiencies for Elementary Teachers&quot;</td>
<td>52</td>
</tr>
<tr>
<td>VII.</td>
<td>Frequency Distribution of Teacher Interpretation Scores on the &quot;Standardized Test Proficiencies for Elementary Teachers&quot;</td>
<td>53</td>
</tr>
<tr>
<td>VIII.</td>
<td>Frequency Distribution of Teacher Application Scores on the &quot;Standardized Test Proficiencies for Elementary Teachers&quot;</td>
<td>54</td>
</tr>
</tbody>
</table>
**LIST OF ILLUSTRATIONS**

<table>
<thead>
<tr>
<th>Figure</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Experimental Design</td>
<td>14</td>
</tr>
</tbody>
</table>
CHAPTER I

INTRODUCTION

A major concern of educators, parents, educational psychologists, and sociologists is how teachers evaluate students. Beginning in the late 1960's, the move toward providing educational opportunities that would benefit all students has received greater stress. These opportunities include provisions for bilingual instruction, the talented and gifted, mainstreaming of handicapped students, individualized programs, Title IX, and busing for integration purposes. Standardized test results are used widely as an aid in making decisions which will influence the ultimate welfare of the individual (15). Buros (1972) offers the following criticism of test publishers who continue:

"... to market tests which do not begin to meet the standards of the rank and file of MMY (Mental Measurements Yearbook) and journal reviewers. At least half of the tests currently on the market should never have been published. Exaggerated, false, or unsubstantiated claims are the rule rather than the exception. Test users are becoming more discriminating, but not nearly fast enough (3, p. xxvii)."

The following studies have produced findings concerning a teacher's evaluation of students and the use of test information. Roeder states in a national survey of 940 accredited institutions,
This writer jestingly hypothesized that upon graduation
from college, most elementary education majors are
better prepared to conduct impromptu art and music
lessons than they are to evaluate pupil performance. . . .
Unfortunately, the data appear to support this writer's
whimsical hypothesis (17, p. 118).

Noll and Rosner expressed the need for increased
teacher competency in the use and interpretation of standard-
ized tests; they felt that undergraduate preparation was not
adequate (10).

In 1967, Mayo concluded that course work in tests and
measurements, and statistics does produce an increase in the
measurement competencies of beginning teachers. He also
stated that graduates of teacher training programs, although
having some measurement competencies, do not become deeply
involved in the problems and practices of evaluation. The
question is raised as to the relevancy and utility of the
measurement competencies taught in such courses (17).

Sorotzkin concluded from a study of teacher knowledge
of standardized test information and its effect on pupil IQ
and achievement that,

Test information is but one factor exerting a press
on the teacher in her/his interactions in the class-
room. It seems that teachers incorporate the infor-
mation they receive from tests into their global view
of the child's ability rather than basing their judge-
ments solely on this one input. Although this finding
places in question the usefulness of test information
in the academic world, massive empirical evidence
dealing with all facets of the test information
phenomenon at all grade, sex, socio-economic, and
teacher characteristic levels must be undertaken
before the educational community commits itself to
the use or elimination of group test instruments (18,
p. 83).
Anttonen reports in a study concerning the use of group administered tests that providing teachers with standardized test information did not do any harm. However, steps should be taken to link test information closer to classroom utilization (2).

Kamin gives a warning to all users of standardized tests that they are not neutral. He feels school personnel would be remiss not to question the assumptions and predictable consequences of current testing practices (13).

Statement of Problem

The problem of this study was to determine the proficiencies of selected elementary classroom teachers of second, fourth, and sixth grades in minimum skill in terminology, interpretation, and use of standardized tests, and to compare the teachers' minimum skills with those defined by a panel of public school administrators using a researcher-developed instrument.

Purposes of Study

The purposes of this study were to determine

1. Elementary teachers' knowledge of the terminology, the interpretation and the use of standardized achievement tests according to the grade level teaching assignment;

2. Elementary teachers' knowledge of terminology, interpretation and use of standardized achievement tests according to highest degree earned;
3. Elementary teachers' knowledge of terminology, interpretation and use of standardized achievement tests according to the number of years teaching experience; and

4. Elementary teachers' levels of standardized test proficiencies as compared with the minimum levels determined by public school administrators on the instrument, *Standardized Test Proficiencies for Elementary Teachers*.

**Hypotheses**

**Hypothesis I:** There will be no statistically significant difference among Group 1, second-grade teachers, Group 2, fourth-grade teachers, and Group 3, sixth-grade teachers on a measure of knowledge of standardized test terminology.

**Hypothesis II:** There will be no statistically significant difference among Group 1, second-grade teachers, Group 2, fourth-grade teachers, and Group 3, sixth-grade teachers on a measure of interpretation of standardized tests.

**Hypothesis III:** There will be no statistically significant difference among Group 1, second-grade teachers, Group 2, fourth-grade teachers, and Group 3, sixth-grade teachers on a measure of standardized test use.

**Hypothesis IV:** There will be no statistically significant difference among Group 1, second-grade teachers, Group 2, fourth-grade teachers, and Group 3, sixth-grade teachers on a measure of total scores.
Hypothesis V: There will be no statistically significant difference among experience levels of teachers on a measure of knowledge of standardized test terminology.

Hypothesis VI: There will be no statistically significant difference among experience levels of teachers on a measure of interpretation of standardized tests.

Hypothesis VII: There will be no statistically significant difference among experience levels of teachers on a measure of standardized test use.

Hypothesis VIII: There will be no statistically significant difference among experience levels of teachers on a measure of total scores.

Hypothesis IX: There will be no statistically significant difference in individual degrees earned by teachers on a measure of knowledge of standardized test terminology.

Hypothesis X: There will be no statistically significant difference in individual degrees earned by teachers on a measure of interpretation of standardized tests.

Hypothesis XI: There will be no statistically significant difference in individual degrees earned by teachers on a measure of standardized test use.

Hypothesis XII: There will be no statistically significant difference in individual degrees earned by teachers on a measure of total scores.

Hypothesis XIII: There will be no statistically significant difference among Group 1, second-grade teachers, Group 2,
fourth-grade teachers, and Group 3, sixth-grade teachers, on a measure of attitude toward standardized tests.

Hypothesis XIV: There will be no statistically significant difference among experience levels of teachers on a measure of attitude toward standardized tests.

Hypothesis XV: There will be no statistically significant difference among highest degree earned by teachers on a measure of attitude toward standardized tests.

Background and Significance of Study

The administration, interpretation and eventual use of test information by the elementary classroom teacher provides opportunity for decision making by personnel who may not be qualified in that field. Cronbach states that untrained users of standardized tests may administer the tests incorrectly, place undue reliance on inaccurate measurements, and reach unsound conclusions (5).

Leiter reports in the Sociology of Education that one of the consequences of standardized tests is to eliminate the bias of the teacher's background knowledge. He proposes that the attempts to remedy teachers' use of background knowledge have failed. The test scores do not tell the teacher what he needs to know, so the scores are equivocal from the teachers' perspective (14).

The guiding principles of a control system set by the Ethical Standards of Psychologists adopted by the American
Psychological Association as long ago as 1950 states, "Tests and diagnostic aids should be released only to a person who can demonstrate that he has the knowledge and skill necessary for his effective use and interpretation" (5, p. 11). The same concern is expressed in the standards as late as 1977.

In 1968, Fleming conducted a study to evaluate the usefulness of various kinds of intelligence test information to teachers. The purpose of the study was to examine an on-going school situation and to determine whether or not the self-fulfilling prophecy would operate when conventional measures were used. The investigator felt that there was a paucity of knowledge about the relationship between a teacher's attitude toward testing and the utilization of test results, as reflected in the performance of children in her classroom (7). The results suggested that teachers assess children, reject discrepant information, and operate on the basis of previously developed attitudes toward and knowledge about children and tests (7). Houts concurs by stating that the tests are not bad news, but misleading and open to misinterpretation (11).

Pedulla reports on an experimental study of forty-seven second-grade students in Ireland to see what impact knowledge of standardized test results would have on teachers' judgments. It was found that teachers receiving test results showed greater positive shifts in their ratings of students than those not receiving results. The conclusion was that
standardized tests results do alter grades--they tend to be raised, not lowered (16).

A doctoral study of New York secondary school English teachers in 1976 by Infantino considered the knowledge of current testing practices and the teachers' attitudes toward testing and accountability. It was found that teachers showed a lack of familiarity with testing terminology and test construction. Teachers were willing to be held accountable for pupil performance in subject matter, but not for performance measured by standardized tests (12). Plumleigh trained administrators and teachers in Los Alamitos, California to utilize test results and found that a noticeable benefit of the program was reduced anxiety over test information for both groups (17).

In a 1972 study of teachers' and counselors' attitudes toward standardized testing, Cormany concluded that colleges, professional organizations and test publishers should cooperate to improve the attitudes and skills of practicing teachers through publications and in-service programs. The study found a need for practical experience at the college level in critiquing standardized tests, writing objectives, matching tests to objectives, and researching the effectiveness of the tests (4).

Gardner, of Syracuse University, in a special report for the National Council on Measurement in Education warns of the misuse of test profiles. He gives nine warnings to assist the
schools in using profiles appropriately. The author stresses the need to base interpretation on the test scores aided by the profile, not on the profile configuration aided by the test scores. Observed differences by a quick scanning of a profile is a common abuse in profile interpretation by teachers, counselors, and school psychologists (8). Airasian studied the magnitude and direction of the effect of standardized test information on teachers' rating of 1,566 pupils. He found that test results can effect teachers' rating for only a small number of students, and that the influence of the results is to raise, not lower, ratings (1).

Definition of Terms

For the purposes of this study, the following definitions were formulated:

**Standardized achievement test.**—A test for which directions, time limits, test booklets, and other conditions of administration have been structured so that they are the same each time the test is administered. The test measures the extent to which a person has gained certain information, or acquired certain skills. The Stanford Achievement Test, used in this study, is designed to measure understandings, skills, and abilities that are desired in the elementary curricula. Normative data are included with the test.

**Classroom teacher.**—The person primarily responsible for instruction in the second-, fourth-, or sixth-grade classroom.
Group 1.---Teachers who are currently assigned to a second-grade classroom in two suburban, metroplex school districts.

Group 2.---Teachers who are currently assigned to a fourth-grade classroom in two suburban, metroplex school districts.

Group 3.---Teachers who are currently assigned to a sixth-grade classroom in two suburban, metroplex school districts.

Standardized Test Proficiencies for Elementary Teachers.---A researcher-developed standardized test skills inventory for determining classroom teacher proficiencies in knowledge, interpretation, and application of standardized achievement test in the public schools.

Basic Assumption

It is assumed that the responses received from the second-, fourth-, and sixth-grade teachers on the survey instrument will be a reflection of the standardized test proficiencies of all elementary school teachers in suburban, metroplex school districts.

Delimitation

This study is limited to classroom teachers of the second, fourth, and sixth grades in two suburban, metroplex independent school districts.
Procedures for Collection of Data

Instrument

The instrument used in this study, Appendix A, was a standardized-achievement-test skills inventory designed by the researcher. The proficiencies selected for the study were submitted to a panel of jurors, consisting of five professors of statistics, for the purpose of establishing content validity.

The instrument was divided into three areas. Area 1 concerned statistical terminology used in standardized test results for school and parent interpretation. Area 2 concerned interpretation of test results, and Area 3 concerned application, or the teacher's use of test results in planning a student's program.

For each of the thirty-nine proficiencies the respondents were asked to respond, "yes," "no," or "don't know." The "don't knows" were included so that the "yes" and "no" answers would be valid. The total score was determined by multiplying the incorrect answers by .5 and subtracting the results from the total correct answers (18). The only variables of interest were the "yes" and "no" answers. The instrument also included three demographic statements which reflected the three independent variables in the study. They were grade level taught, years teaching experience, and highest degree held. The instrument used for assessing the attitudes of the participants was a semantic differential.
Data Collection

To establish test-retest reliability of the instrument, it was administered to a university group of graduate students from North Texas State University during January, 1980. The retest was administered to the same group one week later. This process produced a coefficient of stability between the two sets of scores.

The skills inventory was administered to 122 second-, fourth-, and sixth-grade teachers of two, suburban, metroplex school districts in the spring of 1980. The attitude test was given first, then the inventory administered to the same group of teachers.

Procedures for Analysis of Data

The data collected from the two instruments was taken from the inventories and placed on encoding sheets. From these sheets the information was key-punched on computer cards for use with the SPSS software package on an IBM-370 computer.

The terminology knowledge score, $S_1$, interpretation score, $S_2$, and planning and evaluation scores, $S_3$, were the raw scores, i.e., the number right on each of the respective sub-tests. The total score, $S_4$, was a weighted composite of the $z$ scores for $S_1$, $S_2$, and $S_3$. The weights were inversely related to the three standard deviations (9). An individual's attitude score was the mean of the evaluative scales on the
semantic differential. A group attitude score was the mean of individual scores.

Hypotheses I through XV were tested using five runs on an analysis of variance (see Figure 1). The dimensions of the design were (a) grade level groups, (b) highest degree earned, and (c) number of years teaching experience. The independent variables were the four test scores; interpretation, knowledge, use, total, and the attitude score.

The three criterion variables were (a) knowledge scores, (b) interpretation, and (c) application scores, respectively, for each hypothesis.
Fig. 1—Experimental Design
CHAPTER BIBLIOGRAPHY


11. Houts, Paul L., "Behind the Call for Test Reform and Abolition of the IQ/Culture Issue," Phi Delta Kappan, 57 (June, 1976), 674-676.


CHAPTER II

REVIEW OF RELATED LITERATURE

The elementary classroom teacher is often confronted with the administration of one or more standardized achievement tests to his or her students in a school year. It is not always possible to have the services of a trained test technician to handle the testing program. As a result, the administration and interpretation of tests are usually left to the classroom teacher (2).

The academic test is predominant in public discussions of education. The massive Coleman report, in 1966, relied on correlations of achievement tests results with characteristics of schools, teachers, pupils, and their families. The volume of Jencks and others, in 1972, dealt with correlations of achievement tests results with criteria of accomplishment in later life. Newspaper editorial comments on test results appear when the scores change for better or for worse (23).

Although strong arguments against standardized achievement testing have been presented in educational literature over the past few years, the influence of these tests is not diminishing. The impact of the primary role state assessment programs are assuming in operationalizing educational
accountability is increasing the attention to results of tests and assessments (14).

The analyses of research studies concerning standardized achievement testing in the elementary school concern various themes. Nolte (28), Walden (35), O'Donnell (29), Haag (17), Birchall (4), Holmen (19), Statz (34), Ebel (11), and Walstrom (36) have focused on the use and misuse of standardized test information; Fillos (14), Forsyth (15), Boyd (7), The National Education Association (26), and Beggs (3) examined teacher attitude toward use of standardized test information by school personnel. Numerous studies which include those by Porter (31), Oliver (30), Newman (27), Ayer (1), Long (24), Bolig (5), Stanley (33), Hubbard (20), Funk (16), Jenkins (21), Johnson (22), and Cassidy (9) have examined an area related to teacher attitude, the effects of standardized test results on the school curriculum. Current practices in the publishing of standardized achievement tests are examined by Bossone (6), Hoepfner and Doherty (18), Callenbach (8), Cowie (10), McCraig (25), and Echternacht (13).

Use of Standardized Tests

Setting aside educational implications inherent in grouping, Nolte (28) expresses a concern for the legal problems facing those who make decisions related to sorting children into categories for purposes of instruction. Nolte states the Constitution, not the tools, is the best measure
of whether fundamental fairness, reasonableness and good will are being exercised by those who sort, categorize, classify, group, identify, label, screen, track and compare children. Nolte concludes that the schools may need to place less emphasis on testing and more on the individual rights of the child.

Walden (35) warns that while schools continue to examine standardized testing on philosophical and educational grounds, educators should also be aware of its possible legal consequences. Courts will be interested in how schools use data from standardized tests and what happens to children as a result. The author states, "Courts have not concluded that standardized tests are completely inappropriate for schools, but judges are extremely wary of their use as the principal criterion for pupil placement" (35, p. 81).

In discussing the misuse of standardized tests, O'Donnell (29) states an ability test measures what one has learned through formal teaching methods in school. For a test to be fair, educational models need to be developed which would consider the linguistic and cognitive styles of a child's initial exposure to formal education. The author contends that the classroom contributes to culture bias through standardized tests based on white middle-class norms.

Haag (17) finds a definite need for standardized tests in curriculum planning. These "goals" include individualized instruction, accountability, long-range planning, and
effective teacher education. However, a great concern is expressed on the interpretation of test scores. The current methods of reporting scores—percentiles, standard scores, stanines, and grade equivalencies—present sophisticated problems of interpretation and are likely to be misunderstood, even by teachers. Haag contends that there is evidence that teachers do not learn a great deal about tests and measurement within the typical teacher training program. Birchall (4) agrees in a position paper on reading. The author recommends that more should be learned about test construction and a closer study of test manuals in order to better understand the uses and limitations of these measures.

Among the recommendations of the National Consortium on Testing, in 1978, concerning the use of standardized achievement tests, was that a comprehensive study be undertaken on standardized tests' actual administration and use of scores in schools.

Holmen (19) finds the major criticisms of standardized educational testing as (a) discriminating against some individuals, (b) predict imperfectly, (c) scores may be rigidly interpreted, (d) may influence teacher expectations, (e) have harmful effects on shaping cognitive styles, (f) distort self-concept, and (g) results in select, homogeneous grouping. The author concludes there have been too many serious lapses of professional judgment by individuals who are using tests without the proper qualifications.
Statz (34) surveyed school superintendents and members of the National Council on Measurement in Education concerning the users of educational tests. The study included opinions on the levels of proficiency necessary for users to function effectively. Results of the survey indicated both groups felt teachers, principals and guidance counselors should be required to pass a comprehensive measurement course before being allowed to work with these instruments in schools. Walstrom (36) studied teacher use of standardized tests and the amount of college preparation in test assessment. Most teachers reported having had one course in assessment procedures either at teacher college or a Faculty of Education, although about one in five reported no such training. The conclusion made was that there is a real need for further training in assessment procedures.

**Conclusion**

The use of standardized achievement testing in the public schools receives criticism based on the professionals' lack of understanding of the tests. Critics find unqualified personnel use the test information primarily to group or classify students. This practice is viewed as unfair, unconstitutional, and unnecessary in many instances. Although accepted as appropriate by the schools, the courts view standardized tests as inappropriate when used as a main criterion for pupil placement. The test results are most
effective when correctly interpreted for curriculum planning. Test interpretation by unqualified personnel is a major concern.

Teacher Attitude

Fillos (14) conducted a study of a group of elementary teachers to measure teacher attitude toward standardized testing evaluations. Fillos states that student evaluations implicitly or explicitly reflect on teachers and other responsible personnel. He found, for most teachers, standardized tests did not appear to be a salient organ in the perceived evaluation program. This, Fillos attributes to the fact that (a) teachers believe they are evaluated only by the principal, (b) students are evaluated by teacher judgment, and (c) teachers do not have a clear understanding of any evaluation system in their schools. Among the conclusions the author finds many teachers see no value for standardized testing to anyone but themselves, and they suspect a "hidden agenda" of such testing is held by administrators.

Forsyth (15) studied teacher attitude toward standardized tests questioning if the benefits are worth the costs. It was concluded that too little communication occurs in many schools and teachers professed to be uninformed as to how the test information was used. Few teachers were unalterably convinced that test data are of no use to anyone.
Open discussions by administrators, counselors, and teachers may serve to clarify how test results are used by all groups.

In an evaluation of a district's standardized testing program, Boyd (7) found teachers had little or no in-service to understand test formats and to become familiar with administration procedures. Too much time was required for computing, summarizing and filing results, teachers were able to explain results to parents only in broad, general terms, and most felt the results confirmed what they already knew. Beggs (3) found similar results in a study of various techniques of interpreting test results on teacher perception and pupil achievement. One conclusion was that few teachers base their judgment of the student's ability solely on the information that has been gleaned exclusively from a single test score.

The National Education Association (26) called for a moratorium on standardized testing and created a task force on testing in 1972. The beliefs of the task force included the training of those administering tests was inadequate, and that schools of education, school systems, and the testing industry must take the responsibility. Teachers reported that they are frequently unfamiliar with the standardized tests which they are required to administer, the purposes of the overall evaluation programs they are part of, and the uses that will be made of the results. A survey of requirements in the fifty states for instruction in tests and
measurement as a prerequisite for teacher licensure showed that only thirteen states have such requirements.

Conclusion

Although standardized tests are considered of some value to teachers, they are not relied upon as heavily as might be expected. Teachers view the testing programs as an aid to student evaluations, but rely more on teacher judgment. There is a need for improved communications between administrators, counselors, and teachers as to how the test results are to be utilized.

Teacher training institutions and school districts are lacking in required tests and measurement training to better prepare teachers in the area of standardized testing.

Curriculum Effects

Who controls the schools and how? Porter (32) suggests that the teacher is becoming less an autonomous decision maker and more an agent of public school policy makers. The accountability movement and minimal competency testing are in response to the public demand for knowledge and control of the content of instruction. Oliver (30) states that informal inventories placed students in easier materials at the primary level and increasingly more difficult materials in the intermediate grades. The data suggest the placement is up to two full years in the fifth and sixth grades.
Oliver concludes it is understandable why students are frustrated by reading materials and have problems in the secondary schools.

Newman (27) found standardized achievement tests to be a less valid evaluation when administered to young, low-achieving readers. Newman cautions that invalid scores can be used in a manner harmful to the child if the user of the scores assumes them to be valid. It was concluded that until alternatives of the graded testing plan are found for low achievers, standardized test scores interpretation should be done with caution. The effects of three modes of test administration, approval, standard, and disapproval, were studied by Johnson (22). The findings included the belief that, should teachers deviate from standard examiner behavior during testing periods, there is a good chance that the testing situation would be unstandardized. Johnson recommends in-service training for teachers designed to teach appropriate and uniform test behavior.

Similar studies were conducted by Ayer (1) and Long (24) comparing the assigned reading level of low achievers and the level of standardized achievement tests administered. Ayer found the use of out-of-level tests for reading resulted in depressed scores and concluded that scores are comparable with small groups who have normal distributions. Ayer suggests test publishers develop norms based on out-of-level
administration of their tests. Long questioned whether students would receive the same grade-equivalent score when taking out-of-level and instruction level reading tests. It was concluded that the assumption "data are data" was a fallacious one. Long states the lack of empirically developed norms for out-of-level tests, or tests which are not on the student's reading level, makes interpretation of the scores difficult. Long agrees with Ayer in the need for publishers of achievement tests to develop appropriate out-of-level norms for low achieving groups of students.

Bolig (5), Stanley (33), Funk (16), Hubbard (20), Jenkins (21), and Ebel (11) examined the relationship between teacher opinion and student performance on a standardized achievement test. Bolig studied kindergarteners' teacher-predicted first-grade success as compared with the Metropolitan Readiness Test (MRT). The MRT proved to be as good or better than the ratings of teachers. Stanley found that a score on a difficult standardized mathematics-aptitude test predicted over a two-to-three-year period considerably better than the judgment of the students' mathematics teacher.

Hubbard (20) studied teacher ratings compared with standardized measures of achievement to determine the extent of their relationship. The question of interest was if a teacher's rating of a pupil provides the same information concerning his achievement as his standardized test scores. Teachers' judgments of "high" and "low" third- and sixth-
grade students were consistent with the achievement test results. Hubbard concluded that this suggests teachers use classroom performance as a criterion in making their ratings much more than any knowledge of a recorded IQ score. Cassidy's (9) study of appropriate reporting of student progress in reading found that parents and teachers considered standardized tests comparing students by grade level or percentile rank as less important than data which described the child's actual reading performance in attitude or behavior.

Funk (16) reported an examination of what research says about the relationship between teaching and student performance on achievement tests in the science programs. He contends that present elementary science programs are teaching new processes which are not being measured by standardized tests. A comparison of twelve tests currently used to measure achievement in elementary science revealed much variation among test content, grade level, and cognitive levels. Funk concludes that people who use and administer standardized tests should critique them for their strengths and weaknesses. There should be more effort to match tests to elementary science programs. Ebel (11) concurs in stating that standardized achievement tests enforce curricular conformity and fail to test what local schools or particular teachers have been trying to teach.
Jenkins (21) studied a group of first- and second-grade basal reading series for curriculum bias of achievement tests as they relate to the evaluation of teachers, children and the curricula. A compilation of basal word lists were compared with achievement tests for extent of overlapping in both sources. Specific attention was given to reading placement, and the identification and classification of exceptional children. The results indicated that achievement tests may not be capable of reliably discriminating handicapped children from those who have learned every word taught in the reading curriculum.

Jenkins states that "when achievement tests run counter to teachers' perceptions of children's progress, the achievement test score is usually accepted as the more valid assessment" (21, p. 451). Jenkins concluded that a basic assumption underlying standardized achievement measures—-they representatively sample different curricula—is largely without support; clear, significant biases appear to exist.

Conclusion

The use of standardized achievement tests have certain disadvantages in planning student programs. The results of these tests tend to place students, particularly the low-achievers, in reading levels that are too difficult and result in student frustration. Substituting out-of-level tests for the low-achiever in reading depresses the scores
and there is an apparent need for tests based on out-of-level administration of the tests.

Standardized achievement tests do not necessarily match the subject material taught in the classroom. The teachers tend to use their own judgment of a student's ability when the standardized test scores do not agree with their own.

Standardized Achievement Test Practices

Standardized achievement tests are currently utilized to test children at almost every level of school life: when they enter kindergarten, before graduating from high school, before entering college. Bossone (6) contends that the growing interest in testing is coincident with the growing interest in education in general. Citing the evaluation of New York City's testing program, in 1970, the author describes the concern as a result of the general public, and parents, seeking assurance that tests serve some worthwhile purpose. Bossone's report points out the need for improvement in (a) a lack of communicating test results in "simple English" terminology, (b) a lack of dialogue in the testing field, (c) a need for federal aid based on school and pupil performance, and (d) more reliable instruments. The author concludes that the testing industry must either move toward self-regulation or create a major dialogue with the audience for whom the testing industry performs.
Hoepfner and Doherty (18) studied the profiles of seven elementary-level standardized achievement tests to evaluate the strengths and weaknesses of test procedures. Twenty-four different aspects were grouped into four major categories: Measurement Validity, Examinee Appropriateness, Administrative Usability, and Normed Technical Excellence. The data supported the supposition that the test publishers differ in their priorities as reflected in their tests' characteristics. Publishers can be characterized by groups as (a) producing tests highly usable administratively and fairly good in terms of examinee appropriateness, (b) producing tests with greater relevance and validity, and (c) emphasizing the examinee appropriateness to the neglect of other test qualities. The authors concluded such a profile could be used by test publishers as a guideline for self-improvement and improving the qualities of their product.

Callenbach (8) studied the effects, both immediate and long-range, of practice and instruction in test wiseness for nonverbal materials upon standardized reading test scores of test-naive second-grade students. Lessons provided instruction in following oral-directions, response marking, using time efficiently, and guessing. The data gathered from the investigation demonstrated that the standardized reading test scores of test-naive second-grade students can be significantly raised through instruction. Despite some limitations listed by the author, it is suggested by
Callenbach that periodic practice and instruction in standardized test-taking techniques in the elementary classroom would lend more credence to the practice of classifying children according to standardized achievement test scores. Ebel (11) disagrees in stating that practicing test-taking allows students to become familiar with test items. Their actual test may not reflect previous learning, but an ability to grasp testing skills.

Cowie (10) finds some agreement in a report on intelligence and intelligence testing, and the effects of coaching. Cowie states,

Coached groups (those having been told about tests and had numerous representative items explained by the teacher) gain about five to six IQ points. Practiced but uncoached groups gained about six IQ points. Practiced and coached groups may gain eight to ten points (10, p. 6).

Cowie adds, however, that if coaching for tests occurs there is something wrong with the system.

In evaluating test publishing of standardized English tests for the elementary level, McCraig (25) contends the problem lies in the critical criteria being determined by whether the data can be collected easily and processed in the computer. Specific criticism is directed in the area of punctuation as compared with total writing skills such as unity, organization, and word choice. The author finds fault not only with test publishers but with educators who keep buying the tests. Recommendations include school districts'
evaluation of student writing by looking at actual writing samples.

How well school personnel interpret test results was studied by Echternacht (13). The author found that while most school districts believe they understand grade equivalent scores—test publishers urge the use of percentile standards which they feel the user better understands.

Conclusion

The problem of standardized achievement test construction and publication is growing with the increased interest in test use in the schools. The public demands for knowledge of test results provides a critical evaluation of test production; how valid, reliable and ultimately how useful the instrument.

There is some question as to whether schools should instruct in test-taking procedures. Studies do show scores can be raised with students having prior knowledge of the test, but this may not be indicative of mastery of learning. Critics feel that the publishing system has been established to permit ease in collecting and processing data, rather than testing for skills actually taught in the classroom.
CHAPTER BIBLIOGRAPHY


CHAPTER III

METHODS AND PROCEDURES FOR
COLLECTION OF DATA

The problem of this study was to determine the proficiencies of selected elementary classroom teachers of second, fourth, and sixth grades in minimum skills in terminology, interpretation, and use of standardized achievement tests. These minimum skills, as defined by a panel of jurors consisting of eleven public school administrators, were compared to actual teacher proficiency levels as determined by the "Standardized Test Proficiencies for Elementary Teachers"--a researcher-developed instrument.

Description of the Population

The subjects in this study were 122 elementary classroom teachers within two suburban, metroplex school districts. Second-, fourth-, and sixth-grade teachers participated from twelve elementary schools in one district, two elementaries and one middle school of a second district. The middle school provided the sixth-grade level which was not located in the elementary buildings of the second district.

The population included forty-three second-grade teachers, thirty-nine fourth-grade teachers, and forty-sixth-grade teachers. There was no random selection of
participants—all teachers on the designated grade levels were asked to participate.

Construction of the Initial Skills Inventory

Construction of the initial skills inventory was begun with the determination of the elementary teachers' proficiencies in standardized achievement test procedures and practices implemented in elementary schools. Through an intensive review of literature, a list of possible items was prepared.

The instrument used for the study was the "Standardized Test Proficiencies for Elementary Teachers," an inventory designed by the researcher (Appendix A). It was composed of three major areas of investigation. In each of the areas the respondent had a choice of answering "yes," "no," or "don't know" to a total of thirty-nine items. The "don't knows" were included so that "yes" and "no" answers would be valid. The total score was determined by multiplying the incorrect answers by .5 and subtracting the results from the total correct answers. The only variables of interest were the "yes" and "no" answers.

Area 1, Standardized Test Terminology, included fifteen statements concerning the statistical terminology used in standardized achievement test results for school and parent interpretation. Area 2, Test of Interpretation, included thirteen statements concerning interpretation of test results,
and Area 3, Test of Application, included eleven statements concerning teacher use of test results in planning a student's program.

The original instrument was submitted to a panel of five judges to establish content validity. The criterion used in the selection of the panel of judges was that they were currently professors of statistics. Four of the jurors were from North Texas State University and one was from the University of Dallas.

The initial instrument was presented to the panel of jurors to obtain their opinions for validation of the items. A copy of the instrument accompanied by a cover letter (Appendix C), and a return-addressed stamped envelope were mailed to each member of the panel. The judges were asked to read each statement and circle one of the three numerals in the left margin. The numerals indicated (1) the item is important to the study, (2) the validity could not be readily decided, and (3) the item is not clearly stated, or does not pertain to the study. An additional sheet was provided for comments. A simple majority decided the item's inclusion in the instrument. Items receiving a (1) were accepted, those receiving a (2) were revised as directed, and those receiving a (3) were discounted. Appendix D charts the results. Upon recommendations that additional items were necessary, fourteen items were added and mailed to the jurors for judgment. All fourteen additional items were acceptable. The demographic
portion of the instrument was designed with three items reflecting the three independent variables in the study. These were the respondent's Present Position of second-, fourth-, or sixth-grade teacher; Years Teaching Experience of zero to two years, three to ten years, or eleven or more years; and Highest Degree Earned, a baccalaureate degree, a master's degree, or beyond the master's degree.

To establish test-retest reliability, the instrument was administered to a university group of graduate students from North Texas State University in January, 1980. The test was given in one session and a retest administered one week later to the same group. This time delay was used as suggested by Borg (1) to prevent recall of answers. The correlation between the two scores was significant, \( r = .6503 \), but less than desirable (Appendix H).

**Attitude Instrument**

A semantic differential instrument was included to assess the participant's attitude toward standardized achievement tests in the elementary school (Appendix B). Henderson (2) describes the instrument, for analyzing judgments, as a highly generalized technique of measurement when applied to the requirements of a specific research problem.

The fifteen items were selected, by the researcher, from a list of terms recommended by Henderson (2), for their relevance and semantic stability. They included three "active"
terms, three "potency" terms, and nine "evaluative" terms. Respondents were asked to choose a point on the scale which more closely described their attitude. The scales ranked from one (highly agree) to seven (highly disagree). Items were arranged reversing the positive and negative extremes to prevent the formation of position preference. The factor score, or average of the related scales, was derived to determine the participant's general attitude toward standardized achievement tests.

**Administration Instrument**

The "Standardized Test Proficiencies for Elementary Teachers" (Appendix A), a rating sheet (Appendix E), and a cover letter (Appendix F), were mailed to twelve public school administrators whose position currently involved them in test interpretation or in aiding classroom teachers in the interpretation and use of standardized achievement scores. The administrator group was composed of one director of pupil services, one elementary school counselor, a director of special education, a supervisor of special education, a district reading consultant, six elementary principals, and one secondary principal who had previously served as an elementary principal. Eleven requests were returned, with one unreturned.

The participants were asked to review the instrument and indicate the percentage of items, in their opinion, an
elementary teacher should be able to answer correctly. These results (Appendix G) were averaged and compared with teacher results on the skills inventory. This provided a comparison of "what is," the teachers' scores, with "what should be," the administrators' opinions.

Statistical Procedures

The data collected from the two instruments were taken from the inventory and placed on encoding sheets. From these sheets the information was keypunched on computer cards for use in the SPSS software package on an IBM-370 computer.

The terminology knowledge scores, $S_1$, interpretation score, $S_2$, and planning and evaluation scores, $S_3$, were the raw scores, i.e., the number right on each of the respective subtests. The total score, $S_4$, was a weighted composite of the $z$ scores for $S_1$, $S_2$, and $S_3$. The weights were inversely related to the three standard deviations. The individual attitude score was the mean of the evaluative scales on the semantic differential. The group attitude score was the mean of the individual scores.

Hypotheses I through XV were tested using five runs on an analysis of variance. The dimensions of the design were (1) grade level taught, (2) highest degree earned, and (3) number of years teaching experience.
CHAPTER BIBLIOGRAPHY


CHAPTER IV

PRESENTATION OF DATA

The purpose of this study was to determine elementary teachers' knowledge of test terminology, interpretation, and application of standardized achievement tests according to grade level assignment, highest degree earned, and number of years teaching experience. The participants in the study included second-, fourth-, and sixth-grade classroom teachers in two suburban school districts. A panel of jurors, consisting of eleven school administrators, provided data for a comparison of actual teacher knowledge of standardized achievement tests with the ideal level of proficiencies determined by the administrative group. An analysis of variance was used to determine the correlations among the variables.

The data have been organized into two sections. The first section analyzes the differences in mean scores among the three variables of grade level taught, number of years teaching experience, and highest degree earned, according to teachers' knowledge, interpretation, application, and attitude toward standardized achievement tests. The second section analyzes the frequency distribution of scores among
the three variables compared with administrators' ratings of teachers' proficiencies.

Analysis of the Differences in Mean Scores of Grade Level Taught, Years Teaching Experience, and Highest Degree Earned Versus Knowledge of Standardized Tests

In Hypotheses I, V, and IX, it was predicted that there would be no statistically significant difference in knowledge scores across the three variables of grade level taught, years teaching experience, and highest degree earned. The mean scores, degrees of freedom, and calculated F values used to test these hypotheses are reported in Table I.

TABLE I
ANALYSIS OF VARIANCE FOR TEACHER KNOWLEDGE SCORES ON THE "STANDARDIZED TEST PROFICIENCIES FOR ELEMENTARY TEACHERS"

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grade Level</td>
<td>5.49</td>
<td>2</td>
<td>2.74</td>
<td>.582</td>
</tr>
<tr>
<td>Experience</td>
<td>13.88</td>
<td>2</td>
<td>6.94</td>
<td>1.472</td>
</tr>
<tr>
<td>Degree</td>
<td>13.47</td>
<td>2</td>
<td>6.74</td>
<td>1.429</td>
</tr>
<tr>
<td>Error</td>
<td>542.47</td>
<td>115</td>
<td>4.71</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>579.47</td>
<td>121</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

In Table I, grade level taught, years teaching experience, and highest degree earned, respectively, reached F values of .583, 1.472, and 1.429. With an N of 122, degrees of freedom
2, and $F$ value of 3.93 must be reached to obtain significance at the .05 level. None of the variables obtained this required level of significance and cannot be considered to have shown significant interaction. The test results failed to reject Hypotheses I, V, and IX. In these samples, years teaching experience and highest degree earned had higher means for knowledge scores than grade level taught. Experience does appear to have the strongest influence on knowledge, with degree held following closely. Due to the nature of the statistical analysis, it is not possible to ascertain in what direction this influence takes. In summary, there is no apparent significant difference across the three variables versus knowledge.

Analysis of the Difference in Mean Scores of Grade Level Taught, Years Teaching Experience and Highest Degree Earned Versus Interpretation of Standardized Tests

In Hypotheses II, VI, and X it was predicted that there would be no statistically significant difference in interpretation scores across the three variables, grade level taught, years teaching experience, and highest degree earned. The mean scores, degrees of freedom, and calculated $F$ values to test these hypotheses are reported in Table II.

In Table II, grade level taught, years teaching experience, and highest degree earned, respectively, reached $F$ values of 2.89, .85, and 1.21. With an $N$ of 122, and
TABLE II

ANALYSIS OF VARIANCE FOR TEACHER INTERPRETATION SCORES
ON "THE STANDARDIZED TEST PROFICIENCIES FOR
ELEMENTARY TEACHERS"

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grade Level</td>
<td>29.97</td>
<td>2</td>
<td>14.98</td>
<td>2.89</td>
</tr>
<tr>
<td>Experience</td>
<td>8.83</td>
<td>2</td>
<td>4.42</td>
<td>.85</td>
</tr>
<tr>
<td>Degree</td>
<td>12.57</td>
<td>2</td>
<td>6.28</td>
<td>1.21</td>
</tr>
<tr>
<td>Error</td>
<td>596.35</td>
<td>115</td>
<td>5.19</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>642.16</td>
<td>121</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

degrees of freedom 2, and F value of 3.93 must be reached to obtain significance at the .05 level. The three variables did not obtain the required level of significance, and cannot be considered to have shown significant interaction.

The test results failed to reject Hypotheses II, VI, and X. Grade level taught reached an F value of 2.89, the highest of the three variables, and appears to have the strongest influence on interpretation scores.

Analysis of the Differences in Mean Scores of Grade Level Taught, Years Teaching Experience and Highest Degree Earned Versus Application of Standardized Tests

In Hypotheses III, VII, and XI it was predicted that there would be no statistically significant difference in application scores across the three variables, grade level
taught, years teaching experience, and highest degree earned. The mean scores, degrees of freedom, and calculated F values to test these hypotheses are reported in Table III.

**TABLE III**

**ANALYSIS OF VARIANCE FOR TEACHER APPLICATION SCORES ON THE "STANDARDIZED TEST PROFICIENCIES FOR ELEMENTARY TEACHERS"

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grade</td>
<td>3.73</td>
<td>2</td>
<td>1.87</td>
<td>.46</td>
</tr>
<tr>
<td>Experience</td>
<td>9.10</td>
<td>2</td>
<td>4.55</td>
<td>1.13</td>
</tr>
<tr>
<td>Degree</td>
<td>1.01</td>
<td>2</td>
<td>.51</td>
<td>.13</td>
</tr>
<tr>
<td>Error</td>
<td>464.09</td>
<td>115</td>
<td>4.04</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>480.16</td>
<td>121</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

In Table III, grade level taught, years teaching experience, and highest degree earned, respectively, reached F values of .46, 1.13, and .13. The three variables did not reach the critical value for significance, and cannot be considered to have shown significant interaction.

The test results failed to reject Hypotheses III, VII, and XI. Years teaching experience reached an F value of 1.13, the highest of the three variables. In these samples, years teaching experience had the highest scores for applying standardized test information. Although the differences
are not significant, the teachers' years of experience in
the classroom do effect their knowledge of utilizing students' 
standardized test results in program planning.

Analysis of Variance for Teacher Total Scores on the
"Standardized Test Proficiencies for Elementary
Teachers"

In Hypotheses IV, VIII, and XII it was predicted that
there would be no statistically significant difference on a
measure of total scores across the three variables of grade
level taught, years teaching experience, and highest degree
earned. The mean scores, degrees of freedom, and calculated
F values used to test these hypotheses are listed in Table
IV.

TABLE IV

ANALYSIS OF VARIANCE FOR TEACHER TOTAL SCORES ON THE
"STANDARDIZED TEST PROFICIENCIES FOR ELEMENTARY
TEACHERS"

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grade Level</td>
<td>88.28</td>
<td>2</td>
<td>44.14</td>
<td>2.13</td>
</tr>
<tr>
<td>Experience</td>
<td>38.94</td>
<td>2</td>
<td>19.47</td>
<td>.94</td>
</tr>
<tr>
<td>Degree</td>
<td>59.54</td>
<td>2</td>
<td>29.77</td>
<td>1.44</td>
</tr>
<tr>
<td>Error</td>
<td>2380.70</td>
<td>115</td>
<td>20.70</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>2575.46</td>
<td>121</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
In Table IV the total scores for grade level taught, years teaching experience, and highest degree earned, respectively, reached F values of 2.13, 194, and 1.44. With an N of 122 and 2 degrees of freedom, an F value of 3.93 was necessary to obtain significance at the .05 level. The test results failed to reject Hypotheses IV, VIII, and XII. As in the analysis of interpretation scores, the grade level taught variable reached the highest level with an F value of 2.13. Grade level taught had the greatest influence on total scores.

Analysis of Variance for Teacher Attitude Scores on the "Standardized Test Proficiencies for Elementary Teachers"

In Hypotheses XIII, XIV, and XV it was predicted that there would be no statistically significant difference in scores of attitude toward standardized achievement tests across the three variables of grade level taught, years teaching experience and highest degree earned. The mean scores, degrees of freedom, and calculated F values are listed in Table V.

In Table V the grade level taught, years teaching experience, and highest degree earned, respectively, reached F values of .49, 1152, and 1.90. With an N of 122 and 2 degrees of freedom, an F value of 3.93 must be reached to obtain significance at the .05 level. The three variables did not obtain the critical value of significance, and cannot
### TABLE V

**ANALYSIS OF VARIANCE FOR TEACHER ATTITUDE SCORES ON THE "STANDARDIZED TEST PROFICIENCIES FOR ELEMENTARY TEACHERS"**

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grade Level</td>
<td>.50</td>
<td>2</td>
<td>.25</td>
<td>.49</td>
</tr>
<tr>
<td>Experience</td>
<td>1.57</td>
<td>2</td>
<td>.78</td>
<td>1.52</td>
</tr>
<tr>
<td>Degree</td>
<td>1.97</td>
<td>2</td>
<td>.98</td>
<td>1.90</td>
</tr>
<tr>
<td>Error</td>
<td>59.37</td>
<td>115</td>
<td>.52</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>62.88</td>
<td>121</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The test results failed to reject Hypotheses XIII, XIV, and XV. Highest degree earned reached an F value of 1.90, the highest of the three variables. This variable has the strongest influence on attitude, while grade level taught had the lowest scores.

### Analysis of Frequency Distribution for Knowledge, Interpretation, and Application Scores on the "Standardized Test Proficiencies for Elementary Teachers"

One of the purposes of this study was to determine elementary teachers' levels of standardized test proficiencies as compared with the minimum levels determined by a panel of public school administrators. Appendix G lists the percentage of items in each of the three areas, knowledge, interpretation,
and application, that individual administrators considered teachers should be able to answer correctly. Tables VI, VII, and VIII report the distribution of scores the teachers obtained on knowledge (Table VI), interpretation (Table VII), and application (Table VIII).

**TABLE VI**

**FREQUENCY DISTRIBUTION OF TEACHER KNOWLEDGE SCORES ON THE "STANDARDIZED TEST PROFICIENCIES FOR ELEMENTARY TEACHERS"**

<table>
<thead>
<tr>
<th>Raw Scores</th>
<th>Per Cent</th>
<th>f</th>
<th>rf pct</th>
<th>cf pct</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>.8</td>
<td>.8</td>
</tr>
<tr>
<td>2</td>
<td>13</td>
<td>1</td>
<td>.8</td>
<td>1.6</td>
</tr>
<tr>
<td>3</td>
<td>20</td>
<td>5</td>
<td>4.1</td>
<td>5.7</td>
</tr>
<tr>
<td>4</td>
<td>27</td>
<td>12</td>
<td>9.8</td>
<td>15.6</td>
</tr>
<tr>
<td>5</td>
<td>33</td>
<td>13</td>
<td>10.7</td>
<td>26.2</td>
</tr>
<tr>
<td>6</td>
<td>40</td>
<td>18</td>
<td>14.8</td>
<td>41.0</td>
</tr>
<tr>
<td>7</td>
<td>47</td>
<td>14</td>
<td>11.5</td>
<td>52.5</td>
</tr>
<tr>
<td>8</td>
<td>53</td>
<td>31</td>
<td>25.4</td>
<td>77.9</td>
</tr>
<tr>
<td>9</td>
<td>60</td>
<td>14</td>
<td>11.5</td>
<td>89.3</td>
</tr>
<tr>
<td>10</td>
<td>67</td>
<td>8</td>
<td>6.6</td>
<td>95.9</td>
</tr>
<tr>
<td>11</td>
<td>73</td>
<td>4</td>
<td>3.3</td>
<td>99.2</td>
</tr>
<tr>
<td>12</td>
<td>80</td>
<td>1</td>
<td>.8</td>
<td>100.0</td>
</tr>
<tr>
<td>Total</td>
<td>...</td>
<td>122</td>
<td>100.0</td>
<td>100.0</td>
</tr>
</tbody>
</table>

N = 122, $\bar{X} = 6.93$, SD = 2.19, and $s^2 = 4.79$.

Table VI lists the distribution of scores the teacher obtained in Area 1, Standardized Test Terminology. The mean score was 6.93, or a score of 46 per cent. Administrators (Appendix G) considered a score of 72 per cent necessary for a teacher to be proficient in the area of test terminology.
Only 5 of the 122 subjects reached 73 per cent or above, with the one highest being 80 per cent. The mode was a score of 8, or 53 per cent of the items answered correctly.

Table VII reports the distribution of scores the teachers obtained on Area 2, Test of Interpretation.

**TABLE VII**

FREQUENCY DISTRIBUTION OF TEACHER INTERPRETATION SCORES ON THE "STANDARDIZED TEST PROFICIENCIES FOR ELEMENTARY TEACHERS"

<table>
<thead>
<tr>
<th>Raw Score</th>
<th>Per Cent</th>
<th>f</th>
<th>rf pct</th>
<th>cf pct</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>.8</td>
<td>.8</td>
</tr>
<tr>
<td>1</td>
<td>8</td>
<td>4</td>
<td>3.3</td>
<td>4.1</td>
</tr>
<tr>
<td>2</td>
<td>18</td>
<td>10</td>
<td>8.2</td>
<td>12.3</td>
</tr>
<tr>
<td>3</td>
<td>25</td>
<td>13</td>
<td>10.7</td>
<td>23.0</td>
</tr>
<tr>
<td>4</td>
<td>31</td>
<td>16</td>
<td>13.1</td>
<td>36.1</td>
</tr>
<tr>
<td>5</td>
<td>38</td>
<td>20</td>
<td>16.4</td>
<td>52.5</td>
</tr>
<tr>
<td>6</td>
<td>46</td>
<td>24</td>
<td>19.7</td>
<td>72.1</td>
</tr>
<tr>
<td>7</td>
<td>54</td>
<td>12</td>
<td>9.8</td>
<td>82.1</td>
</tr>
<tr>
<td>8</td>
<td>62</td>
<td>12</td>
<td>9.8</td>
<td>91.8</td>
</tr>
<tr>
<td>9</td>
<td>69</td>
<td>5</td>
<td>4.1</td>
<td>95.9</td>
</tr>
<tr>
<td>10</td>
<td>77</td>
<td>3</td>
<td>2.5</td>
<td>98.4</td>
</tr>
<tr>
<td>11</td>
<td>85</td>
<td>2</td>
<td>1.6</td>
<td>100.0</td>
</tr>
<tr>
<td>Total</td>
<td>...</td>
<td>122</td>
<td>100.0</td>
<td>100.0</td>
</tr>
</tbody>
</table>

\[ N = 122, \bar{X} = 5.31, SD = 2.30, \text{ and } s^2 = 5.31. \]

The mean score was 5.31 or 40 per cent. Administrators' opinions considered a score of 82.4 per cent necessary for teachers to be proficient in the area of interpretation. Two scores of 85 per cent were the only scores within the
administrators' expectations. The mode for interpretation scores was 6, or 46 per cent of the items answered correctly.

Table VIII reports the distribution of scores the teachers obtained on Area 3, Test of Application.

**TABLE VIII**

**FREQUENCY DISTRIBUTION OF TEACHER APPLICATION SCORES ON THE "STANDARDIZED TEST PROFICIENCIES FOR ELEMENTARY TEACHERS"

<table>
<thead>
<tr>
<th>Raw Scores</th>
<th>Per Cent</th>
<th>f</th>
<th>rf</th>
<th>cf</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>8</td>
<td>6.6</td>
<td>6.6</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>9</td>
<td>7.4</td>
<td>13.9</td>
</tr>
<tr>
<td>2</td>
<td>18</td>
<td>16</td>
<td>13.1</td>
<td>27.0</td>
</tr>
<tr>
<td>3</td>
<td>27</td>
<td>27</td>
<td>22.1</td>
<td>49.2</td>
</tr>
<tr>
<td>4</td>
<td>36</td>
<td>18</td>
<td>14.8</td>
<td>63.9</td>
</tr>
<tr>
<td>5</td>
<td>45</td>
<td>24</td>
<td>19.7</td>
<td>83.6</td>
</tr>
<tr>
<td>6</td>
<td>55</td>
<td>8</td>
<td>6.6</td>
<td>90.2</td>
</tr>
<tr>
<td>7</td>
<td>64</td>
<td>8</td>
<td>6.6</td>
<td>96.7</td>
</tr>
<tr>
<td>8</td>
<td>73</td>
<td>4</td>
<td>3.3</td>
<td>100.0</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>. . .</strong></td>
<td><strong>122</strong></td>
<td><strong>100.0</strong></td>
<td><strong>100.0</strong></td>
</tr>
</tbody>
</table>

N = 122, $\bar{X} = 3.69$, SD = 1.99, and $s^2 = 3.97$.

The mean score was 3.69, or 33 per cent. Administrators considered a score of 79 per cent necessary for teachers to be considered proficient in application of standardized test information. None of the scores reached the administrators' expectations. The highest score of 73 per cent was reached by only 4 subjects. The mode was 3, or 27 per cent.
Discussion

The data reported in this chapter provide an examination of teachers' utilization of standardized test information. How knowledgeable teachers are in standardized tests is most affected by the number of years teaching experience, and the least affected by the grade level taught. This is probably due to having had more exposure to the testing procedures over a period of time. However, the teachers' ability to interpret the test scores is least affected by experience and more by the grade level taught.

Proper application of the test information to the students' program is the desired outcome of any testing program. In this area, teaching experience has the most influence on the teacher's knowledge of this skill. The degree held has the least influence. This should have some effect on the administrators' approach to the schools' in-service programs, particularly for beginning teachers.

The highest degree earned has the most influence on teachers' attitudes toward standardized tests, with the grade level taught having the least influence. Increased knowledge and years teaching experience do provide additional skills, confidence, and a resulting more positive attitude toward standardized achievement tests.
Summary

In this chapter the data obtained for the study have been presented, analyzed, and discussed. The hypotheses were presented and analyzed to determine if they would be rejected or accepted.

Hypotheses I, V, and IX, which involved knowledge of standardized test terminology scores across the three variables, were not significant in any of the areas. Hypotheses II, VI, and X, which involved interpretation of standardized test scores across the three variables, were not significant. Hypotheses III, VIII, and XI, which involved application of standardized tests scores across the three variables, were not significant. Hypotheses XIII, XIV, and XV, which involved teacher attitude toward standardized tests across the three variables, were not significant. None of the null hypotheses were rejected.
CHAPTER V

SUMMARY, FINDINGS, AND RECOMMENDATIONS

Summary

The purpose of this study was to determine elementary teachers' knowledge of terminology, interpretation, and application of standardized achievement tests according to grade level assignment, highest degree earned, and years teaching experience.

The stated hypotheses were,

I. There will be no statistically significant difference among Group 1, second-grade teachers, Group 2, fourth-grade teachers, and Group 3, sixth-grade teachers on a measure of knowledge of standardized test terminology.

II. There will be no statistically significant difference among Group 1, second-grade teachers, Group 2, fourth-grade teachers, and Group 3, sixth-grade teachers on a measure of interpretation of standardized tests.

III. There will be no statistically significant difference among Group 1, second-grade teachers, Group 2, fourth-grade teachers, and Group 3, sixth-grade teachers on a measure of standardized test use.

IV. There will be no statistically significant difference among Group 1, second-grade teachers, Group 2, fourth-grade
teachers, and Group 3, sixth-grade teachers on a measure of total scores.

V. There will be no statistically significant difference among experience levels of teachers on a measure of knowledge of standardized test terminology.

VI. There will be no statistically significant difference among experience levels of teachers on a measure of interpretation of standardized test.

VII. There will be no statistically significant difference among experience levels of teachers on a measure of standardized test use.

VIII. There will be no statistically significant difference among experience levels of teachers on a measure of total scores.

IX. There will be no statistically significant difference in individual degrees earned by teachers on a measure of knowledge of standardized test terminology.

X. There will be no statistically significant difference in individual degrees earned by teachers on a measure of interpretation of standardized tests.

XI. There will be no statistically significant difference in individual degrees earned by teachers on a measure of standardized test use.

XII. There will be no statistically significant difference in individual degrees earned by teachers on a measure of total scores.
XIII. There will be no statistically significant difference among Group 1, second-grade teachers, Group 2, fourth-grade teachers, and Group 3, sixth-grade teachers, on a measure of attitude toward standardized tests.

XIV. There will be no statistically significant difference among experience levels of teachers on a measure of attitude toward standardized tests.

XV. There will be no statistically significant difference among degree levels of teachers on a measure of attitude toward standardized tests.

The subjects who participated in the study were 122 elementary classroom teachers within two suburban, metroplex school districts. Forty-three were second-grade teachers, thirty-nine were fourth-grade teachers, and forty were sixth-grade teachers. The subjects were administered the "Standardized Test Proficiencies for Elementary Teachers" and the Teacher Concepts of Standardized Achievement Tests in the Public Schools, a semantic differential test, by building groups—one session per building.

Findings

Hypotheses I, V, and IX were accepted. There was no significant difference between knowledge test scores and the variables grade level taught, years teaching experience, and highest degree earned.
Hypotheses II, VI, and X were accepted. There was no significant difference between interpretation test scores and the variables grade level taught, years teaching experience, and highest degree earned.

Hypotheses III, VII, and XI were accepted. There was no significant difference between application test scores and the variables of grade level taught, years teaching experience, and highest degree earned.

Hypotheses IV, VIII, and XIII were accepted. There was no significant difference between total scores and the variables grade level taught, years teaching experience, and highest degree earned.

Hypotheses XIII, XIV, and XV were accepted. There was no significant difference in attitude scores and the variables grade level taught, years teaching experience, and highest degree earned.

Administrators' ratings of necessary teacher proficiencies in standardized achievement test use and actual teacher scores differed, with administrators' expectations averaging 37 per cent higher scores.

Conclusions

The findings in this study support the following conclusions:

1. Teacher knowledge, interpretation, and application of standardized achievement tests are not significantly
influenced by grade level taught, years teaching experience, or highest degree earned.

2. Elementary teachers' attitudes toward standardized achievement tests are not highly influenced by grade level taught, years teaching experience, or highest degree earned. Teachers, generally, do not hold strong negative or strong positive attitudes toward standardized achievement tests.

3. Elementary teachers' knowledge and use of standardized achievement tests are far below the teacher expectations held by public school administrators.

Recommendations

Based on the results of this study, it is recommended that

1. Schools of education include test and measurement course(s) for students in teacher preparation programs.

2. Local school districts include test and measurement instruction for teachers through in-service education.
APPENDICES
STANDARDIZED TEST PROFICIENCIES FOR ELEMENTARY TEACHERS

DIRECTIONS

STEP 1: Please check the appropriate space in each of the three sections below:

(4) PRESENT POSITION

| _1. 2nd Grade Teacher | _1. 0-2 years |
| _2. 4th Grade Teacher | _2. 3-10 years |
| _3. 6th Grade Teacher | _3. 11 or more years |

(6) HIGHEST DEGREE EARNED

| _1. Baccalaureate |
| _2. Masters Degree |
| _3. Beyond Masters |

Listed below are 39 proficiency statements. Note that the proficiency statements have been separated into three main areas of skills and are numbered consecutively within the areas.

STEP 2: Read the proficiency statement. Then, check the appropriate blank (Yes, No, or Don't Know) which you feel is the best answer.

STEP 3: Repeat STEP 2 for each area presented.

A STUDY OF STANDARDIZED TEST KNOWLEDGE AND INTERPRETATION BY ELEMENTARY CLASSROOM TEACHERS

AREA 1 STANDARDIZED TEST TERMINOLOGY

1.1
A test that measures acquired information or skills as a result of specific instruction is called an achievement test.

<table>
<thead>
<tr>
<th>YES</th>
<th>NO</th>
<th>DON'T KNOW</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

1.2
A group of several tests standardized on the same population so that several tests are comparable is called a battery.

<table>
<thead>
<tr>
<th>YES</th>
<th>NO</th>
<th>DON'T KNOW</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

1.3
An IQ score is a measure of brightness that takes into account both score on an intelligence test and age.

<table>
<thead>
<tr>
<th>YES</th>
<th>NO</th>
<th>DON'T KNOW</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>YES</td>
<td>NO</td>
</tr>
<tr>
<td>---</td>
<td>-----</td>
<td>----</td>
</tr>
<tr>
<td>1.4</td>
<td>Norms are to regarded as desirable levels of attainment.</td>
<td>10.</td>
</tr>
<tr>
<td>1.5</td>
<td>The extent to which a test does the job for which it is used is called test reliability.</td>
<td>11.</td>
</tr>
<tr>
<td>1.6</td>
<td>A percentile score is the rank expressed in percentage terms.</td>
<td>12.</td>
</tr>
<tr>
<td>1.7</td>
<td>A raw score is a direct numerical report of a person’s performance.</td>
<td>13.</td>
</tr>
<tr>
<td>1.8</td>
<td>Grade equivalent scores are obtained by averaging the raw scores of samples of pupils at different grade levels in the publisher’s norm group.</td>
<td>14.</td>
</tr>
<tr>
<td>1.9</td>
<td>The terms norms and standards are synonymous.</td>
<td>15.</td>
</tr>
<tr>
<td>1.10</td>
<td>The p value is an index commonly associated with the difficulty of an item.</td>
<td>16.</td>
</tr>
<tr>
<td>1.11</td>
<td>Stanines are obtained by averaging a designated percentage of raw scores in a distribution to each stanine category.</td>
<td>17.</td>
</tr>
<tr>
<td>1.12</td>
<td>The difference between the lowest and highest scores obtained on a test by some group is called the range.</td>
<td>18.</td>
</tr>
<tr>
<td>1.13</td>
<td>The score in a distribution which divides the group into two equal parts, is called the mean.</td>
<td>19.</td>
</tr>
<tr>
<td>1.14</td>
<td>The score or value that occurs most often in a distribution is called the mode.</td>
<td>20.</td>
</tr>
<tr>
<td>1.15</td>
<td>The symbol N is commonly used to represent the number of cases in a distribution.</td>
<td>21.</td>
</tr>
</tbody>
</table>
AREA 2  TEST OF INTERPRETATION

**Directions:** For items 2.1 through 2.4 use the test data furnished below.

![Test Data Table]

2.1  According to the test results, this student should be performing at the third grade level in math.  

2.2  The student's PR of 76 in reading comprehension indicates a low average academically.  

2.3  The DIQ is the student's true IQ.  

2.4  This student might be labeled as an underachiever in reading comprehension.  

2.5  The GE score is the most reliable score.  

2.6  A grade equivalent score indicates a steady increase in the average performance throughout the grades.  

2.7  A stanine of 8 on reading comprehension indicates high average ability.  

2.8  An individual's obtained score on a test is usually not meaningful until it is compared with other examinees' scores.  

2.9  The intelligence quotient is a simple method of expressing the relationship between a pupil's mental age and chronological age.
2.10
The two types of derived scores based on average or median performance are the grade score and the age score.

2.11
Subtest scores may be too unreliable to be used effectively.

2.12
Raw scores, by themselves, have little meaning.

2.13
Scores used in reporting standardized tests are absolute rather than relative.

AREA 3 TEST OF APPLICATION

3.1
A fourth grade student who earns a grade score of 6.7 in arithmetic has shown that he has the arithmetical ability of an average sixth grade student.

3.2
It would be wise to compare the GS scores in vocabulary and reading comprehension for instructional planning.

3.3
The intelligence quotient is a good predictor of probable success in school.

3.4
Grade based performance data are helpful in forming subgroups within the classroom for specific instructional activities.

3.5
The percentile rank has the disadvantage that units are not equal all along the scale measurement.

3.6
Schools should identify underachievers and overachievers on the basis of intelligence test scores.
<table>
<thead>
<tr>
<th>Question</th>
<th>YES</th>
<th>NO</th>
<th>DON'T KNOW</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.7 Grade and age norms are most appropriate for elementary school subjects that are studied continuously over a long series of grades.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.8 Grade and age equivalents are usually indicated to tenths of a unit.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.9 Scores on one standardized test may be substituted for scores on another standardized test since they are equivalent.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.10 Performance on intelligence tests is related to achievement.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.11 A given difference in percentile rank reflects a large change in raw scores near the average.</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Appendix B

Teacher Concepts of Standardized Achievement Tests in the Public Schools

Directions: The purpose of this survey is to measure your concept of standardized tests use in the elementary schools of your school district. Complete the rating sincerely, fairly rapidly and spontaneously. If you feel that your concept of standardized tests use is Very Closely related to the higher end of the scale, mark at the appropriate end. If you feel that your concept of standardized tests use is Neutral, mark the middle space. If you feel your concept is Very Closely related to the lower end of the scale, mark at the appropriate point on the scale toward that end.

Standardized Achievement Tests

1. Stimulating ; ; ; ; ; ; ; ; Dull
2. Fair ; ; ; ; ; ; ; Unfair
3. Ugly ; ; ; ; ; ; ; Beautiful
4. Hindering ; ; ; ; ; ; ; Helpful
5. Fresh ; ; ; ; ; ; ; Stale
6. Tense ; ; ; ; ; ; ; Relaxed
7. Valuable ; ; ; ; ; ; ; Worthless
8. Clear ; ; ; ; ; ; ; Hazy
9. Dishonest ; ; ; ; ; ; ; Honest
10. Angry ; ; ; ; ; ; ; Calm
11. Kind ; ; ; ; ; ; ; Cruel
12. Practical ; ; ; ; ; ; ; Impractical
13. Sour ; ; ; ; ; ; ; Sweet
14. Static ; ; ; ; ; ; ; Dynamic
15. Relevant ; ; ; ; ; ; ; Irrelevant
Appendix C

Dallas, Texas
October 1, 1979

Dear

First, may I express my appreciation for your cooperation and aid in the validation of the instrument for my dissertation study.

My study is entitled, A Study of Standardized Test Knowledge and Interpretation by Elementary Classroom Teachers. The purposes of the study are to determine the following: (1) elementary teachers' knowledge of standardized test terminology, (2) elementary teachers' interpretation of standardized test scores, (3) elementary teachers' use of test information for evaluation and planning, and (4) elementary teachers' attitude toward standardized tests in the public schools. (The last purpose will be determined by a separate instrument.) The instrument will be administered to second-, fourth-, and sixth-grade teachers in two metroplex, suburban school districts in the Fall of 1979.

For each of the items on the instruments please circle one of the three numerals on the left side of the page. The numerals indicate (1) the item is important to the study, (2) you cannot decide on the validity of the item, or (3) the item is not clearly stated, or does not pertain to the study. An additional sheet is provided for any additions and/or corrections you may feel are necessary for the questionnaire to be valid. A self-addressed envelope is enclosed for returning the instrument.

Thank you, again, for your time and expertise.

Respectfully,
Appendix D

Validity Test of the "Standardized Test Proficiencies for Elementary Teachers"

<table>
<thead>
<tr>
<th>Item</th>
<th>Juror 1</th>
<th>Juror 2</th>
<th>Juror 3</th>
<th>Juror 4</th>
<th>Juror 5</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Area 1</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1.2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1.3</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1.4</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1.5</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1.6</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1.7</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1.8</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1.9</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1.10</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1.11</td>
<td></td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1.12</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1.13</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1.14</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1.15</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td><strong>Area 2</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2.2</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2.3</td>
<td>3</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>1*</td>
</tr>
<tr>
<td>2.4</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2.5</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2.6</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>2.7</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2.8</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2.9</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2.10</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2.11</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2.12</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2.13</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2.14</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Item</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>------</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
</tr>
<tr>
<td>3.1</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>1*</td>
</tr>
<tr>
<td>3.2</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>3.3</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>3.4</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>3.5</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>3.6</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>3.7</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>3.8</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>3.9</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>3.10</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>3.11</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>3.12</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>1</td>
</tr>
</tbody>
</table>

*Items dropped.
Appendix E

Standardized Test Proficiencies for Elementary Teachers

Administrator Questionnaire

Name ____________________________

Position ____________________________

School District ____________________________

Directions: Please read the enclosed instrument and determine how many questions a teacher should answer correctly to be considered proficient in standardized test use in the public schools.

<table>
<thead>
<tr>
<th>Area</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Area 1: Standardized Test Terminology</td>
<td></td>
</tr>
<tr>
<td>Area 2: Test of Interpretation</td>
<td></td>
</tr>
<tr>
<td>Area 3: Test of Application</td>
<td></td>
</tr>
</tbody>
</table>

Signature ____________________________

Date ____________________________
Appendix F

Dallas, Texas
December 31, 1979

Dear

The enclosed instrument is to be administered to second-, fourth-, and sixth-grade teachers in suburban public school districts as part of a dissertation study. The purpose is to determine teacher competencies in standardized test use. What is (teacher scores) will be compared with what should be (administrator opinion).

May I request your expertise in providing what you would consider sufficient competencies for teachers to use standardized test information effectively? This is simply your opinion of what would constitute a passing score (percentage) on each area of the test. This request will be sent to twelve area administrators who participate in student program decisions based on his or her school district's standardized test results.

Thank you for your time and expertise. If you have any questions please contact me at 492-3237.

Respectfully,
### Appendix G

Administrator Ratings of the "Standardized Test Proficiencies for Elementary Teachers" for Desired Teacher Levels of Proficiency

<table>
<thead>
<tr>
<th>Administrator</th>
<th>Percentages</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Area 1</td>
<td>Area 2</td>
<td>Area 3</td>
<td>Average</td>
</tr>
<tr>
<td>1</td>
<td>53</td>
<td>92</td>
<td>91</td>
<td>79</td>
</tr>
<tr>
<td>2</td>
<td>47</td>
<td>92</td>
<td>73</td>
<td>71</td>
</tr>
<tr>
<td>3</td>
<td>60</td>
<td>85</td>
<td>73</td>
<td>73</td>
</tr>
<tr>
<td>4</td>
<td>80</td>
<td>85</td>
<td>82</td>
<td>82</td>
</tr>
<tr>
<td>5</td>
<td>80</td>
<td>75</td>
<td>75</td>
<td>77</td>
</tr>
<tr>
<td>6</td>
<td>80</td>
<td>85</td>
<td>90</td>
<td>85</td>
</tr>
<tr>
<td>7</td>
<td>73</td>
<td>92</td>
<td>90</td>
<td>85</td>
</tr>
<tr>
<td>8</td>
<td>87</td>
<td>54</td>
<td>82</td>
<td>74</td>
</tr>
<tr>
<td>9</td>
<td>66</td>
<td>85</td>
<td>45</td>
<td>65</td>
</tr>
<tr>
<td>10</td>
<td>86</td>
<td>76</td>
<td>100</td>
<td>87</td>
</tr>
<tr>
<td>11</td>
<td>80</td>
<td>85</td>
<td>63</td>
<td>76</td>
</tr>
<tr>
<td><strong>Mean</strong></td>
<td>72</td>
<td><strong>82.4</strong></td>
<td>79</td>
<td>78</td>
</tr>
</tbody>
</table>
Appendix H

"Standardized Test Proficiencies for Elementary Teachers"

Test-Retest Reliability

<table>
<thead>
<tr>
<th>Subject</th>
<th>Score X</th>
<th>Score Y</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>25</td>
<td>21</td>
</tr>
<tr>
<td>2</td>
<td>24.5</td>
<td>20</td>
</tr>
<tr>
<td>3</td>
<td>21</td>
<td>26.5</td>
</tr>
<tr>
<td>4</td>
<td>19.5</td>
<td>18.5</td>
</tr>
<tr>
<td>5</td>
<td>19</td>
<td>17.5</td>
</tr>
<tr>
<td>6</td>
<td>18.5</td>
<td>23.5</td>
</tr>
<tr>
<td>7</td>
<td>16.5</td>
<td>13</td>
</tr>
<tr>
<td>8</td>
<td>16</td>
<td>14.5</td>
</tr>
<tr>
<td>9</td>
<td>15.5</td>
<td>20</td>
</tr>
<tr>
<td>10</td>
<td>15</td>
<td>17</td>
</tr>
<tr>
<td>11</td>
<td>15</td>
<td>22.5</td>
</tr>
<tr>
<td>12</td>
<td>15</td>
<td>14</td>
</tr>
<tr>
<td>13</td>
<td>14.5</td>
<td>17</td>
</tr>
<tr>
<td>14</td>
<td>14</td>
<td>10</td>
</tr>
<tr>
<td>15</td>
<td>13.5</td>
<td>17</td>
</tr>
<tr>
<td>16</td>
<td>13</td>
<td>10</td>
</tr>
<tr>
<td>17</td>
<td>12.5</td>
<td>14</td>
</tr>
<tr>
<td>18</td>
<td>11.5</td>
<td>6</td>
</tr>
<tr>
<td>19</td>
<td>11</td>
<td>12</td>
</tr>
<tr>
<td>20</td>
<td>10.5</td>
<td>17</td>
</tr>
<tr>
<td>21</td>
<td>10.5</td>
<td>11.5</td>
</tr>
<tr>
<td>22</td>
<td>9.5</td>
<td>13.5</td>
</tr>
</tbody>
</table>

\[ y_c = 4.584 \]
\[ t(20) = 3.824 \]
\[ r = .6503 \]
\[ r^2 = .4222 \]
\[ t^*(.05, 20) \]
\[ r = \text{significant} \]
BIBLIOGRAPHY

Books


Articles


Cassidy, Jack, "Reporting Pupil Progress in Reading--Parents vs. Teachers," The Reading Teacher, 31 (December, 1977), 294-296.


Hopper, Gordon, "Parental Understanding of Their Child's Test Results as Interpreted by Elementary School Teachers," Measurement and Evaluation in Guidance, 10 (July, 1977), 84-89.

Houts, Paul L., "Behind the Call for Test Reform and Abolition of the IQ/Culture Issue," Phi Delta Kappan, 57 (June, 1976), 674-676.

Hubbard, R. E. and W. R. Fletcher, "Intelligence Tests and Intelligent Teachers, Do They Agree?" Educational Research Bulletin, 32 (1953), 113-140.


Nolte, Chester, "Use and Misuse of Tests in Education: Legal Implications," Educational Horizons, 54 (Fall, 1975), 10-16.

Oliver, Jo Ellen and Richard D. Arnold, "Comparing A Standardized Test, An Informal Inventory and Teacher Judgment on Third-Grade Reading," Reading Improvement, 15 (Spring, 1978), 56-59.

Plumleigh, George, "Achievement Test Scores: Are They of Any Use?" Phi Delta Kappan, 58 (April, 1977), 614-615.


Reports


Publications of Learned Organizations


"Testing," National Education Association Division of Instruction and Research, 1973, p. 44.

Public Documents


Unpublished Materials